SYSTEMATIC REVIEWS

Prediction of chronic kidney disease and its progression by artificial intelligence algorithms

Francesco Paolo Schena[1,2] · **Vito Walter Anelli**[2] · **Daniela Isabel Abbrescia**[3] · **Tommaso Di Noia**[2]

Abstract

Background and objective Aim of nephrologists is to delay the outcome and reduce the number of patients undergoing renal failure (RF) by applying prevention protocols and accurately monitoring chronic kidney disease (CKD) patients. General practitioners and nephrologists are involved in the first and in the late stages of the disease, respectively. Early diagnosis of CKD is an important step in preventing the progression of kidney damage. Our aim was to review publications on machine learning algorithms (MLAs) that can predict early CKD and its progression.

Methods We conducted a systematic review and selected 55 articles on the application of MLAs in CKD. PubMed, Medline, Scopus, Web of Science and IEEE Xplore Digital Library of the Institute of Electrical and Electronics Engineers were searched. The search terms were chronic kidney disease, artificial intelligence, data mining and machine learning algorithms.

Results MLAs use enormous numbers of predictors combining them in non-linear and highly interactive ways. This ability increases when new data is added. We observed some limitations in the publications: (i) databases were not accurately reviewed by physicians; (ii) databases did not report the ethnicity of the patients; (iii) some databases collected variables that were not important for the diagnosis and progression of CKD; (iv) no information was presented on the native kidney disease causing CKD; (v) no validation of the results in external independent cohorts was provided; and (vi) no insights were given on the MLAs that were used.

Overall, there was limited collaboration among experts in electronics, computer science and physicians.

Conclusions The application of MLAs in kidney diseases may enhance the ability of clinicians to predict CKD and RF, thus improving diagnostic assistance and providing suitable therapeutic decisions. However, it is necessary to improve the development process of MLA tools.

Keywords Chronic kidney disease · Renal failure · Artificial intelligence · Machine learning models · Prediction

Introduction

Chronic kidney disease (CKD) is characterized by reduced glomerular filtration rate (GFR) (≤ 60 ml/min/1.73 m^2) and increased proteinuria (≥ 0.5 g/day), as established by the Kidney Disease Improving Global Outcomes (KDIGO) study group on the evaluation and management of CKD

✉ Francesco Paolo Schena
paolo.schena@uniba.it

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

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

3 Schena Foundation, Bari, Italy

[1]. In 2015, the Global Burden of Disease Study [2] report announced that the worldwide prevalence of CKD ranged between 8 and 16% of the causes of the total number of global deaths. Practically, 750 million people worldwide suffered from CKD [3] and this morbidity was ranked among the first 20 causes of deaths.

Unfortunately, CKD is a common disorder that is underestimated by patients and health care providers. Many patients are unaware of their CKD diagnosis [4]. Therefore, early detection of CKD and rapid intervention are essential to slow down disease progression and improve both the outcome and quality of life.

The outcome of CKD is renal failure (RF). Therefore, the aim of nephrologists is to delay the outcome and reduce the number of patients reaching RF through prevention programs and accurate monitoring of CKD patients. This

process involves general practitioners with regard to the first CKD stages (1 and 2), and nephrologists in the later stages (3 to 5) of the disease.

Health informatics has produced a great number of electronic medical records (EMRs) and processing of these data generates new knowledge. In general, datasets suffer from missing data and noisy values; therefore, data mining techniques are used to reduce errors, thereby improving data quality. Artificial intelligence (AI) uses data mining systems to discover meaningful patterns in a database. Furthermore, there are different techniques for extracting knowledge from databases. Feature selection is a technique used to reduce the number of variables in a database, remove noise, and improve accuracy. All these methods improve learning performance and build better machine learning algorithms (MLAs) which can be divided into two groups according to the kind of learning: supervised and unsupervised learning.

In the supervised techniques, the task outcome (i.e. a label) is associated with each sample, whereas there is no ground truth for the task outcome in the unsupervised learning strategy. Supervised techniques are often used for classification to distinguish different patterns. Feature selection for the classification tasks consists of four basic steps: subset generation, subset evaluation, stopping criterion and result validation. These techniques can be used to develop clinical decision-making support systems that can be adopted in clinical practice. These tools support clinicians in diagnosis or in the prediction of prognosis. Finally, deep learning models improve the MLAs because they automatically compute an "abstract" interpretation of data, resulting in accurate algorithms.

In this systematic review our aim was to analyze different AI algorithms that predict CKD and RF.

Methods

We conducted a systematic review following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [5].

Strategy of research

We developed a data form (Supplementary Table 1) to extract the main information from PubMed, Medline, Scopus, Web of Science and IEEE *Xplore* Digital Library of the Institute of Electrical and Electronics Engineers. The search terms were chronic kidney disease, artificial intelligence, data mining and machine learning algorithms. We searched for additional publications by checking the references quoted in the studies.

Study selection

Titles and abstracts of all publications were screened for eligibility. The potentially relevant publications were retrieved and assessed for final eligibility. Disagreements were resolved through discussion and consultation of a third person.

Data extraction

Institutions, dataset (origin), number of patients, ethnicity, number of attributes, training set, test set, validation set, metrics (accuracy, sensitivity, specificity, precision, recall, F measure), comparison with other algorithms (classifiers), machine learning algorithms with best accuracy, and type model of task were collected. Finally, remote clinical impact connectivity (availability of a website and/or App), acceptability by physicians and patients, and clinical impact were evaluated.

Study quality assessment

Risk of bias of included publications, mainly focusing on the selection of participants, definition of variables, and appropriate use of MLAs were appraised. Disagreements were resolved through the discussion and consultation of a third person.

Data synthesis

The findings of the publications were synthesized in tables described in the MLA sessions. Meta-analysis was not performed due to the wide variability of the publications with regard to study design, population and MLAs.

The authors, FPS, VWA and DIA, identified studies by searching PubMed, Medline, Scopus, Web of Science and IEEE *Xplore* Digital Library of the Institute of Electrical and Electronics Engineers.

Two authors (FPS and VWA) independently screened the titles and abstracts of all publications for eligibility. The potentially relevant publications were retrieved and assessed for final eligibility. Disagreements were resolved through discussion and consultation of a third author (TD).

FPS and VWA independently appraised the risk of bias of the included publications, mainly focusing on the selection of participants, definition of variables, and appropriate use of the appraised MLAs. Disagreements were resolved through the discussion and consultation of a third author (DT).

Results

Searching analysis

Figure 1 shows the flow diagram of our systematic review. The websites and the ISEE digital library quoted 387 articles between January 1998 and June 2021, and 52 other studies were extracted from the chosen publications. We screened 439 publications, and excluded 346 articles such as duplications, review articles, editorials, abstracts and publications on MLAs in dialysis and renal transplantation. Ninety-three records were assessed for eligibility but 38 were excluded because CKD in kidney diseases was not evaluated. Finally, 55 studies were selected.

Different algorithms were used by investigators, therefore classifier distribution in AI subsets are summarized in Fig. 2. AI includes Fuzzy logic systems (FLSs), random forest classification (RFC), decision tree (DT), and naïve Bayes (NB).
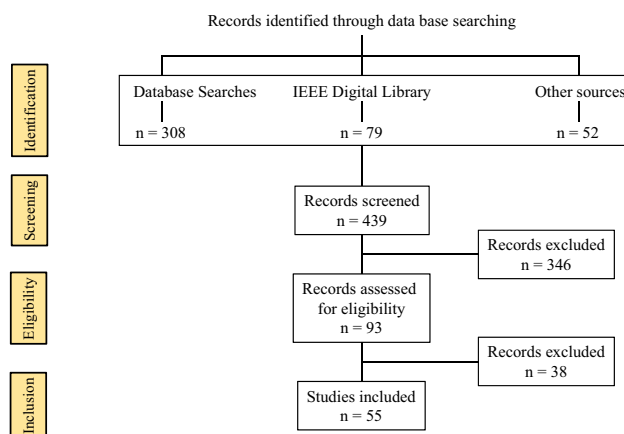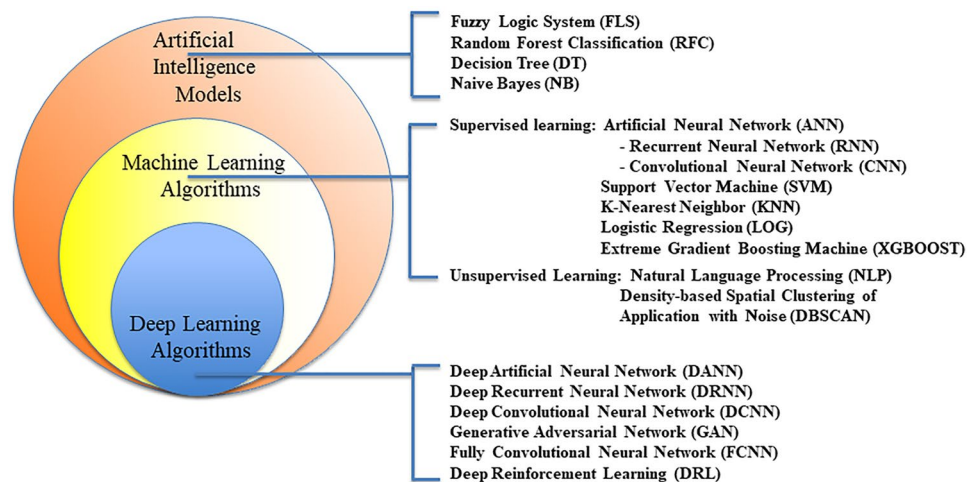

Fig. 1 PRISMA flow diagram of the review process

The MLA subset includes supervised learning algorithms, such as artificial neural networks (ANNs), support vector machine (SVM), k-nearest neighbors (KNNs), logistic regression (LOG) and gradient boosting (GB).Density-based spatial clustering of application with noise (DBSCAN) and various techniques belonging to the natural language processing (NLP) research field are among the unsupervised learning algorithms. The deep learning (DL) subset includes algorithms such as deep recurrent neural networks (DRNNs) and deep convolutional neuronal networks (DCNNs).

