

Artificial Intelligence for the Diagnosis and Treatment of Diabetes Kidney Disease: a systematic review

Shams Mohammad Abrar

Email: shams.abrar2611@gmail.com

Affiliation: Independent researcher

ORCID iD: 0000-0002-3594-1113

Abstract:

Background: Diabetic nephropathy (DN) is a serious microvascular complication that affects 40% of diabetes patients. In the last decade, artificial intelligence (AI) has been widely used in both structured and unstructured clinical data to improve the treatment of patients/potential patients with DN.

Methods: This systematic review aims to cover all applications of AI in the clinical use of DN or related topics. Studies were searched in four open-access databases (Pubmed, IEEE Xplore, DBLP Computer Science Bibliography, and ACM digital library). Finally, the author manually searched the reference list of included studies in the study for additional relevant articles.

Results: Finally, a total of 24 original peers reviewed articles were included in this study. Through a manual data extraction, the summary of key information such as applied AI algorithm, main outcomes, performance evaluation etc. was taken. Then the included studies underwent a quality assessment criterion, assessing the reproducibility, generalizability etc. Most of the included studies revealed that the AI frameworks outperformed conventional statistical methods. A summary of the limitations, such as lack of data availability or external validation of the framework, in the included studies, was also included.

Conclusion: The rapid advancement of the AI framework and the exponential data generation in healthcare can be utilized and applied in clinical practices. The aid of AI can be instrumental in the treatment of DN.

Keywords: Diabetic kidney disease, diabetes, artificial intelligence, nephropathy, systematic review, machine learning, deep learning

1. Introduction

Diabetes Mellitus (DM), a devastating incurable health complication, is caused due to the incapability of beta cells (β cells) in the pancreas to produce adequate effective insulin for the body to utilize - leading to chronic high levels of blood glucose (1). Globally diabetes is considered a major public threat. According to the International Diabetes Federation (IDF), every eleventh person on the planet (463 million, 95% confidence interval: 361-601 million) has diabetes, and half of these patients are undiagnosed due to the disease's complex pathogenesis. Furthermore, diabetes accounts for ten per cent of global health expenditure (USD 760 billion) (2). Several factors have been linked to the onset of DM in patients. Inactive lifestyles, obesity, a lack of awareness, and other factors have all been linked to the recent global increase in diabetes (3).

DM, on the other hand, results in a variety of macro and microvascular complications as a result of hyperglycemia or hypoglycemia due to the lack of blood glucose (BG) control. Diabetic Nephropathy (DN)/ diabetic kidney disease (DKD) is one of the microvascular complications, which is the most common cause of end-stage renal disease in the world (4-7). DN is caused due to increased urine albumin excretion and/or impairment. It is characterized by gradual loss of kidney function from hyperfiltration at the early stage (8). The principal risk factors involving DKD are hyperglycemia, hypertension and genetic predisposition. To diagnose DN early on, it is recommended to screen microalbuminuria five years after diagnosis for type 1 diabetes (T1D) and yearly screening for type 2 diabetes (T2D) (9). However, the amount of DM patients being affected by kidney disease is rising rapidly. Prevalence of microalbuminuria is seen in about one-third of DM patients after fifteen years of disease duration, and almost half of them develop DN. Often DN occurs in patients of a certain age group or demographics. It is found that DKD is mostly common among elderly patients, very unlikely to develop among patients with less than one decade of diabetes duration (10). Meanwhile, the risk of developing DKD is higher among patients from developing countries compared to developed countries (26). The cost of bearing dialysis and other casualties is very high. As DN is more prevalent in developing countries, it almost becomes impossible for the patient's family to bear the expenses. Meanwhile, the mortality rate of DN patients is very high, and early diagnosis or adequate treatment is necessary. Around 10% of the death in T2D is due to kidney failure (11).

Despite growing concerns and preventative measures of DKD, the rate of chronic kidney disease (CKD) among DM patients remains unchanged as of 20 years ago. Even though DN might be controlled by the maintenance of blood glucose, blood pressure, blood cholesterol and widespread use of renin-angiotensin-aldosterone system inhibitors (9), it only works at the early stages. On the other hand, the progression of DKD of individual patients shows a large variation due to the complex heterogeneous nature of the disease (41). At later stages of DN, when CKD is well developed, it is harder to control (12). It is found that early diagnosis of microalbuminuria among DM patients and monitoring will postpone or prevent overt nephropathy (13,14). As a result of risk verification or early diagnosis of DN a large sum of the socioeconomic burden suppressed on the patients and their families can be avoided - where the further progression of DN could be stopped (12).

With the help of machine learning (ML), a branch of artificial intelligence (AI) diagnosis, prognosis and clinical management of DKD patients without direct human intervention is possible. These algorithms are capable of extracting clinically relevant information from medical data with great accuracy (8). Without being explicitly programmed, machine learning algorithms can produce non-linear correlations and patterns from raw data (15). Deep learning (DL), a branch of machine learning that is aided by increased computational capabilities, entails training a large number of

neural networks with many layers on large data sets. As a result, deep learning algorithms are useful for image classification in medical datasets (11). This systematic review covers the application of AI in DKD, which includes the early detection of DN to the management of DN through the application of AI.

Several systematic reviews have examined the use of AI in diabetes. Zhu et al. investigated the use of DL in DM, searching multiple databases and obtaining 40 studies (16). A total of 107 peer-reviewed research articles were included in a study by Jyotismita Chaki et al. on the use of AI in the detection and self-management of diabetic patients (17). Ioannis Kavakojis et al. discussed the use of machine learning and data mining in diabetes care (18). These studies, on the other hand, cover the broader application of AI in DM patients. They are not focused on any specific complications. Meanwhile, Katrine et al. conducted a systematic review on the use of DL algorithms in the screening for diabetic retinopathy (DR), evaluating the diagnostic performance of DL algorithms. Diabetic retinopathy is the most common diabetic microvascular complication. The study included a total of 20 articles (19). Miguel Tejedor et al. discussed the use of reinforcement learning, a type of machine learning in which the agent learns in an interactive environment through trial and error and feedback, in diabetic glucose control (20). A systematic review with meta-analysis was performed by Md Mohaimenul et al. about the detection of DR from retinal fundus photographs, where a total of 20 studies were included (21). However, there has yet to be a systematic review of AI's application in DN. This research aims to fill that void. This systematic review includes studies that use AI algorithms to address the early detection of DN, the prognosis of DN, DN management, and other related topics.