Machine learning models to predict CKD

Various AI models were employed to predict CKD. In seven publications, which are illustrated in Table 1, the researchers used *Fuzzy Learning Systems*. Akgundogdu et al. [6] used an adaptive neuro-Fuzzy inference system (ANFIS) to predict CKD in 112 patients, of whom 64 had renal damage. This system took advantage of the ANN theory (i.e. processing data) to determine the FLS truth values. Thus, the machine learning system was able to acquire knowledge from the database within a shorter training time and to represent it in fuzzy form. Seven variables were included in the algorithm. The number of individuals included in the training and test set are shown in Table 1. The model was compared with other machine learning models (SVM and ANN). The results showed that the ANFIS model was more efficient (accuracy 100%) than the others. The low number of patients in the training and test sets means low number of patients in the two cohorts of patients.

A FLS containing seven variables was used by Ahmed et al. [7] in 817 patients. The model was implemented by the Mamdani system, and the results ranged between 1 (healthy) and 6 (extremely sick). The model, which was tested on 30 random patients in the same cohort, had an accuracy of 86.37%. However, the study had important limitations. First,

Fig. 2 List of algorithms in artificial intelligence, machine learning and deep learning used in the reviewed publications

Table 1 Seven selected publications of Fuzzy Logic System (FLS) applications to predict Chronic Kidney Disease

Authors (year)	Dept IEEE	Dept Medicine	Country	Data set repository	Data set size (subjects)	Population	Input variables	sCr	GFR	Training set	Testing set	Accuracy (%)	Comparison with other MLAs	Task model
Akgundogdu et al. (2010)	✓	✓	Turkey	Istanbul	112	Turkey	7	Y	N	85	27		SVM, ANN	C
Ahmed et al. (2014)	✓		Bangladesh	Dhaka	817	Bangladesh	7	N	Y	787	30	86.7		R
Muslim et al. (2015)	✓		Indonesia	Indonesia		Indonesia	6	Y	Y	100	75	98.9	KNN, ANN	C
Chen et al. (2016)	✓		China	UCI	400	India	25	Y	N			99.6		C
Zarandi et al. (2018)	✓		Iran	Chamran	400	Iran	9	Y	N	320	80	80.0		C
Abdolka-rimzadeh et al. (2018)	✓		Iran	Chamran	600	Iran	15	Y	N	480	120	90.0		C
Hamedan et al. (2020)		✓	Iran	Tehran	216	Iran	16	Y	Y			92.1		

ANNs artificial neural networks, *C* classifier, *FLS* fuzzy logic system, *GFR* glomerular filtration rate, *IEEE* institute of electrical and electronics engineers, *KNN* K-nearest neighbor, *MLA* machine learning algorithm, *N* no, *R* regression *sCR* serum creatinine, *SVM* support vector machine, *UCI* university of California Irvine, *Y* yes

proteinuria was not evaluated among the considered variables. Second, the system was not validated in an external cohort of patients.

An expert system based on the Mamdani Fuzzy inference system was developed to predict CKD by Muslim et al. [8]. The design of CKD was developed using the fuzzy logic toolbox in Matlab R2009a. After feeding the system the demographic and laboratory data, the use of the defuzzification method improved the diagnosis of CKD with a 98.86% accuracy level. Surprisingly, sex and proteinuria were not considered.

Chen et al. [9] used two in-house fuzzy classifiers (fuzzy rule-building expert system and fuzzy optimal associative memory) to predict CKD. Both fuzzy tools were compared with partial least squares discriminant analysis. The CKD database of the Sri Lanka cohort (UCI repository) contained 400 patients, of whom 251 with CKD were used. Some patients and attributes were removed due to missing data; thus, 386 patients were included in the study, and 21 variables were used. The performance metrics demonstrated that both fuzzy classifiers had the same accuracy. However, the fuzzy rule-building expert system was more robust than the others. Cross-validation was realized by performing 200 bootstraps with four partitions in the original dataset. The system was not validated in an external cohort of patients.

Zarandi and Abdolkarimzadeh [10] generated a type-1 fuzzy inference system (FIS), and they improved the FIS with ANFIS. They generated an algorithm that was trained and tested on a cohort of 400 patients. Thus, a neuro Fuzzy classification to diagnose CKD was developed. Then, the same group of researchers [11] proposed a new model, the interval type-II fuzzy system, which analyzed 15 variables in a dataset of 600 patients with two classes: CKD and not CKD. They obtained an accuracy of 90%. The validation process in external cohorts was not considered in either of the publications.

Recently, Hamedan et al. [12] developed a fuzzy system that was evaluated using data extracted from 216 randomly selected medical records of patients with and without CKD. After identifying diagnostic variables carried out by reviewing the literature and organizing a survey among 18 nephrologists, 16 variables were selected and used to predict CKD. High values of accuracy, sensitivity and specificity were reported. However, the algorithm did not detect the different stages of CKD, the degree of disease, or the recommended treatment.

Table 2 shows the *Random Forest classification* algorithm used in six studies. The model is an accurate and noise-resistant classifier in which data inclusion and random feature selection are combined. Siyad and Manoj [13] studied the prediction of CKD in a cohort of 400 patients of whom 150 were healthy children, using four classification strategies: NB, RFC, J48 (i.e. a decision tree) and

LOG classifier. They found the best accuracy (99.75%) with RFC. However, the paper has several limitations, mainly due to the presence of adult patients and children in the same cohort. No data are reported on the training and test cohort. Moreover, the accuracy of RFC was not validated in an external cohort of patients.

Salekin and Stankovic [14] compared three classifiers (KNN, RFC and ANN) to predict CKD in the dataset of the Sri Lanka cohort that included 400 patients of whom 200 were in the early stages of CKD. Before moving on with the comparison of the classifiers, they reduced the number of variables using the wrapper method and the LASSO regularization. Thus, they identified the most important predictive variables for CKD. The results showed that the RFC algorithm, with 12 attributes can detect CKD with the highest accuracy. Next, the authors reduced the number of attributes to five (urinary specificity gravity, albumin, diabetes mellitus, hypertension and hemoglobin) to reduce the computational cost. However, the predictive value of these five attributes, of which sex and age were not considered, is questionable.

Different machine learning techniques (ANN, SVM, KNN, C4.5 DT, RFC) were used by Subas et al. [15] to predict CKD in the dataset of the Sri Lanka cohort. Performance analysis showed 100% accuracy in the RFC algorithm, but it was not validated in another cohort of patients.

Gunarathne et al. [16] developed an automated machine learning system to predict CKD in a cohort of 400 patients living in Sri Lanka, where the disease has high prevalence. Fourteen variables were selected (proteinuria was not included). Four models (multiclass decision forest, multiclass decision jungle, multiclass logistic regression, and multiclass neural network) were tested. The decision forest models showed the best accuracy (99.1%).

Lakshmi et al. [17] used the same system used by Subas et al. [15]. Ten variables were selected from a total of 25 attributes. The dataset was divided into a training set (300 individuals) and a test set (100 individuals) The RFC algorithm distinguished subjects with CKD from individuals without CKD well. However, the accuracy value of the approved model was not reported. Moreover, the article showed many limitations in the description of the patient cohort and the validation process.

Qin et al. [18] used six machine learning techniques (LOG, RFC, SVM, KNN, NB, ANN) to determine the most accurate learning system to diagnose CKD for a dataset of the Sri Lanka cohort (400 patients). The authors analyzed 24 variables, some of which were questionable (e.g. pedal edema can be replaced by spot urinary protein or daily proteinuria). The RFC algorithm had the best accuracy (99.75%). The authors suggested an integrated model that combines logistic regression and RFC which increased the

Table 2 Six selected publications of Random Forest Classification (RFC) applications to predict Chronic Kidney Disease

Authors (year)	Dept IEEE	Dept Medicine	Country	Data set repository	Data set size (subjects)	Population	Input variables	sCR	GFR	Training set	Testing set	Accuracy (%)	Comparison with other MLAs	Task model
Siyad et al. (2016)	✓		Kerala	UCI	400	India	25	Y	N			99.8	NB, J48, LOG	C
Salekin et al. (2016)	✓		U.S.A	UCI	400	India	25	Y	N			99.8	KNN, ANN	C
Subas et al. (2017)	✓		Saudi Arabia	UCI	400	India	25	Y	N			100.0	ANN, SVM, KNN, DT-C4.5	C
Gunarathne et al. (2017)	✓		Sri Lanka	UCI	400	India	25	Y	N			99.1	LOG, ANN	C
Lakshmi et al. (2019)	✓		India	UCI	400	India	25	Y	N	300	100			C
Qin et al. (2019)	✓		China	UCI	400	India	25	Y	N			99.8	ANN, NB, KNN, SVM, LOG	C

ANNs artificial neural networks, *C* classifier, *DT-C4.5* decision tree-C4.5 algorithm, *GFR* glomerular filtration rate, *IEEE* institute of electrical and electronics engineers, *J48* J48 algorithm, *KNN* K-nearest neighbor, *LOG* logistic regression, *MLA* machine learning algorithm, *N* no, *NB* naïve bayes, *RFC* random forest classification, *sCR* serum creatinine, *SVM* support vector machine, *UCI* university of California Irvine, *Y* yes

accuracy (99.83%). However, the model was not validated in a large cohort of patients.