2. Overview of Big Data analytics and AI:

Researchers have been craving, since the mid-twentieth century, for a support system that can make clinical decisions while processing growing quantities of clinical data (22). As the medical records are often high dimensional, heterogeneous and sparse they are often not utilized properly in clinical practices (9). Applied biology is advancing to the era of 'big data' due to the remarkable advances in biotechnology. The principal reason behind the establishment of big data in medicine is high throughput sequencing due to the advancement of biotechnology (23,24). Myriads of large-scale and real-world data are generated from the likes of hospital records, patients' medical records, biomedical research etc. However, to utilize the data generated requires proper management and analysis (25). The proper application of big data in healthcare is executed by extracting patterns and trends thus creating models from the data. Consequently, big data analytics will yield more throughput insightful diagnoses, prognosis and most clinically effective treatments in a cheaper method, and it will improve healthcare marginally (26). An abundance of clinical data and information regarding diabetes patients is already available and it is increasing at an exponential rate (27).

Artificial Intelligence (AI) is a branch of computer science that aims to mimic the cognitive function of human beings. There are two main branches of AI: virtual and physical. The physical branch covers the aspects of robotics, which can assist in surgery in clinical practice. On the other hand, the virtual branch of AI includes the likes of ML. This informatics branch of AI is capable of assisting physicians in clinical diagnosis, treatment decisions etc. (28). As a result of the 'big data' era's emergence, AI brings forth permanent change in the healthcare system. In addition, AI applies to both structured and unstructured data. ML and Natural Language Processing (NLP) is applied in structured and unstructured data respectively (29). The raw data available from EHR usually consisted of structured data. Thus, ML algorithms are mostly used as predictive performance models in different studies. ML is classified into three different categories: supervised learning, unsupervised learning and reinforcement learning. In both supervised and

unsupervised learning, the system infers a function from labelled and unlabeled data respectively. ML workflow often involves data harmonization, preparation learning, model fitting and evaluation (30). Figure 1 demonstrates the learning process of both supervised and unsupervised learning (31). The theoretical construct of ML models gives them an edge over conventional statistical models. As they are efficient at analyzing large complex datasets, while identifying hidden associations ML is proved to be superior. Thus, over the recent decade, ML has evolved dramatically, and it is commonly used in clinical practices (16).

On the other hand, DL is a subfield of ML. The field of DL has advanced massively in the ability of machines to understand and data including the likes of images, language and speech. In contrast to ML, DL is a form of representation learning composed of numerous layers - gaining the ability to learn highly complex functions (33). DL algorithms such as Convolutional Neural Network (CNN) is designed to process data that exhibits natural spatial invariances. Therefore, DL models have achieved the accuracy of physicians in identifying cardiovascular risk from fundus images (34). Compared to the traditional Artificial Neural Networks (ANN), which are limited to three layers and trained to obtain supervised representation for specific tasks only (35), DL layers of features are not designed by anyone, but they are learned from data using a general-purpose learning procedure. The inputs of deep neural networks are processed in a non-linear manner to activate the nodes in hidden layers to learn the 'deep structures and representations. Next, the final representation is utilized in a supervised layer to fine-tune the network using backpropagation algorithms toward representations, which are optimized for an end-to-end task (36).

2. Methods:

A qualitative systematic review of the application of AI for the diagnosis, prognosis and management of DN was conducted. This systematic review was conducted by following the reporting checklist of Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA 2020) (37). Consequently, comprehensive literature was undertaken with extensive efforts to identify the articles. This review was conducted from 2020 to 2021 (considering all articles published from 1970 - 2021), and the databases searched include PubMed, DBLP Computer Science Bibliography, IEEE Xplore and ACM digital library. The four databases used are interfaces that do not require institutional subscriptions unlike other databases such as Ovid, Scopus and Web of Science. The purpose of using open-access databases is to enhance the reproducibility of the search results (16). The last search performed was on 12th July 2021.

2.1 Search Strategies:

The steps of searches were performed using related keywords combinations based on mesh terms. For this systematic review, related keywords of diabetes and nephropathy and AI were combined using Boolean operators (AND/OR). The search strategy is given in Table-1.

At the outset, the duplicate articles from different sources were excluded from the initial collection of the obtained articles. The records were screened a total of three times by the author, to avoid any biases in the systematic review. Additional potentially eligible articles were searched manually by the author by screening the reference list of included articles.

Table-1: The databases and the search strategy employed

	Search Strategies
Database	PubMed, ACM Digital Library, IEEE Xplore, DBLP Computer Science Bibliography
#1	diabet* OR DM OR T1D OR T2D OR Hyperglycemia OR “glucose intolerance”
#2	Nephropathy OR “kidney disease” OR kidney OR Renal
#3	“Artificial Intelligence” OR “machine learning” OR “deep learning”
Search	#1 AND #2 AND #3

2.2 Inclusion and Exclusion Criteria:

The most relevant and potential articles from the initial screening were kept for full-text review. To be included in the review, the articles had to fulfil certain criteria. The study must be in English and peer-reviewed. In addition, application of AI algorithms with performance evaluation is necessary. On the other hand, the study must be applicable for Diabetic patients with potential kidney disease or persisting kidney disease. Any study excluding the application of AI in diabetes and kidney disease were not included. If the AI algorithm only applies to DM but is not related to nephropathy is excluded - vice versa is also applicable. In addition, other types of study such as letters to editors, posters and short communications were excluded. As well, the included articles had described thoroughly the algorithms used and the data processing of datasets. Traditional methods used only in DN are excluded. The author, himself, had thoroughly examined the selected articles by screening the article titles and abstracts to identify the eligible articles. If the article is not relevant to the topic at all then it is excluded at the initial phase which is conducted by screening the title and abstract.

2.3 Data extraction:

From the final collection of the studies, key information was extracted to assess the application of AI algorithms. The data extraction was divided into 10 different categories. The categories and their information are given below:

- 1) Study information: the first author’s name, year, country.
- 2) Cases: summary of the specific application in different cases i.e., detection of DN, management of DN etc. If applicable the type of diabetes of the cohort was mentioned as in T1D or T2D.
- 3) Data sources: The name and description of the dataset’s source is obtained and mentioned. It can be an EHR dataset from a regional hospital to national surveys.
- 4) Research type: The type of study conducted, whether it was an observational study or experimental study, is mentioned in this category.
- 5) Baseline characteristics: The clinical variables used to form the ML models were mentioned here. For DL algorithms the type of image taken along with basic characteristics is mentioned.
- 6) Cohort size/sample size: The population included in the cohort studies is mentioned. On the other hand, the sample size for other studies is mentioned here.
- 7) AI algorithm: In this category, all of the ML or DL algorithms used are mentioned.
- 8) Development process: The size of the training and testing dataset is mentioned here.

9) Performance assessment: The evaluation metrics of AI algorithms are included here. Some of the performance assessment techniques include accuracy test, sensitivity, specificity, Cohen's k, the area under the curve (AUC) etc.

10) Limitation: The limitations of the included studies, mostly mentioned by the authors themselves, are included in this category.

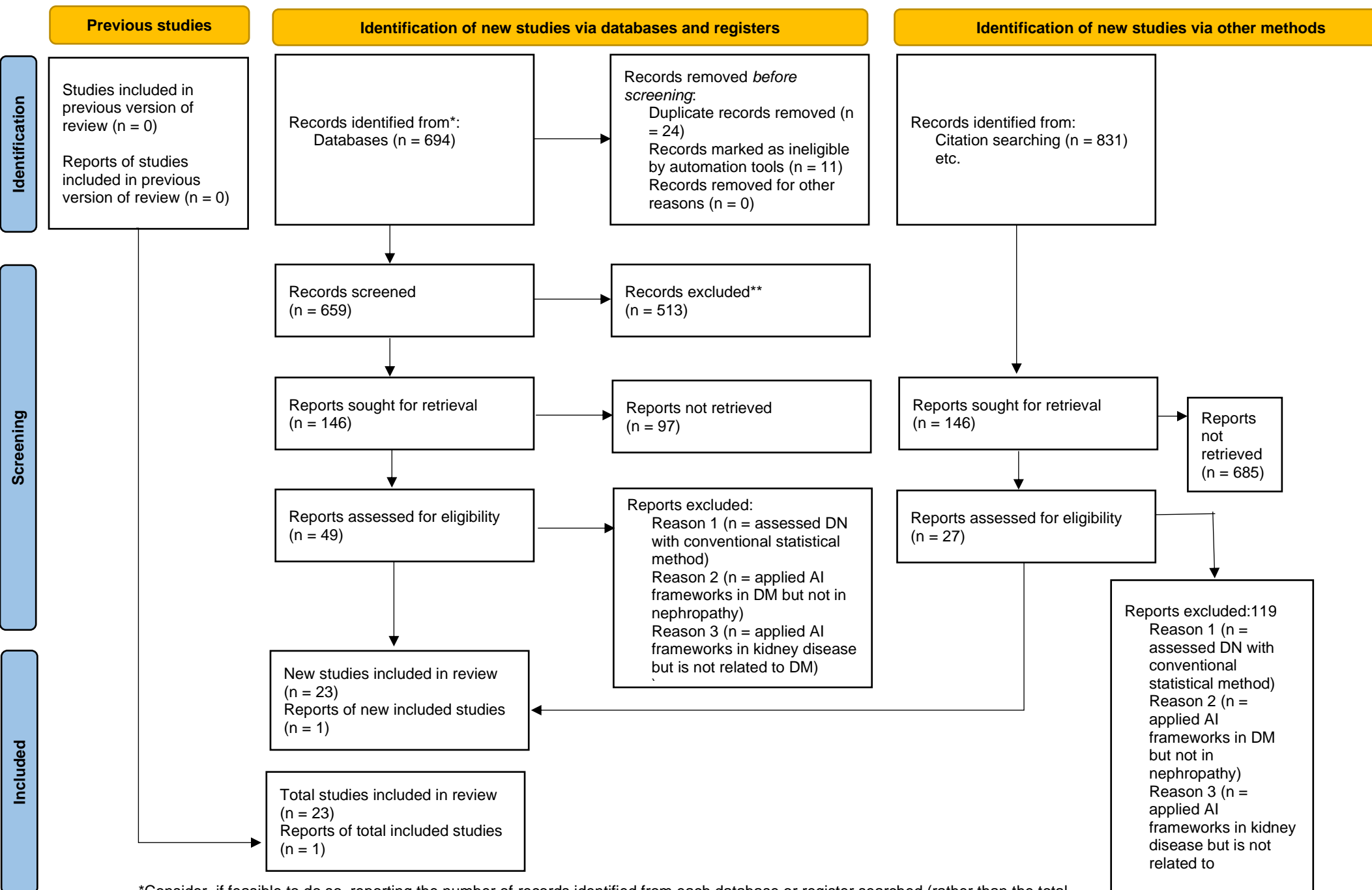
2.4. Quality Assessment:

Studies were assessed from the following categories: limits in current non-machine learning tools, feature engineer methods employed in the AI algorithm before training the dataset, the hyperparameters used in the study, valid methods to overcome overfitting and the external validation of the data. The quality assessment table of the studies consisted of binary answers of yes or no.