In Table 3 we list seven studies considering three types of *Decision Tree* algorithms (ID3, CART and C4.5) that can process both numerical and categorical data. Al-Hyari et al. [19] compared the ANN model with two clinical decision support systems (NB and DT) for the diagnosis of CKD in the UCI dataset. The DT algorithm was the most accurate classifier with the highest value of accuracy value (92.2%). It predicted normal renal function or CKD in the first step and CKD stage in the second step. Unfortunately, the algorithm was not externally validated.

Tazin et al. [20], after preprocessing the data, compared four algorithms (NB, DT, SVM and KNN) using the WEKA tool in a dataset of 400 patients from the UCI repository, USA [21]. Twenty-five variables were provided, and 10 were chosen to detect CKD. The results comparing the four algorithms showed that DT had the best AUROC curve and accuracy (90%). The authors observed that the correct number of variables included in the algorithms dramatically changed the performance of prediction. The DT algorithm was not validated in an external cohort of patients.

Furthermore, Sharma et al. [22] analyzed twelve different classifiers based on five techniques (DT, SVM, DA, KNN, ANN) using the UCI dataset containing 400 people living in the southern part of India (Sri Lanka cohort). Twenty-four attributes were considered. The DT technique performed better than the others, with an accuracy of 98.6%.

Padmanaban and Parthiban [23] tested two machine learning techniques (NB and DT) to build training data, classify the predictive model, and test the classification efficiency. Twelve attributes were used for the comparison. The DT algorithm appeared to have the best performance accuracy (91%). No source concerning the dataset that was used was reported.

Serpen [24] carried out a study on the dataset of the Sri Lanka cohort which was made up of 400 patients, of whom 200 were affected by CKD. He used the C4.5 DT algorithm, which is a classifier composed of nodes that divide the pattern space into more subspaces depending on a specific discrete function of the input attribute variables. The leaves were then assigned to one class denoting the most suitable target value. Classification of the patterns is performed by directing them from the tree root down to a leaf, based on the test outcomes along the pathway [25]. Data of 393 patients were processed. Then, the algorithm was validated by four physicians. The conclusion was that a set of eight variables can be used to predict CKD with an accuracy of 98.25%.

Boukenze et al. [26] used different machine learning techniques (SVM, MLP, DT-C4.5, NB, and KNN) to predict CKD in a dataset of 400 patients (Sri Lanka cohort). Twenty-four variables and two classes (CKD and not CKD) were considered. The experimental study was developed

Table 3 Seven selected publications of Decision Trees (DTs) to predict Chronic Kidney Disease

Authors (year)	Dept IEEE	Dept Medicine	Country	Data set repository	Data set size (subjects)	Population	Input variables	sCR	GFR	Accuracy (%)	Comparison with other MLAs	Task model
Al-Hyari et al. (2013)	✓		Jordan	UCI	102	Jordan	25	Y	Y	92.2	ANN, NB	C, R
Tazin et al. (2016)	✓		Bangladesh	UCI	400	India	25			99.0	NB, SVM, KNN	
Sharma et al. (2016)	✓	✓	India	UCI	400	India	25			98.6	SVM, DA, KNN, ANN	
Padmanaban et al. (2016)	✓		India	nd	nd	nd	12			91.0	NB	
Serpen et al. (2016)	✓	✓	U.S.A	UCI	400	India	25			98.3		
Boukenze et al. (2017)	✓		Morocco	UCI	400	Morocco	25			63.0	SVM, MLP, NB, KNN	
Shi et al. (2020)	✓	✓	Taiwan	Taiwan	19,720	Taiwan	8+1			82.0	CART, ELM, LDA	

ANNs artificial neural networks, *C* classifier, *CARTs* classification and regression trees, *DA* discriminant analysis; *DT* decision tree, *ELM* extreme learning machine, *GFR* glomerular filtration rate, *IEEE* institute of electrical and electronics engineers, *KNN* K-nearest neighbor, *LDA* linear discriminant analysis, *MLA* machine learning analysis, *MLP* multilayer perceptron, *NB* naïve bayes, *R* regression, *sCR* serum creatinine, *SVM* support vector machine, *UCI* university of California Irvine, *Y* yes

along three steps: pre-process, classification, and evaluation. The best value of accuracy (63%) was reached by DT-C4.5.

Shih et al. [27] tested four data mining algorithms (classification and regression tree, C4.5 decision tree, linear discriminant analysis, and extreme learning machine) to predict early CKD in a large cohort of individuals (19,270 Taiwanese adult patients of whom 5,101 affected by CKD). Eight independent variables that were significantly different in CKD patients and one dependent outcome variable (GFR) were used to compare the robustness of the four algorithms. Training and testing samples were used to measure the predictive ability of the models. Results showed that the C4.5 algorithm performed better than the other models with an accuracy of 82%.

The *Naïve Bayes* algorithm, based on Bayes' theorem with the assumption of independence between every pair of data [28], was used in three publications, which are shown in Table 4. Kunwar et al. [29] compared two different algorithms (NB and ANN) to predict CKD in a dataset of 400 subjects. The accuracy of NB was 100%. Dulhare and Ayesha [30] performed data processing to clean and transform the missing values, to eliminate noisy data such as outliers, and normalize and balance unbalanced data. Next, they realized a feature selection using the OneR algorithm in the dataset. Finally, the NB classifier was used to predict CKD in the dataset that contained 25 variables. The NB algorithm combined with the One R technique reduced the number of attributes from 25 to only 5 and improved the accuracy (97.5%).

Singh N and Singh P [31] performed a comparative study using nine rule-based classification methods to predict CKD on a scarcely described dataset. The hybrid decision table-naïve Bayes classifier had the highest AUROC value (99.9%). The authors of these three articles used different algorithms, however, the results were not validated in external independent cohorts.

The *Artificial Neural Network* models are inspired by the neuronal network structure of our brain. An ANN is an adaptive system that changes its structure based on external and internal information that flows through the network during the learning phase. Kriesel [32] stated that an ANN is a bioinspired mechanism of data processing that enables computers to learn similarly to a human-being's brain. The power of the ANN algorithm comes from its ability to represent linear and non-linear relationships and to learn them directly from the input data. The ANN model works well with noisy data and is used for both supervised and unsupervised learning. The common ANN model is also called multilayer perceptron (MLP). ANN algorithms have been used for pattern recognition, classification, and prediction of CKD based on the established variables. We referenced thirteen publications, which are illustrated in Table 5. Chiu et al. [33] compared three different types of ANN models

Table 4 Three selected publications of Naïve Bayes (NB) to predict Chronic Kidney Disease

Authors (year)	Dept IEEE	Dept Medicine	Country	Data set repository	Data set size (subjects)	Population	Input variables	sCr	GFR	Accuracy (%)	Comparison with other MLAs	Task model
Kunwar et al. (2016)	✓		India	UCI	400	India	25	Y	N	100.0	ANN	C
Dulhare et al. (2016)	✓		India	UCI	400	India	25	Y	N	97.5		C
Singh et al. (2017)	✓		India	UCI	400	India	25	Y	N	98.3	CART, DT- C4.5, NNge, CR, OneR, Ridor, PART	C

ANNs artificial neural networks, *C* classifier, *CARTs* classification and regression trees, *CR* conjunctive rule algorithm, *DT-C4.5* decision tree-C4.5 algorithm, *GFR* glomerular filtration rate, *IEEE* institute of electrical and electronics engineers, *MLA* machine learning algorithm, *N* no, *OneR* one rule classification algorithm, *NB* naïve bayes, *NNge* non-nested generalization algorithm, *PART* partial decision tree algorithm, *R* regression, *Ridor* ripple down rule learner, *sCR* serum creatinine, *UCI* university of California Irvine, *Y* yes

Table 5 Thirteen selected publications of Artificial Neural Networks (ANNs) to predict Chronic Kidney Disease