3. Results:

After obtaining the results from the four databases, the duplicate articles were excluded. Then a manual inspection by the author was conducted to evaluate the remaining articles based on inclusion criteria. From the four databases a total of 694 (IEEE = 91, PubMed = 202, ACM Digital Library = 393, DBLP = 8) studies were obtained, and 11 of them were marked as ineligible by endnote. In addition, a total of 24 duplicate studies were found. The remaining 659 studies were manually screened by the author, and they were evaluated based on the inclusion and exclusion criteria. Consequently, a total of 146 studies were retrieved for eligibility assessment via full-text inspection - all of these studies either included diabetes patients and/or kidney patients and/or AI algorithms. In the end, a total of 23 articles met the inclusion criteria. The references cited in those 23 studies were screened by the author himself. Among the 831 references listed only one of the studies met the complete inclusion criteria. The details of the selected articles are presented in Table-3.

Table-2: PRISMA flow diagram for updated systematic reviews.



*Consider, if feasible to do so, reporting the number of records identified from each database or register searched (rather than the total number across all databases/register).

**If automation tools were used, indicate how many records were excluded by a human and how many were excluded by automation tools.

First Author, Year of Publication, Country	Case	Data Sources	Research type	Clinical variables/ baseline characteristics/ image type	Cohort size/sample size	AI algorithm	Development process	Performance Assessment	Limitations
Lili Chan et al. ⁽³⁹⁾ , 2020 USA	Predicting the progression of DKD	BioMe Biobank and PMBB dataset (using novel biomarkers and EHR data)	Retrospective cohort study	Age, sex, race, BMI, BP, hypertension, eGFR, CAD, HF, uACR, HbA1c, follow up (months), Plasma biomarkers, medication, smoking status	1146 patients with DKD	KidenyIntelX RF model LR	The derivation set = 60% The validation set = 40% From the derivation set, training set = 70%, the testing set = 30%.	AUC = 0.77 AUC = 0.75	The uACR of 38% of the cohort was unavailable. Possibility of bias in ML model due to no protocolized follow-up, resulting in lack of kidney biopsies and missing data. The models were not tested in a multiethnic cohort.
Hee-Sung Ahn et al. ⁽⁴⁰⁾ , 2020 Korea	Prognosis to DKD for T2D patients	An EHR dataset from the Human Swissport proteome database using the SequestHT search engine	Retrospective cohort study	Age at diagnosis of DKD, sex, BMI, BP, LDL, HDL, triglycerides, eGFR (one year), HbA1c, ACR, NAPCR, PCR, RAS inhibitor, Anti-hypertensive agent, Lipid-lowering agent, duration of follow-up, the status of DR	Urine samples from 54 patients with or without renal dysfunction, where 1296 protein samples were identified	RF SVM with linear kernel	Not mentioned	AUC = 1 AUC = 0.935	Homogenous population small sample size. Further validation in multiethnic cohorts is needed. DKD was clinically diagnosed in the absence of renal biopsies. The organ deriving urinary protein signature is unclear.

Sunil Belur Nagaraja et al. (41). 2020 Netherlands	Early detection of ESRD of DN patients.	Data were collected from three different clinical trials of T2D patients with nephropathy (RENAAL, IDNT, ALTITUDE).	Clinical trial	Age, sex, BMI, smoking status, BP, serum creatinine, serum potassium, haemoglobin, HbA1c, serum albumin, serum calcium, phosphorus, serum uric acid, HDL, LD, UACR, patient records	11789 patients from the multiethnic cohort. (RENAAL = 1513, IDNT = 1715, ALTITUDE = 8561)	Feed forward neural network	K-nearest algorithm to impute missing variables. Fivefold cross- validation where training set = 80% and validation set = 20%	RENAAL AUC = 0.82 IDNT AUC = 0.81 ALTITUDE AUC = 0.84 AUC (0.78, 0.78, 0.81)	The variability in medication adherence of patients is avoided. Performance can be enhanced by adding additional molecular and cellular biomarkers while increasing the sample size of the training set. AUC (0.78, 0.78, 0.81)
						SVM with Gaussian Kernel			
						RF		RENAAL AUC = 0.8 IDNT AUC = 0.79 ALTITUDE AUC = 0.82	
						LR		RENAAL AUC = 0.78 IDNT AUC = 0.78 ALTITUDE AUC = 0.81	
Shinji Kitamura et al. (42) 2020 Japan	Diagnosis of DN from immunofluorescent images	An EHR dataset consisting of kidney samples of renal biopsies from the patients administered in a regional Hospital (Okayama	Retrospective cohort study	Images of renal immunofluorescent of patients' renal biopsy. Each patient's sample contained IgG, IgA, IgM, C3, C1q and fibrinogen immunofluorescent type.	885 renal immunofluorescent images (image resolution = 256 * 256)	39 programs out of 419 AI programs trained on training set	Training images and testing images ratio = 8:2	Accuracy rate = 83.28 ± 11.64% Precision rate = 80.56 ± 21.83% Recall rate = 79.87 ± 15.65% AUC = 0.93 R ² = 0.4586 P < 0.0001	The model was only applied to immunofluorescent images from a small dataset. The algorithm was not validated externally. The sample only consisted of images from a single hospital.

		University Hospital).				6 programs among the 39 programs		Accuracy, precision, recall = 100% AUC = 1 $R^2 = 1$ $P < 0.001$	
Jinxia Chen et al. ⁽⁴³⁾ 2015 China	Prediction of glomerular filtration rate of T2D patients	An EHR dataset from a regional Hospital (Third Affiliated Hospital of Sun Yatsen University, China)	Cross-sectional study	Age, sex, diabetes duration, weight, height, BSA, BMI, serum creatinine median, sGFR, UACR, HbA1c,	519 T2D patients.	ANN1 ANN2 ANN3 ANN4 ANN5 ANN6 ANN7 ANN8	Developmental dataset, n = 276 Internal validation dataset, n = 138 External validation dataset = 105	Accuracy (%) = 87.6 Accuracy (%) = 83.8 Accuracy (%) = 88.6 Accuracy (%) = 84.8 Accuracy (%) = 85.7 Accuracy (%) = 88.6 Accuracy (%) = 87.6 Accuracy (%) = 87.6	The dataset employed consists of a very small number of samples and it was restricted to a single centre. The model should be validated by employing it on a multiethnic cohort.
Xiaoyuan Bao et al. ⁽⁴⁴⁾ China 2019	Predicting DN among DM patients through a clinical text document	An EHR dataset extracted from a grade 3A tertiary hospital	Retrospective longitudinal study	The patients' chronic disease course, family history of chronic disease, personal smoking, drinking history etc. from the EHR records.	3422 admission notes from EHR.	LR SVM RF	8 target vectors were extracted from the dataset (patients' diabetes, hypertension and heart disease course with personal drinking/smoking; patients' parents' diabetes, hypertension and heart disease course with	F1 = (0.94, 0.93, 0.91, 0.85, 0.84, 0.72, 0.35, 0.91, 0.87) F1 = (0.91, 0.9, 0.9, 0.83, 0.83, 0.73, 0.93, 0.92) F1 = (0.89, 0.94, 0.89, 0.85, 0.86, 0.7, 0.92, 0.9	The dataset contained far too few samples. The information extracted was not a big part of the clinical data. The corpus is not externally validated.

Author	Study Title	Dataset	Study Design	Features	Sample Size	Model	Additional Info	Performance	Notes	
Hayashi et al. (45) 2019 Japan	Early detection of DKD	National Health and Nutrition Survey (NHANES) (1994-2004) dataset	Prospective cross-sectional study	Age, sex, race, BP, waist circumference, BMI, total cholesterol, uACR, tobacco use, alcohol use, Exercise to lose weight, triglycerides, LDL, direct HDL, fasting plasma glucose, insulin, total bilirubin	942 DM patients 524 pre-DM patients	DT MLP NB Adaboost KNN	personal drinking/smoking)	F1 = (0.95, 0.97, 0.92, 0.84, 0.86, 0.72, 0.91, 0.88) F1 = (0.95, 0.97, 0.92, 0.8, 9.8, 0.62, 0.92, 0.92) F1 = (0.81, 0.8, 0.64, 0.65, 0.6, 0.52, 0.92, 0.92) F1 = (0.96, 0.97, 0.93, 0.83, 0.85, 0.69, 0.91, 0.89) F1 = (0.8, 0.77, 0.59, 0.51, 0.71, 0.35, 0.91, 0.92)	Accuracy = 77.56 ± 2.19 AUC-ROC = 75%	The dataset consists only of self-reported data, and it was relatively small. Only T2D patients were in the study.

Baek Hwan Cho et al. ⁽⁴⁶⁾ 2007 Korea	Predicting the onset of DN months before diagnosis.	An EHR dataset from a regional clinic (Samsung Medical Center) of DM outpatient for 10 years.	Cross-sectional study	Age, diabetes duration, white blood cell count, haemoglobin, platelet count, BMI, BP, HDL, LDL, HbA1c, sex, microalbumin, serum triglycerides level, serum cholesterol level, serum aspartate aminotransferase level, serum alkaline phosphate level, blood urea nitrogen, creatinine, uric acid, Na+, K+	Medical data of 292 patients with DM initially from 4321 samples.	Ridge LR	The quantity of training and testing set is not mentioned.	<p>AUC = 0,57 Accuracy = 0.77 HMSS = 0.56 BER = 0.38 Sensitivity = 0.42 Specificity = 0.82 C+ = n/a</p> <p>AUC = 0.68 Accuracy = 0.89 HMSS = 0 BER = 0.5 Sensitivity = 0 Specificity = 1 C+ 1</p> <p>AUC = 0.64 Accuracy = 0.88 HMSS = 0 BER = 0.5 Sensitivity = 0 Specificity = 1</p>	The medication information of the patients in the dataset is missing, the dataset is relatively small. The model scores are inadequate to supplement decision making by physicians.
Brandon Ginley et al. ⁽⁴⁷⁾ 2019 USA	Digital pipeline to classify renal biopsies from patients with DN	The whole-slide image data were obtained from Kidney and Translational Research Core (KTRC) and Vanderbilt University Medical Center (VUMC)	Cross-sectional pilot study	Images of renal tissue samples.	Renal tissue sections image of n = 54 human beings Renal tissue sections image of n = 25 mice	RNN	Out of 54 samples, 6 samples were control tissues. A single batch of training data consisted of 256 feature sequences/	Cohen's k between RNN and ground truth renal pathologist 0.55 (0.50 to 0.6) Glomerular boundaries detection accuracy = 0.93 ± 0.04; Glomerular nuclei	The precision could be improved by continuous risk prediction models. The training set for the model is relatively small.