Authors (year)	Dept IEEE	Dept Medicine	Country	Data set repository	Data set size (subjects)	Population	Input variables	sCR	GFR	MLA	Trainig set	Testing set	Accuracy (%)	Comparison with other classifiers	Task model
Chiu et al. (2013)	✓		Taiwan	Taiwan Hospital	430	Taiwan	7	Y	Y	BPN+GA	300	100	91.7	BPN, GFNN, MNN	C
Di Noia et al. (2013)	✓	✓	Italy	Italy	575	Italy	10			MLP			87.7		C
Neves et al. (2015)	✓		Portugal	Portugal	558	Europe	9			ANN			90.0		
Rubini et al. 2015	✓		India	UCI	400	India	25	Y	N	MLP			99.7	RBF, LR	C
Ramya et al. (2016)	✓		India	Coimbatore	1000	India	15	Y	N	RBF			85.3	ANN, BPNN, RFC	C
Borisagar et al. (2017)	✓		India	UCI	400	India	25	Y	N	LMA	280	60	99.8	MNN	C
Yildirim, (2017)	✓		Turkey	UCI	400	India	25	Y	N	RM			99.8	SMOTE, SSS	C
Aljaaf et al. (2018)	✓		Iraq	UCI	400	India	25	Y	N	MLP			99.5	RT, SVM, LR	C
Zeynu et al. (2018)	✓		India	UCI	400	India	25	Y	N	ANN			99.5	KNN, J48, ANN, NB, SVM	C
Abdelaziz et al. (2019)	✓		Egypt	UCI	400	India	25	Y	N	ANN	280	120	97.8		C
Almansour et al. (2019)	✓		Saudi Arabia	UCI	400	India	25	Y	N	ANN			99.7	SVM	C
Manonmani et al. (2020)	✓		India	UCI	400	India	25	Y	N	CNN			97.7	SVM, GB	C
Sharifi et al. (2020)	✓		Iran	Iran	140	Iran	12			ANN			98.0		C

ANNs artificial neural networks, *BPN* back propagation network, *BPNN* back propagation neural network, *C* classifier, *CNN* convolutional neural network, *GA* genetic algorithm, *GFNN* generalized feed forward neural network, *GB* gradient boosting, *GFR* glomerular filtration Rate, *IEEE* institute of electrical and electronics engineers, *J48* J48 algorithm, *KNN* K-nearest neighbor, *LMA* levemberg-marquardt algorithm, *LR* logistic regression, *MLA* machine learning algorithm, *MLP* multi-layer perceptron, *MNN* modular neural network, *N* No, *NB* naïve bayes, *RA* resemble algorithm, *RBF* radial basis function, *RFC* random forest classification, *RM* resample method, *RT* regression tree, *sCR* serum creatinine, *SMOTE* synthetic minority over-sampling technique, *SSS* spread sub sample, *SVM* support vector machine, *UCI* university of California Irvine, *Y* yes

to predict CKD in a dataset of 430 patients of whom 285 were affected by CKD. Seven variables were chosen by the investigators. The model was developed by considering three steps associated with different sets of patients: 300 for the training test, 100 for the test set, and 30 for the validation set. The NBuilder tool was adopted for the development of the model. The neural network, combined with the genetic algorithm, showed the best accuracy (91.7%). Clinical data regarding the cohort of patients were not reported and the tool was not validated in an external independent cohort.

Di Noia et al. [34] designed a tool to predict CKD in patients affected by biopsy-proven immunoglobulin A nephropathy (IgAN). The ANN model was developed using a dataset of 575 Italian IgAN patients and 10 variables. A multilayer perceptron architecture with 10 inputs, 6 hidden neurons, and 1 output neuron with a sigmoid activation function was chosen. The tool was available online as a web application and as an Android mobile app.

Neves et al. [35] developed a hybrid decision support system based on logic programming and complemented with ANNs. The latter were considered due to their adaptability, robustness and flexibility. The study was carried out on a population of 558 adult patients of whom 175 had CKD. The dataset contained 24 variables, which were reduced to nine. The tool had good performance on the training and test set but was not validated in an external cohort of patients.

Rubini and Eswaran [36] compared three classifiers (radial basis function networks, multilayer perceptron and logistic regression) to predict CKD in a dataset of the UCI repository in which 63% of patients were affected by CKD. Performance analysis showed that the ANN model, which used the Levenberg–Marquardt algorithm, had the best accuracy (99.8%). The ANN was not validated in an external cohort of patients.

Ramya and Radha [37] used four machine learning algorithms, including ANN, Radial Basis Function, and RFC, to diagnose CKD in a dataset of 1,000 patients who were referred to different laboratories in Coimbatore, India. Fifteen variables were considered; however, proteinuria was not included. The comparative study demonstrated that the radial basis function algorithm had the best accuracy (85.3%) when predicting CKD.

Borisagar et al. [38] used a multilayer ANN model to detect CKD in a dataset of the UCI machine learning repository that contained 24 input variables and one class attribute. Missing values in the dataset were replaced with the most probable value. Four different back-propagation algorithms (Levenberg–Marquardt, scaled conjugate, Bayesian regularization, and resilient back propagation) were used. The processes of training, validation and testing were carried out on 70%, 15% and 15% of cases, respectively. The Levenberg–Marquardt algorithm had the best accuracy. The algorithm was not validated in an external cohort of patients.

Furthermore, some variables (pedal edema, bacteria and pus cell) were found to have no influence on the development of CKD.

Yildrim [39] compared three over-sampling algorithms (resample, synthetic minority over-sampling technique, and spread sub-sample) combined with ANNs with different learning rate values in the dataset of the Sri Lanka cohort composed of 400 patients, of whom 200 were affected by CKD. The WEKA 3.7.3 software was used. The results showed that the resample algorithm reached the best accuracy (99.8%). The best execution time was obtained by the spread sub-sample method. The algorithm was not validated in an external cohort of patients.

Zeynu et al. [40] used five algorithms (KNN, J48, ANN, NB and SVM) to predict CKD in a dataset of 400 patients (Sri Lanka cohort), of whom 200 were affected by CKD. Twenty variables were included in the selected models. The system combined algorithms implemented by the authors in Java and others using the WEKA tool. The results showed that accuracy in predicting CKD increased when algorithms were used together. The study was realized solely using training and test sets.

Aljaaf et al. [41] conducted an accurate statistical analysis of 24 attributes of the UCI dataset repository containing data from a cohort of 400 Indian patients. Only seven informative variables were used for the four machine learning algorithms (regression tree, SVM, logistic regression and ANN) to predict CKD. The results showed that the multilayer perceptron neural network had the highest AUROC value and the lowest type I error. Therefore, they concluded that this model had the best performance accuracy (99.5%). The model was validated in an external cohort of patients.

Abdelaziz et al. [42] developed a hybrid intelligent model based on linear regression and ANN to predict CKD in smart cities using cloud computing and internet of things. The authors found 13 out of 24 attributes to be critically influential factors for CKD diagnosis. Unfortunately, proteinuria was not included. The ANN model was implemented in Windows Azure. The accuracy of the model was 97.8%. The validation test was conducted on only three patients.

Almansour et al. [43] studied CKD prediction using two machine learning algorithms, ANN and SVM, in a cohort of 400 Indian subjects of whom 200 were affected by CKD. The dataset, released by the Apollo Hospitals, Tamil Nadu, India, was first preprocessed and then prepared for the study. Several experiments were carried out to determine the optimized parameters for the two algorithms, and the final models were developed. The results showed that ANN performed better than SVM. Unfortunately, deep learning systems were not used in this study. The tool was not validated in an external cohort of patients.

Manonmani and Balakrishnan [44] used an array of techniques to select features involving density-based feature

selection as a filter to rank the attributes of the CKD dataset of the UCI repository. Then, the results of this filtering process were given to a wrapper-based optimization technique to determine the optimal feature set for CKD prediction. The results showed that the feature subset was made up of 19 attributes combined with a convolutional neural network (CNN), and it had the best accuracy (97.75%) compared with SVM and GB. The results were not validated in an external cohort of patients.

Sharifi and Alizadeh [45] predicted CKD using a novel classifier model based on a multilayer perceptron-ANN algorithm in a population of 50 healthy subjects and 90 patients from Iran. Twelve attributes were adopted. The results showed that this algorithm had high accuracy (98%). However, the major limitations of this study were due to the low number of patients, and proteinuria was not present in the list of attributes. The ANN model was not validated.

Support Vector Machines are supervised learning techniques that perform classification and regression tasks. The algorithms find the optimal hyperplane that separates one-class samples from those of other classes, while maximizing the margin between the two classes [46]. SVM algorithms were used in five studies, as shown in Table 6. Al-Hyari et al. [47] analyzed the performance of different classifiers (ANN, NB, DT, SVM and LOG) to predict CKD in a dataset of 102 patients collected at the Nephrology Department in Prince Hamza Hospital in Amman, Jordan. Fifteen attributes were chosen with the aid of nephrologists. The classifiers were implemented using the Waikato Environment for Knowledge Analysis tool. SVM showed the best accuracy (93.14%). The authors adopted the inaccurate term CRF (chronic renal failure) instead of CKD. Among the 15 chosen attributes, proteinuria, which plays an important role in the progression of renal damage, was not considered. Furthermore, the machine learning algorithm was not validated.

Chen et al. [48] studied three supervised models as SVM, KNN and soft independent modeling of class analogy (SIMCA) to evaluate the CKD risk in a cohort of 386 patients. The robustness and abilities of the models to tolerate noise data were investigated. The SVM model performed better than other algorithms and tolerated noise disturbance better. However, the algorithm was not validated in an external cohort.

Charleonnan et al. [49] evaluated the prediction of CKD in a cohort of 400 Indian patients of the Apollo Hospitals deposited in the UCI repository. The dataset contained 11 numeric and 14 nominal attributes. Four machine learning algorithms (SVM, LOG, DT and KNN) were trained and tested. The SVM classifier showed the best accuracy (99%). The model was not validated in an external cohort of patients.

Polat et al. [50] used wrapper and filter approaches to remove non-appropriate inputs and reduce the

Table 6 Five selected publications of Support Vector Machine (SVM) to predict Chronic Kidney Disease

Authors (year)	Dept IEEE	Dept Medicine	Country	Data set repository	Data set size (subjects)	Population	Input variables	sCR	GFR	Accuracy (%)	Comparison with other MLAs	Task model
Al-Hyari et al. (2014)	✓	✓	Jordan	Jordan	102	Jordan	15	Y	N	93.1	ANN, LOG, NB, DT	C
Chen et al. (2016)	✓		China	UCI	400	nd	25	Y	N	99.0	KNN, SIMCA	C
Charleonnan et al. (2016)	✓		Thailand	UCI	400	Indian	27	Y	N	99.0	KNN, DT, LOG	C
Polat et al. (2017)	✓		Turkey	UCI	400	nd	25	Y	N	98.5		
Ravindra et al. (2018)	✓		India	UCI	400	nd	25	Y	N	93.8		

ANNs artificial neural networks, *C* classifier, *DTs* decision trees, *GFT* glomerular filtration rate, *IEEE* institute of electrical and electronics engineers, *KNN* K-nearest neighbor, *LOG* logistic regression, *MLA* machine learning algorithm, *NB* naïve bayes, *N* no, *sCR* serum creatinine, *SIMCA* soft independent modeling by class analogy, *SVM* support vector machine, *UCI* university of California Irvine, *Y* yes

dimensionality of the dataset obtained from the Sri Lanka cohort composed of 400 patients, of whom 200 were affected by CKD. The SVM algorithm had an accuracy of 98.5% after filtering the dataset. The limitations of this study were the low number of patients and the non-validation via an external cohort of patients.

Ravindra et al. [51] used the SVM classifier to predict CKD in the UCI dataset containing data of 400 subjects, of whom 200 were affected by CKD, and 24 attributes. The SVM classifier had an accuracy of 93.75%. The model was not validated in an external independent cohort.

The *K-Nearest Neighbors* algorithm belongs to the class of memory-based algorithms. They directly exploit data to perform classification without first building a model [52]. The only hyperparameter in the model is k, the number of nearest neighbors to include in the estimate of class membership. By varying k, the model can consider more or less neighbors. It is robust when applied to noisy training data and effective with large training datasets if the algorithm is implemented using sparse matrices. There are different types of similarity to be used with KNN such as fine, medium, cosine, cubic, and weighted. We referenced four publications exploiting KNNs, as shown in Table 7. Sinha and Sinha [53] conducted a comparative study to predict CKD using two different algorithms (SVM and KNN). They used a dataset designed with findings obtained by many laboratories, centers and hospitals. The work, carried out using the Matlab tool, showed that the KNN classifier had the best accuracy (80.90%). No description of the cohort was reported, and no validation of the machine learning algorithm was carried out.

Chetty et al. [54] applied three classifiers (NB, IBK and sequential minimal optimization) in the dataset of 400 patients with 24 variables and one class attribute. After reducing the attributes from 24 to seven, they found the best accuracy using the IBK algorithm. The attributes that were used were not described, and the model was not validated in an external cohort of patients.

Ani et al. [55] compared five algorithms (NB, KNN, DT, random subspace classification algorithm, LDA) in a population of 400 Indian patients, of whom 200 were affected by CKD. The random subspace classification algorithm with KNN had the best accuracy (94%). The model was not validated.

Wibawa et al. [56] compared KNN with other algorithms (NB, SVM) after a selection of the attributes using correlation-based feature selection and AdaBoost to improve CKD prediction. The UCI dataset had 24 input attributes and one class attribute. Correlation-based feature selection was applied to multivariate data. The authors evaluated the ability of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy among them. AdaBoost was used to

Table 7 Four selected publications of K-Nearest Neighbors (KNNs) to predict Chronic Kidney Disease

Author(s)	Dept IEEE	Dept Medicine	Country	Data set repository	Data set size (Subjects)	Population	Input variables	sCR	GFR	Accuracy (%)	Comparison with other MLAs	Task model
Sinha et al. (2015)	✓		India	UCI	400	India	25	Y	N	78.8	SVM	C
Chetty et al. (2015)	✓		India	UCI	400	India	25	Y	N	100.0	NB, SMO	C
Ani et al. (2016)	✓		India	UCI	400	India	25	Y	N	94.0	ANN, NB, KNN, RSCA, LDA	C
Wibawa et al. (2017)	✓		Indonesia	UCI	400	India	25	Y	N	98.1	NB, KNN	C

ANNs artificial neural networks, *C* classifier, *GFR* glomerular filtration rate, *IEEE* institute of electrical and electronics engineers, *KNN* K-nearest neighbor, *LDA* linear discriminant analysis, *MLA* machine learning algorithm, *N* no, *NB* naïve bayes, *RSCA* random subspace classification algorithm, *sCR* serum creatinine, *SMO* sequential minimal optimization, *SVM* support vector machine, *UCI* university of California Irvine, *Y* yes

select class prediction and the weight of each model. Seven attributes were removed, and among the three algorithms, KNN showed the best accuracy (98.10%); CFS and Adaboot improved CKD diagnosis.

Machine learning models to predict RF

The doctor-patient relationship is based on mutual trust. However, the first question a CKD patient usually asks the physician is: "Doctor, will I develop RF?" If the answer is yes, the second question will be "Doctor, when will I need renal replacement therapy?". The answers to these questions affect the patient's quality of life. Patients with kidney diseases have high heterogeneity of clinical symptoms, progression and response to treatment. Therefore, in the last few decades, many risk prediction models, based on mathematical models, have been developed for some kidney diseases in order to be one step ahead of RF and to enable more personalized therapy. The decline in renal function is slower in the early stages of CKD (G1 and G2) and faster in the later stages (G3 and G4); therefore, accurate long-term follow-up is needed.

We reviewed ten publications, as listed in Table 8. Pesce et al. [57] developed a new model based on two neural networks, the first predicted RF and the second predicted the time needed to develop RF, ranging between three and more than 8 years. The ANN model, based on six variables, was developed using an international cohort of 1040 IgAN patients from Italy, Norway and Japan. The tool performance was validated by five experienced nephrologists.

Singh et al. [58] used three different models to incorporate longitudinal electronic medical records obtained from 6,435 patients. Nine variables were considered. Proteinuria was not present in the list of variables. They found that the multitask-temporal learning model was able to exploit the temporal dynamics in the data and improve the prediction of short-term loss in kidney function.

Longitudinal inpatient and outpatient EMRs were combined with narrative clinical documents using automated information extraction as the natural language processing. Perotte et al. [59] used this approach to monitor the progression of CKD from stage 3 to stage 4 in a cohort of 2,908 primary-care clinic patients. Twenty-one variables were included in the model. Ethnicity was not considered. Thus, they developed a risk model of prediction using three types of information (demographic, laboratory, and clinical documentation). Three Kalman filter risk models were tested. The laboratory test combined with the Kalman filter test had the highest agreement (85%) regarding CKD progression. In conclusion, this was the first predictive model for CKD progression from stage 3 to ESKD that included comorbidities, common laboratory tests, and other vital signs.

Norouzi et al. [60] proposed an ANFIS to predict RF in CKD patients. The learning algorithm was a hybrid model in which an ordinary least square algorithm was used to update coefficients of output functions, while the back propagation algorithm was used to update the weights of the system. EMRs collected in a cohort of 465 CKD patients over a period of 10 years were used for the development of the tool. The model was able to predict GFR with high accuracy at intervals of 6, 12, and 18 months. Of the 10 independent variables selected, the model used only four (weight, diastolic blood pressure, diabetes mellitus and time GFR) which had the highest correlation with estimated MDRD-GFR. The authors concluded that this model could be used by physicians in the clinical practice of monitoring the impact of each variable by routinely measuring CKD in patients during follow-up. However, the study has important limitations, such as the low number of patients used for the development of the model; furthermore, the tool was not validated in an external cohort of patients. Another limit was the absence of some variables in the model that can influence kidney damage, such as age, gender, and proteinuria.

An ANFIS-based medical decision support system was developed by Yadollahpour et al. [61] in a cohort of 465 patients to predict, first, the early stages of CKD and, next, RF in sequential periods of 6, 12, and 18 months. This system integrates ANN and FLS principles into a single framework to use non-linear functions. The model was available to monitor CKD progression and predict the time for renal replacement therapy. The authors used three predictive models and found that ANFIS predicted CKD better than linear regression and multilayer perceptron neural networks. A user-friendly graphical interface was designed in MATLAB to build a medical decision support system to predict the GFR values in CKD patients. This approach could reduce both the cost of CKD management and the mortality rate, and generally improve the quality of life of patients.