								detection sensitivity and specificity (0.94, 0.93); Glomerular structural component detection sensitivity and specificity (0.95, 0.99)	
Jialing Huang et al. ⁽⁴⁸⁾ 2020 Germany	Early detection of CKD among pre-diabetes and T2D patients.	A survey conducted in South Germany (KORA)	Longitudinal cohort study	Age, sex, BMI, HbA1c, Fasting glucose, 2-h glucose, BP, triglyceride, HDL, LDL, total cholesterol, Baseline eGFR, Follow-up eGFR, Smoking status, Medication usage	2142 individuals were 385 hyperglycemic patients and 1453 participants with normal glucose tolerance	SVM	Data of DM patients were partitioned into 10-non overlapping subsets, where 9 of them are training data and the remaining one is the testing data	Reference Predictors AUC = 0.8 (95% CI) Developed set AUC = 0.825 (95% CI)	The external validation of the study is missing. There is a missing replication of 10 patients in the study.
						RF		Reference Predictors AUC = 0.789 (95% CI) Developed set AUC = 0.818 (95% CI)	
						Adaboost		Reference Predictors AUC = 0.798 (95% CI) Developed set AUC = 0.814 (95% CI)	
Yong-Mi Kim et al. ⁽⁴⁹⁾ USA 2017	Comparing and detecting markers to predict ESRD among African	An EMR dataset from the Cerner database.	Comparative research	All the complications found in DM patients (eg. Low	4623 DM outpatients. Here, 3550 Caucasian patients	Apriori	The exact quantity of training and testing set is not mentioned	Support > 0.00001 Confidence = 1.00 Lift > 35	The study was focused on the Caucasian group with eGFR ≥60 mL/min.

	Americans and Caucasians.			hematocrit, low lymphocytes etc.)	and 1073 African American patients.			744 rules found for African American patients	The data was not externally validated.
Ji Eun Kim et al. ⁽⁵⁰⁾ Korea 2020	Detection of posttransplant DM in kidney transplant recipients based on quantification of visceral fat (VF) volumes	An EHR dataset (2003-2017) derived from a national hospital (Seoul National University Hospital)	Retrospective study	Age, sex, BMI, percent of the deceased donor, type of pre-transplant dialysis, Cause of kidney failure, induction agent, a calcineurin inhibitor, total cholesterol, LDL, HDL, uric acid	718 nondiabetic kidney transplant recipients (39268 labelled whole-body CT images)	DNN	The exact quantity of training and testing set is not mentioned	Average segmentation accuracy of VF = 92.4–98.9% Average segmentation accuracy of subcutaneous fat = 94.1–99.7%	One of the main factors of measuring fat volume, waist circumference, was not taken into account as a clinical variable. The model was not tested on a multiethnic population, and further external validation is needed before applying it in clinical factors.
Xing Song et al. ⁽⁵¹⁾ USA 2020	Predicting CKD in DM patients	An EHR dataset from the data repository of University of Kansas Medical Center's (regional clinic) Healthcare Enterprise Repository for Ontological Narration (HERON).	Retrospective cohort study	Demographics, diagnoses, history, laboratory tests, medications, procedures, alerts, allergy, visit details, orders, Vizient	14039 DM patients	Temporal-Enhanced GBM	Training set = 80% Validation set = 20%	AUROC of predicting CKD risk with the automatic accumulation of new data in 2, 3, 4 years since the onset of diabetes (AUC = 0.83, 0.78, 0.82)	The model is not externally validated, and it is not robust against population drift.

Xing Song et al. ⁽⁵²⁾ USA 2018	Detecting robust set of discriminant factors for DKD	An EHR dataset (2007-2017) from the data repository of University of Kansas Medical Center's (regional clinic) Healthcare Enterprise Repository for Ontological Narration (HERON).	Retrospective cohort study	Different sources of a diagnosis, different aspects of medication, different types of encounters where procedures were performed, different states of an alert	15645 T2D patients	GBM with weighted mean feature ensemble technique	Training set = 44.4% Internal validation set = 45.5% Temporal validation = 10.1%	Internal and external validation (AUC = 0.82, 0.71)	Lack of robustness persists against population shift. The differences in the feature selection algorithm are not taken into account.
Andreas Heinzl et al. ⁽⁵³⁾ Austria	Predicting eGFR declination in T2D patients	A multinational cohort study from PROVALID	Prospective cohort study	Age, sex, serum cholesterol, uACR, HbA1c, mean arterial pressure, BMI, eGFR at the baseline level.	2560 patients with 17 biomarkers measured at the baseline.	Logistic Regression	The exact quantity of training and testing set is not mentioned	Adjusted R ² = 62.5% 12 out of 17 biomarkers were associated with 34% of eGFR decline.	A possibility of over-optimistic results in the selection of patients from the dataset. The follow-up result of patients is assumed to be relatively short.
Masaki Makino et al. ⁽⁵⁴⁾ Japan 2019	Prediction of DKD progression within 6 months	An EMR dataset from a regional hospital's (Fujita Health University Hospital) EHR repository	A retrospective longitudinal cohort study	. structured data (lab tests, medication, anamnesis etc.), text data (diagnosis, treatments records etc.), longitudinal data (laboratory test etc.)	64059 T2D patients	Convolutional autoencoder	The exact quantity of training and testing set is not mentioned	Accuracy = 71% AUC = 0.743	The extracted data were from different sources; thus, it was not unified. The samples are from a single centre, and the model is not validated in other samples from different hospitals. The progression of

									DKD for six months and medication's relationship is unavailable.
Violeta Rodriguez-Romero et al. ⁽⁵⁵⁾ USA 2019	Prediction of DKD in T2D patients	Dataset obtained from Biological Specimen and Data Repository Information Coordinating Center of the National Heart, Lung and Blood Institute (NHLBI)	Retrospective longitudinal study	Age, sex, Race, history of CV event, GHb, eGFR, UCr, SCr, Ualb, CPK, HDL, K, LDL cholesterol, uACR, BP, FPG, triglycerides	10251 T2D patients	One Rule (1R) J48 DT RF Simple Logistic Sequential Minimal Optimization Naïve Bayes	Training set = 66% Testing and validation set = 34%	Sensitivity > 0.78 Accuracy > 0.77 Sensitivity > 0.72 Accuracy > 0.74 Sensitivity > 0.76 Accuracy > 0.92 Sensitivity > 0.74 Accuracy > 0.77 Sensitivity > 0.75 Accuracy > 0.73 Sensitivity > 0.67 Accuracy > 0.73	The data of patients having developed DN was not considered in the analysis. The novel biomarker variables that are related to the DN pathogenesis is not applied. The model is not externally validated.
Sitong Lin et al. ⁽⁵⁶⁾ China 2021	Early renal damage detection in DM patients	The sample size is obtained from EI database	Retrospective cohort study	Age, gender, FPG, 2h PPG, HbA1c variables	100 DM patients	LeNet DetectionNet	The exact quantity of training and testing set is not mentioned	The evaluation is not mentioned	To verify and analyze the renal biopsy of DM patients a large sample is needed.
Kang Zhang et al. ⁽⁵⁷⁾ China 2021	Detecting CKD and T2DM through retinal fundus images	Retinal fundus image dataset from the China Consortium of Fundus	Retrospective longitudinal study	Retinal fundus images, and metadata (age, sex, height, weight, BP	57672 patients where 115344 retinal fundus photographs were obtained	RF using clinical metadata	Training, tuning and internal testing ratio 7:1:2	Early CKD detection on internal test set and external test set (AUC = 0.805, 0.8)	The algorithm should be tested in a multiethnic cohort. There is a possibility of inaccurate