Kanda et al. [62] studied the progression of CKD in a healthy Japanese population, which included 7,465 subjects who received annual check-ups and for whom data was collected over a period of seven years. Variables and their causal relationships were studied using the Bayesian network, and the bootstrapping method was used to average the networks. Then, they used the SVM models to detect outcome (CKD development) in both high and low risk subjects at baseline. The results suggested that it was necessary to follow-up both high and low risk individuals for three years. The suggested check-up was questionable as it is very expensive. For this reason, check-ups were recommended for subjects who were at high risk of kidney damage, such as aged individuals with hypertension, cardiovascular diseases, and diabetes, as well as those with proteinuria. The principal limitation of this study is

Table 8 Ten selected publications of machine learning algorithms to predict Chronic Kidney Disease and its progression

Authors (year)	Dept IEEE	Dept Medicine	Country	Data set repository	Data set size (subjects)	Population	Input variables	sCR	GFR	MLA	Training set	Testing set	Accuracy (%)	Comparison with other MLAs	Task model
Pesce et al. (2015)	✓	✓	Italy, Norway, Japan	Italy, Norway, Japan	1040	Italy, Norway, Japan	6			ANN			91.6		C, R
Singh et al. (2015)	✓	✓	USA	USA	6435	USA	9	N	Y	MTLM	5148	1287			C, R
Perotte et al. (2015)	✓	✓	USA	USA	2908	USA	21	Y	Y	NLP	2617	291	84.9		R
Norouzi et al. (2016)	✓	✓	Iran	Tehran	465	Iran	24	Y	Y	FLS	327	139	95.0		R
Yadollahpour et al. (2018)	✓	✓	Iran	Tehran	465	Iran	25	Y	Y	FLS	325	140		LOG, MLP	C,R
Kanda et al. (2019)	✓	✓	Japan	Japan	7465	Japan	13	N	Y	SVM					R
Xiao et al. (2019)	✓	✓	China	China	551	China	18	Y	Y	LOG	440	111	87.3	EN, XGB, LASSO R, SVM, RFC, ANN	R
Weber et al. (2020)	✓	✓	Germany	Germany	785	Germany	4	N	Y	LOG, RF, ANN					R
Song et al. (2020)	✓	✓	USA	USA	14,039	USA	8	N	Y	XGB	11,184	2855			C, R
Schena et al. (2021)	✓	✓	Italy	Europe	948	Europe	7	Y	Y	ANN	758	190	89.0		C

ANNs artificial neural networks, *EN* elastic network, *FLS* fuzzy logic system, *IEEE* institute of electrical and electronics engineers, *KNN* K-nearest neighbor, *LOG* logistic regression, *LASSO R* lasso regression, *MLA* machine learning algorithm, *MLP* multi-layer perceptron, *MTLM* multitask temporal learning model, *N* no, *NLP* natural language processing, *RFC* random forest classification, *RG* ridge regression, *sCR* serum creatinine, *SVM* support vector machine, *XGB* X gradient boost, *Y* yes

the short follow-up and the fact that the SVM technique was not compared with other models.

Xiao et al. [63] compared nine techniques using linear models (LOG, elastic net, lasso and ridge regression) and non-linear models (ANN, SVM, RF, XGBoost and KNN) to predict RF in 551 Chinese patients with proteinuria. Eighteen variables were considered. They found that the LOG had the best AUROC (0.87) and accuracy (82%). A Prediction System, which is based on a combined model (LOG, Elastic Net and XGboost), was developed for clinical practice. However, this tool was not validated in an external independent cohort of patients.

Weber et al. [64] combined some laboratory EMRs (serum creatinine and eGFR) and ICD-10 billing codes to detect CKD in a cohort of 785 elderly multimorbid patients with a mean age of 75 years, of whom 373 had CKD stage 3 or higher. The laboratory values had the best performance compared to the ICD-10 billing codes. LOG and some machine learning algorithms (generalized linear model via penalized maximum likelihood, RFC, and ANN) improved the distinction between CKD (Stage 3 or higher) and non-CKD. The limitation of this study is that the model focused on patients with advanced CKD but not on patients with CKD 1 or 2.

Song et al. [65] proposed a temporal-enhanced GB machine model that incorporated the patients' longitudinal data. The model predicted 1-year CKD risk. The study, which was conducted on a retrospective cohort of 14,039 adult diabetic patients, demonstrated that exploratory historical temporal EMRs in the landmark-boosting model would significantly improve prediction performance. This model continuously calculated kidney risk in patients with an automatic collection of new EMRs include demographic and laboratory data during the follow-up. The strength of the model was the use of the patients' longitudinal data and long-term follow-up. The limitation was the reduced sensitivity of the models' overtime.

Recently, Schena et al. [66] developed an ANN model to predict CKD in a cohort of 986 patients with IgAN. The model was based on two ANNs; the first one predicted CKD (yes or no); the second one predicted the time it would take to develop RF. The classifier model had a performance of 0.82 in patients with a follow-up of five years and 0.89 in those with a follow-up of 10 years. Similar results were obtained over a 25-year follow-up period. The model was validated in an external independent cohort of 167 IgAN patients. The tool (DialCheck) is available on the web site (www.poliba.igan.it) and on an Android or iOS mobile app.

Discussion

CKD is a term that includes various kinds of kidney diseases. It is asymptomatic in the early stages and symptomatic when the disease becomes more severe. CKD can be prevented by early diagnosis and appropriate therapy. Therefore, the prediction of CKD is important for the life of the patients. Clinical practice in nephrology is based on the collection of EMRs that can be analyzed by standard statistical models or by AI. In the last few decades, clinical observational studies, randomized clinical trials and renal registries of patients with CKD have generated large amounts of clinical data. AI can use non-linear relations between variables that are very common in the outpatient clinical practice.

In this review, we analyzed the sequential steps for promoting the application of machine learning techniques to predict CKD and its impairment. In our opinion, AI may change our methodological approach in clinical activities because the virtual branch, based on informatics, may improve diagnosis, prognosis and therapeutic decision making.

We reviewed 55 published articles and paid particular attention to the following key points: (1) collaboration between experts in AI- and physicians; (2) datasets used to train and test the classifier; (3) intrinsic correlations and the target to be investigated; (4) development of the algorithms; (5) cross-validation test; (6) user-friendly web-servers for physicians and patients.

Data analysis showed that several MLAs were used by investigators who obtained a high level of accuracy; however, several relevant drawbacks were found in these studies.

First, the studies were designed, carried out, and reviewed by experts in intelligent systems. There was sporadic collaboration among experts in electronics, computer science, and physicians. A limited number of studies, i.e., 16/55 (29%), reported collaboration among experts. Therefore, few patient cohorts were revised by physicians. In many publications, the database was obtained from different sources, and the variables were chosen based on their statistical relevance. The database obtained from the Sri Lanka cohort, deposited in the UCI repository, was frequently used by the investigators, but there was a great deal of criticism with regard to the medical terms used and to the data collected. Sex was not considered, even though CKD is more frequent in men than in women [67]. The term albumin is not correct; it should have been replaced by albuminuria or daily proteinuria. Two attributes describe the hypertensive state of a subject: blood pressure and hypertension. The term hypertension (yes or no) is sufficient to describe the status. The low specific gravity of the urine occurs only in the late stages of CKD; therefore, this attribute is not necessary to predict early CKD. Pus cell, pus cell clumps, and bacteria are etiologic factors that have nothing to do with CKD; these laboratory findings are only related to urinary infections, which are responsive to targeted antibiotics. The number of white blood cells does not change in CKD. The term "pedal edema" should have been replaced by "swelling lower limbs". The anemia attribute is unnecessary if hemoglobin values and the number of

Table 9 Recommendations to improve the development of the MLA tool to predict CKD and RF

1. Strict collaboration between artificial intelligence- and physician- investigators
2. Accurate screening of combined variables included in the MLAs
3. Use of database from large cohorts of subjects (potentially, 1000 or more)
4. Validation of the tool by an external independent cohort (preferentially, in prospective clinical studies)

red blood cells were present. In conclusion, the most important variables in predicting CKD that should be taken into consideration are age, sex, blood pressure, serum creatinine, eGFR, and 24 h proteinuria. These variables, as described in the KDIGO guidelines [68], have been used in a limited number of mathematical models and in a few papers on machine learning techniques.

In many instances, neither serum creatinine nor proteinuria were included among the chosen attributes for the machine learning algorithms. Proteinuria is the first known biomarker responsible for tubulointerstitial damage and CKD progression [69].

The etiopathogenesis of CKD depends on the type of kidney disease, therefore collaboration with physicians is necessary in order to select the correct attributes. In our opinion, studies predicting CKD should be conducted in homogeneous cohorts of patients, i.e., patients with the same kidney disease.

Second, many reviewed studies contained analyses performed on databases containing a limited number of patients; moreover, they suffered from noisy and missing data. To build an accurate data mining model with high accuracy ($> 90\%$), it is necessary to collect clinical data from one thousand or more patients with a limited amount of missing data. Furthermore, data must be validated in an external independent cohort of patients. This process was present in only three publications [24, 57, 66].

Third, in many publications ethnicity was not reported; however, the important role of this factor in estimating GFR was recently discussed [70], and it should be taken into consideration when applying MLAs [71]. Another aspect that has not been considered is the geographic area (developing or developed country) as hygienic factors can influence the onset and outcome of different forms of glomerulonephritis [72]. This aspect should be considered in the future, especially when it comes to using the general term of CKD, that could be secondary to different kinds of kidney diseases.

Fourth, in many of the above mentioned publications, there is no information on the type of kidney disease diagnosed in patients. This point is relevant because the clinical course of CKD varies depending on the cause: glomerular diseases, diabetes, obesity, hypertension, and other cardiovascular diseases. The generic term 'CKD', used in many of the reviewed publications contains different etiologies and consequently different prognoses. Accurate classification of CKD patients, based on the type of kidney disease,

may improve the use of machine learning techniques, thus prolonging survival and improving target therapy. Management of these patients might be different because disease heterogeneity is a major hurdle [73].

Fifth, the studies analyzed in this review developed classification models and did not offer the insight gained from the analysis of the dataset on which the classifier models were used. In other words, once these classifier models were induced from the dataset, they were used as black-box decision making tools.

Sixth, the validation process was conducted via an external cohort of patients in only a few of the reviewed publications. This process is important to estimate the model's performance and should be interpreted by two different models (reproducibility and transportability). Debray et al. [74] proposed a framework of the methodological steps in an empirical example of deep vein thrombosis.

Seventh, the development of user-friendly web-servers for physicians and patients is the main aim of AI studies. We found that this was clarified in only four publications [7, 34, 57, 66].

The strength of our study is that, to our knowledge, we conducted the first systematic review on MLAs in CKD using search terms in multiple databases following the PRISMA guidelines. We also carefully considered many different aspects illustrated in the enclosed tables. The main limitation is that we did not perform a meta-analysis due to the heterogeneity of the reviewed studies, tspecifically related to the design, population and comparators used.

Compared to classical mathematical models, MLAs have the ability to explicitly outline the strong non-linear relations between variables that are common in outpatient clinical practice. Therefore, as shown in Table 9, we suggest (i) collaboration between experts in artificial intelligent systems and physicians to design and develop user-friendly, clinical decision support systems to help physicians in clinical practice; (ii) accurate screening of variables; (iii) use of databases from large cohorts of patients (potentially, 1000 or more), and lastly, (iv) the tools must be validated in external independent prospective cohorts of patients.

Declarations

Conflict of interest All authors declare no competing interests.

Ethical approval This is a systematic review carried out on papers which received ethical approval.

References

1. Levey AS, Atkins R, Coresh J et al (2007) Chronic kidney disease as a global public health problem: approaches and initiatives—a position statement from kidney disease improving global outcomes. Kidney Int 72:247–259. https://doi.org/10.1038/sj.ki.5002343
2. Global burden of disease study (2013) Collaborators (2015) global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the global burden of disease study 2013. Lancet 386:743–800. https://doi.org/10.1016/S0140-6736(15)60692-4
3. Lozano R, Naghavi M, Foreman K et al (2012) Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010. Lancet 380:2095–2128. https://doi.org/10.1016/S0140-6736(12)61728-0
4. McIntyre NJ, Fluck R, McIntyre C, Taal M (2012) Treatment needs and diagnosis awareness in primary care patients with chronic kidney disease. Br J Gen Pract 62:e227–e232. https://doi.org/10.3399/bjgp12X636047
5. Liberati A, Altman DG, Tetzlaff J et al (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 339:b2700. https://doi.org/10.1136/bmj.b2700
6. Akgundogdu A, Kurt S, Kilic N, Ucan ON, Akalin N (2010) Diagnosis of renal failure disease using adaptive neuro-fuzzy inference system. J Med Syst 34:1003–1009. https://doi.org/10.1007/s10916-009-9317-2
7. Ahmed ATMS, Kabir T, Mahmood NT, Rahman RM (2014) Diagnosis of kidney disease using fuzzy expert system. 8th Int Conf Softw, Knowl, Information Manag Appl (SKIMA 2014). https://doi.org/10.1109/SKIMA.2014.7083522
8. Muslim MA, Kurniawati I, Sugiharti E (2015) Expert system diagnosis chronic kidney disease based on mamdani fuzzy inference system. J Theor Appl Inf Technol 78:70–75
9. Chen Z, Zhang Z, Zhua R, Xiang Y, Harringtonb PB (2016) Diagnosis of patients with chronic kidney disease by using two fuzzy classifiers. Chemometr Intell Lab Syst 153:140–145. https://doi.org/10.1016/j.chemolab.2016.03.004
10. Zarandi MHF, Abdolkarimzadeh M (2018) Fuzzy rule based expert system to diagnose chronic kidney disease. In: Melin P, Castillo O, Kacprzyk J, Reformat M, Melek W (eds) Fuzzy Logic in intelligent system design NAFIPS 2017 advances in intelligent systems and computing. Springer, Cham. https://doi.org/10.1007/978-3-319-67137-6_37
11. Abdolkarimzadeh M, Zarandi MHF, Castillo O (2018) Interval type II fuzzy rough set rule based expert system to diagnose chronic kidney disease. In: Barreto G, Coelho R (eds) Fuzzy information processing NAFIPS 2018 communications in computer and information science. Springer, Cham. https://doi.org/10.1007/978-3-319-95312-0_49
12. Hamedan F, Orooji A, Sanadgol H, Sheikhtaheri A (2020) Clinical decision support system to predict chronic kidney disease: a fuzzy expert system approach. Int J Med Inform 138:104134. https://doi.org/10.1016/j.ijmedinf.2020.104134
13. Siyad MB, Manoj M (2016) Fused features classification for the effective prediction of chronic kidney disease. IJIRST 2:44–48
14. Salekin A, Stankovic J (2016) Detection of chronic kidney disease and selecting important predictive attributes. IEEE (ICHI). https://doi.org/10.1109/ICHI.2016.36
15. Subasi A, Alickovic E, Kevric J (2017) Diagnosis of chronic kidney disease by using random forest. IFMBE Proc. https://doi.org/10.1007/978-981-10-4166-2_89
16. Gunarathne WHSD, Perera KDM, Kahandawaarachchi KADCP (2017) Performance evaluation on machine learning classification techniques for disease classification and forecasting through data analytics for chronic kidney disease (CKD). IEEE Int Conf (BIBE). https://doi.org/10.1109/BIBE.2017.00-39
17. Lakshmi SV, Meena FK, Kiruthika NS (2019) Diagnosis of chronic kidney disease using random forest algorithms. Int J Eng Sci Manag 2:559–562
18. Qin J, Chen L, Liu Y, Liu C, Feng C, Chen B (2019) A machine learning methodology for diagnosing chronic kidney disease. IEEE Access 8:20991–21002. https://doi.org/10.1109/ACCESS.2019.2963053
19. Al-Hyari AY, Al-Taee AM, Al-Taee MA (2014) Diagnosis and classification of chronic renal failure utilising intelligent data mining classifiers. (IJITWE) 9:1–12. https://doi.org/10.4018/ijitwe.2014100101
20. Tazin N, Sabab SA, Chowdhury MT (2016) Diagnosis of chronic kidney disease using effective classification and feature selection technique. Int Conf Med Eng, Health Informatics Technol (MediTec). https://doi.org/10.1109/MEDITEC.2016.7835365
21. Scholar LR UCI-Machine Learning Repository available at https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease. Accessed Mar 2021
22. Sharma S, Sharma V (2016) Performance based evaluation of various machine learning classification techniques for chronic kidney disease diagnosis. Int J Mod Comput Sci Appl 4:11–16
23. Padmanaban KRA, Parthiban G (2016) Applying machine learning techniques for predicting the risk of chronic kidney disease. Indian J Sci Technol 9:1–5. https://doi.org/10.17485/ijst/2016/v9i29/93880
24. Serpen AA (2016) Diagnosis rule extraction from patient data for chronic kidney disease using machine learning. Int J Biomed Clin Eng 5:64–67. https://doi.org/10.4018/IJBCE.2016070105
25. Maimon O, Rokach L (2010) Data mining and knowledge discovery handbook, 2nd edn. Springer, New York, Dordrecht, Heidelberg, London. https://doi.org/10.1007/978-0-387-09823-4
26. Boukenze B, Haqiq A, Mousannif H (2017) Predicting chronic kidney failure disease using data mining techniques. In: El-Azouzi R, Menasche D, Sabir E, De Pellegrini F, Benjillali M (eds) Advances in ubiquitous networking 2 UNet 2016 lecture notes in electrical engineering. Springer, Singapore
27. Shih C-C, Lu C-J, Chen G-D, Chang C-C (2020) Risk prediction for early chronic kidney disease: results from an adult health examination program of 19,270 individuals. Int J Env Res Public Health 17:4973. https://doi.org/10.3390/ijerph17144973
28. Domingos P, Pazzani M (1997) On the optimality of the simple bayesian classifier under zero-one loss. Mach Learn 29:103–130. https://doi.org/10.1023/A:1007413511361
29. Kunwar V, Chandel K, Sabitha AS, Bansal A (2016) Chronic kidney disease analysis using data mining classification techniques. 6th Int Conf—Cloud Syst Big Data Eng (Conflu). https://doi.org/10.1109/CONFLUENCE.2016.7508132
30. Dulhare UN, Ayesha M (2016) Extraction of action rules for chronic kidney disease using naïve bayes classifier. IEEE (ICCIC). https://doi.org/10.1109/ICCIC.2016.7919649

31. Singh N, Singh P (2017) Rule based approach for prediction of chronic kidney disease: a comparative study. Biomed Pharm J 10:867–874. https://doi.org/10.13005/bpj/1179

32. Kriesel D (2007) A brief introduction to neural networks. http://www.dkriesel.com. Accessed Mar 2021

33. Chiu RK, Chen RY, Wang S-A, Chang YC, Chen LC (2013) Intelligent systems developed for the early detection of chronic kidney disease. Adv Artif Neural Syst. https://doi.org/10.1155/2013/539570 (**Article ID 539570**)