								Image Investigation		
									estimation of eGFR among patients with rapid deterioration of renal function.	
								DL model with fundus image	Identifying T2D on internal test set and external test set (AUC = 0.828, 0.796)	
									Early CKD detection on internal test set and external test set (AUC = 0.839, 0.829)	
									Identifying T2D on internal test set and external test set (AUC = 0.923, 0.854)	
								Combined DL and RF model	Early CKD detection on internal test set and external test set (AUC = 0.864, 0.849)	
									Identifying T2D on internal test set and external test set (AUC = 0.929, 0.871)	
Lijun Zhao et al. ⁽⁵⁸⁾ China 2020	Predicting ESRD through the severity of DR during biopsy	An EHR dataset from a regional hospital (West China)	Retrospective longitudinal study	Colour photograph of each retina of the patients using 45° using a fundus camera	91 T2D patients with renal biopsy confirmed DN	RetinalNET (Lesion-aware DL system)	The exact quantity of training and testing set is not mentioned		HR = 2.18 P = 0.04	The casual relationship between DN and DR were not inferred. There

		Hospital of Sichuan University)							was the existence of selection bias. The renal outcomes were infrequent.
Hongxia Xu et al. ⁽⁵⁹⁾ China 2020	Predictive model for the early diagnosis of DKD	An EMR dataset from a regional hospital (Beijing Pinggu Hospital)	Retrospective longitudinal study	Age, haemoglobin, hematocrit, average red blood cell volume, mean haemoglobin content, mean haemoglobin concentration, red blood cell distribution width, lymphocyte count, monocyte count, eosinophilic fine count, basophilic granule count, percentage of lymphocyte, monocyte percentage	1177 DM patients-914 DM patients with non-concurrent DKD and 263 DM patients with DKD	RF LR KNN DT	Training set = 685 patients with non-concurrent DKD and 197 DKD patients Testing set = 229 patients with non-concurrent DKD and 66 DKD	Precision = 0.9 Recall = 0.9 F1-score = 0.89 Precision = 0.88 Recall = 0.88 F1-score = 0.88 Precision = 0.88 Recall = 0.88 F1-score = 0.88 Precision = 0.89 Recall = 0.89 F1-score = 0.89	The AI algorithm was not externally validated and the sample size is relatively small.
Jivan Parab et al. ⁽⁶⁰⁾ India 2021	Monitoring blood urea and Glucose for DM patients with CKD	Laboratory samples and analytical grade compounds from Sigma Aldrich Ltd	Experimental study	Laboratory samples resembling blood mixed with glucose, ascorbate, urea, lactate, and alanine in the proportion found in the blood in an aqueous solution	Spectra of 57 samples	PLSR BP-ANN	Training set = 53 samples Testing set = 4 samples	Prediction of blood urea and glucose (R2 = 0.93, 0.97) (accuracy = 94.2%, 90.14%) Prediction of blood urea and glucose (R2 = 0.96, 0.99) (Accuracy =	The blood samples used in both training and testing data is made from a laboratory. The model should be applied in human tissue to properly estimate urea and glucose.

Jinghe Zhang et al. ⁽⁶¹⁾ USA 2020	Early detection of CKD in DM patients	An EHR dataset from the University of Virginia Health System	Prospective longitudinal data	Sparse longitudinal representation of patients' medical records	395 DM patients who are diagnosed with CKD at least 18 months after DM diagnosis.	Bagged SLR	Exact figures not mentioned	95.96%, 98.65%) Sensitivity = 0.829 ± 0.023 Specificity = 0.865 ± 0.020 AUC = 0.842 ± 0.020 F2 score = 0.818 ± 0.028	The sample-set employed is relatively small and the model is not externally validated
						WB-SLR		Sensitivity = 0.835 ± 0.025 Specificity = 0.852 ± 0.012 AUC = 0.891 ± 0.018 F2 score = 0.820 ± 0.027 AUC = 0.77	
Kinsuk Chauhan et al. ⁽⁶²⁾ 2020 USA	Predicting kidney outcomes of T2D patients	An EHR dataset and plasma specimens from BioMe Biobank	Retrospective Longitudinal study	Age, sex, race, BMI, hypertension, CAD, heart failure, BP, follow-up, eGFR, uACR, Hb1Ac, ARB, plasma biomarker concentration	871 T2D patients	KidneyIntelX	Training set = 80% Testing set = 20%	Positive predictive values = 62% Negative predictive values = 92%	Further validation of the model is necessary in geographically diverse population. Overfitting of the model is possible.