34. Di Noia T, Ostuni VC, Pesce F 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

35. Neves J, Martins MR, Vilhena J et al (2015) A soft computing approach to kidney diseases evaluation. J Med Syst 39:131. https://doi.org/10.1007/s10916-015-0313-4

36. Rubini LG, Eswaran P (2015) Generating comparative analysis of early stage prediction of chronic kidney disease. Int J Mod Eng Res Technol 5:49–55

37. Ramya S, Radha N (2016) Diagnosis of chronic kidney disease using machine learning algorithms. IJIRCCE 4:812–820. https://doi.org/10.15680/IJIRCCE.2016

38. Borisagar N, Barad D, Raval P. (2017) Chronic Kidney Disease Prediction Using Back Propagation Neural Network Algorithm. In: Modi N, Verma P, Trivedi B (eds) Proceedings of international conference on communication and networks advances in intelligent systems and computing, vol 508. Springer, Singapore. https://doi.org/10.1007/978-981-10-2750-5_31

39. Yildirim P (2017) Chronic kidney disease prediction on imbalanced data by multilayer perceptron: chronic kidney disease prediction. IEEE 41st Annual Computer Softw Appl Conf (COMPSAC). https://doi.org/10.1109/COMPSAC.2017.84

40. Zeynu S, Patil S (2018) Prediction of chronic kidney disease using feature selection and ensemble method. Int J Pure Appl Math 118:1–16

41. Aljaaf AJ, Al-Jumeily D, Haglan HM et al (2018) Early prediction of chronic kidney disease using machine learning supported by predictive analytics. IEEE Congr Evolut Comput (CEC). https://doi.org/10.1109/CEC.2018.8477876

42. Abdelaziz A, Salama AS, Riad AM, Mahmoud AN (2019) A machine learning model for predicting of chronic kidney disease based internet of things and cloud computing in smart cities. In: Hassanien A, Elhoseny M, Ahmed S, Singh A (eds) Security in smart cities: models, applications, and challenges lecture notes in intelligent transportation and infrastructure. Springer, Cham. https://doi.org/10.1007/978-3-030-01560-2_5

43. Almansour NA, Syed HF, Khayat NR et al (2019) Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study. Comput Biol Med 109:101–111. https://doi.org/10.1016/j.compbiomed.2019.04.017

44. Manonmani M, Balakrishnan S (2020) An ensemble feature selection method for prediction of CKD. (ICCCI). https://doi.org/10.1109/ICCCI48352.2020.9104137

45. Sharifi A, Alizadeh K (2020) A novel classification method based on multilayer perceptron-artificial neural network technique for diagnosis of chronic kidney disease. Ann Mil Health Sci Res 18:e101585. https://doi.org/10.5812/amh.101585

46. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20:273–297. https://doi.org/10.1007/BF00994018

47. Al-Hyari AY, Al-Taee AM, Al-Taee MA (2013) Clinical decision support system for diagnosis and management of chronic renal failure. IEEE Jordan Conf (AEECT). https://doi.org/10.1109/AEECT.2013.6716440

48. Chen Z, Zhang X, Zhang Z (2016) Clinical risk assessment of patients with chronic kidney disease by using clinical data and multivariate models. Int Urol Nephrol 48:2069–2075. https://doi.org/10.1007/s11255-016-1346-4

49. Charleonnan A, Fufaung T, Niyomwong T, Chokchueypattanakit W, Suwannawach S, Ninchawee N (2016) Predictive analytics for chronic kidney disease using machine learning techniques. 2016 (MITicon). https://doi.org/10.1109/MITICON.2016.8025242

50. Polat H, Danaei Mehr H, Cetin A (2017) Diagnosis of chronic kidney disease based on support vector machine by feature selection methods. J Med Syst 41:55. https://doi.org/10.1007/s10916-017-0703-x

51. Ravindra BV, Sriraam N, Geetha M (2018) Classification of non-chronic and chronic kidney disease using SVM neural networks. Int J Eng Technol 7:191–194. https://doi.org/10.14419/ijet.v7i1.3.10669

52. Peterson LE (2009) K-nearest neighbour. Scholarpedia 4:1883. https://doi.org/10.4249/scholarpedia.1883

53. Pa S, Po S (2015) Comparative study of chronic kidney disease prediction using KNN and SVM. IJERT 4:608–612. https://doi.org/10.17577/IJERTV4IS120622

54. Chetty N, Vaisla KS, Sudarsan SD (2015) Role of attributes selection in classification of chronic kidney disease patients. (ICCCS). https://doi.org/10.1109/CCCS.2015.7374193

55. Ani R, Sasi G, Sankar UR, Deepa OS (2016) Decision support system for diagnosis and prediction of chronic renal failure using random subspace classification. (ICACCI). https://doi.org/10.1109/ICACCI.2016.7732224

56. Wibawa MS, Maysanjaya IMD, Putra IMAW (2017) Boosted classifier and features selection for enhancing chronic kidney disease diagnose. Int Conf (CITSM). https://doi.org/10.1109/CITSM.2017.8089245

57. Pesce F, Diciolla M, Binetti G et al (2015) 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

58. Singh A, Nadkarni G, Gottesman O, Ellis SB, Bottinger EP, Guttag JV (2015) Incorporating temporal EHR data in predictive models for risk stratification of renal function deterioration. J Biomed Inform 53:220–228. https://doi.org/10.1016/j.jbi.2014.11.005

59. Perotte A, Ranganath R, Hirsch JS, Blei D, Elhadad N (2015) Risk prediction for chronic kidney disease progression using heterogeneous electronic health record data and time series analysis. J Am Med Inform Assoc 22:872–880. https://doi.org/10.1093/jamia/ocv024

60. Norouzi J, Yadollahpour A, Mirbagheri SA, Mazdeh MM, Hosseini SA (2016) Predicting renal failure progression in chronic kidney disease using integrated intelligent fuzzy expert system. Comput Math Methods Med 2016:6080814. https://doi.org/10.1155/2016/6080814

61. Yadollahpour A, Nourozi J, Mirbagheri SA, Simancas-Acevedo E, Trejo-Macotela FR (2018) Designing and implementing an ANFIS based medical decision support system to predict chronic kidney disease progression. Front Physiol 9:1753. https://doi.org/10.3389/fphys.2018.01753

62. Kanda E, Kanno Y, Katsukawa F (2019) Identifying progressive CKD from healthy population using bayesian network and artificial intelligence: a worksite-based cohort study. Sci Rep 9:5082. https://doi.org/10.1038/s41598-019-41663-7

63. Xiao J, Ding R, Xu X et al (2019) Comparison and development of machine learning tools in the prediction of chronic kidney disease progression. J Transl Med 17:119. https://doi.org/10.1186/s12967-019-1860-0

64. Weber C, Röschke L, Modersohn L et al (2020) optimized identification of advanced chronic kidney disease and absence of kidney disease by combining different electronic health data resources

and by applying machine learning strategies. J Clin Med 9:2955. https://doi.org/10.3390/jcm9092955

65. Song X, Waitman LR, Yu AS, Robbins DC, Hu Y, Liu M (2020) Longitudinal risk prediction of chronic kidney disease in diabetic patients using a temporal-enhanced gradient boosting machine: retrospective cohort study. JMIR Med Inform 8:e15510. https://doi.org/10.2196/15510

66. Schena FP, Anelli VW, Trotta J 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

67. Neugarten J, Golestaneh L (2019) Influence of sex on the progression of chronic kidney disease. Mayo Clin Proc 94:1339–1356. https://doi.org/10.1016/j.mayocp.2018.12.024

68. Levey AS, Eckardt KU, Tsukamoto Y et al (2005) Definition and classification of chronic kidney disease: a position statement from kidney disease: improving global outcomes (KDIGO). Kidney Int 67:2089–2100. https://doi.org/10.1111/j.1523-1755.2005.00365.x

69. Cravedi P, Remuzzi G (2013) Pathophysiology of proteinuria and its value as an outcome measure in chronic kidney disease. Br J Clin Pharm 76:516–523. https://doi.org/10.1111/bcp.12104

70. Ahmed S, Nutt CT, Eneanya ND et al (2021) Examining the potential impact of race multiplier utilization in estimated glomerular filtration rate calculation on African-American care outcomes. J Gen Intern Med 36:464–471. https://doi.org/10.1007/s11606-020-06280-5

71. Lee CS, Cha RH, Lim YH et al (2010) Ethnic coefficients for glomerular filtration rate estimation by the modification of diet in renal disease study equations in the korean population. J Korean Med Sci 25:1616–1625. https://doi.org/10.3346/jkms.2010.25.11.1616

72. Luyckx VA, Tonelli M, Stanifer JW (2018) The global burden of kidney disease and the sustainable development goals. Bull World Health Organ 96:414-422D. https://doi.org/10.2471/BLT.17.206441

73. Abeysekera RA, Healy HG, Wang Z, Cameron AL, Hoy WE (2021) Heterogeneity in patterns of progression of chronic kidney disease. Intern Med 51:220–222. https://doi.org/10.1111/imj.14770

74. Debray TPA, Vergouwe Y, Koffijberg H et al (2015) A new framework to enhance the interpretation of external validation studies of clinical prediction models. J Clin Epidemiol 68:279–289. https://doi.org/10.1016/j.jclinepi.2014.06.018