Abb: PMBB (Penn Medicine Biobank), CAD (coronary artery disease), HF (Heart Failure), BP (blood pressure), BSA (body surface area), uACR (urine albumin creatinine ratio), sGFR (standard glomerular filtration rate), FPG (fasting plasma glucose), GHb (glycosylated haemoglobin), CV (cardiovascular event), CPK (creatinine phosphokinase), K (potassium), UA1b (Urinary albumin), UCr (urinary creatinine), SCr (serum creatinine), True Positive Rate (TPR), receiver operating characteristic curve (ROC), SLR (Sparse Longitudinal Representations), WB (weighted bagging), AUC-ROC (area under the receiver operating characteristics curve), GBM (Gradient Boosting Machine), P (null hypothesis test), R^2 (coefficient of determination), BMI (body mass index), HbA1c (Hemoglobin A1c),

Table-3: The summary of findings in the included articles

The scheme of this study is summarized into four different categories based on the AI framework in use. The categories include 1) predictive AI model for the early detection of DN (40, 44, 45, 46, 48, 51, 54, 55, 59, 61), 2) AI Framework for the diagnosis of DN (42, 50, 57), 3) predictive model for the progression of DKD (39, 41, 49, 58, 62), 4) the management of DN for existing patients OR early management indication to DM patients without DKD with the help of AI (43, 47, 52, 53, 56, 60). The studies included in this systematic review are from various parts of the world. Seven of the included studies are from the USA (29%), six of them are from China (25%). In addition, three studies are from Japan (12.5%) and Korea (12.5%). Although the number of observational studies included was greater, the comprehensive review consisted of both observational study and experimental study.

A total of 8 studies (39, 40, 44, 45, 48, 49, 53, 59) included a cohort population/ sample size of 1000+ patients/samples to develop the AI model. Meanwhile, 7 studies (41, 50, 51, 52, 54, 55, 57) included a cohort size/ sample size over 10,000 patients/sample size. In addition, a large sum of 14 studies (39, 40, 42, 44, 48, 49, 51, 52, 53, 55, 56, 57, 59, 62) used ML to build the predictive model. On the other hand, 5 studies (47, 50, 54, 58, 60) used DL algorithms as the AI method in their study. Finally, 5 other studies used both ML and DL as separate baselines or in an integrated system (41, 43, 45, 46, 61). It is worth noting that a large number of the studies used EHR datasets from a regional hospital/clinic. A total of nine studies (42, 43, 46, 51, 52, 54, 58, 59, 61) used an EHR dataset from a regional hospital.

Different AI algorithms such as RF (Random Forest), LR (Logistic Regression), SVM (Support Vector Machine), feed-forward network, ANN (Artificial Neural Network), KNN (k-nearest neighbours), MLP (Multilayer Perceptron), RNN (Recurrent Neural Network), Naive Bayes (NB), Adaboost, DT (Decision Tree), Ridge LR, Apriori, DNN (Deep Neural Network), GBM (Gradient Boosting Machine), convolutional autoencoder, One rule, Simple Logistic, Sequential Minimal Optimizer, LeNet, DetectionNet, Lesion aware DL system, PLSR (Partial Least Square Regression), Backpropagation Artificial Neural Network (BP- ANN). Above all, RF is the most used model for predictive models being used 7 times.

As most of the included studies are predictive models, they require a certain criterion to be met. The data should be pre-processed with the necessary imputation, dimensionality reduction, transformation etc. The training model used in the dataset is used to find the patterns from the feature vectors with the output vectors with the help of the AI algorithm(s). A testing dataset is also necessary from the same source as the training that is unseen to the algorithm. Therefore, the accuracy of the AI algorithm can be assessed. However, a lot of the studies did not mention the exact quantity of the development process of the dataset i.e the training and testing sample size (40, 45, 46, 49, 50, 53, 54, 56, 58, 61). As a result, if the performance of the AI model is overfitting or not cannot be determined.

Table 4: Quality Assessment of the studies:

	Limits in current non-AI approach	Reproducibility		External data validation
		Feature Engineering	Hyperparameters	
Lili Chan et al. ⁽³⁹⁾	Yes	No	Yes	No
Hee-Sung Ahn ⁽⁴⁰⁾	Yes	Yes	Yes	Yes
Sunil Belur Nagaraja ⁽⁴¹⁾	Yes	Yes	Yes	No
Jinxia Chen ⁽⁴³⁾	Yes	Yes	Yes	Yes
Xiaoyuan Bao ⁽⁴⁴⁾	Yes	N/A	Yes	No
Baek Hwan Cho ⁽⁴⁶⁾	Yes	Yes	Yes	No
Shinji Kitamura ⁽⁴²⁾	Yes	Yes	Yes	Yes
Brandon Ginley ⁽⁴⁷⁾	Yes	Yes	Yes	Yes
Jialing Huang ⁽⁴⁸⁾	Yes	Yes	Yes	No
Lijun Zhao ⁽⁵⁸⁾	Yes	No	No	Yes
Sitong Lin ⁽⁵⁶⁾	N/A	Yes	Yes	N/A
Jivan Parab ⁽⁶⁰⁾	Yes	No	N/A	No
Hongxia Xu ⁽⁵⁹⁾	No	Yes	Yes	No
Kang Zhang ⁽⁵⁷⁾	Yes	Yes	Yes	Yes
Jinghe Zhang ⁽⁶¹⁾	Yes	Yes	Yes	No
Violeta Rodriguez-Romero ⁽⁵⁵⁾	Yes	Yes	No	No
Xing Song ⁽⁵¹⁾	Yes	Yes	Yes	No
Xing Song ⁽⁵²⁾	Yes	Yes	Yes	Yes
Yoichi Hayashi ⁽⁴⁵⁾	No	Yes	Yes	No
Masaki Makino ⁽⁵⁴⁾	Yes	Yes	No	No
Ji Eun Kim ⁽⁵⁰⁾	Yes	Yes	No	No
Andreas Heinzl ⁽⁵³⁾	Yes	Yes	No	Yes
Yong-Mi Kim ⁽⁴⁹⁾	Yes	No	Yes	No
Kinsuk Chauhan ⁽⁶²⁾	Yes	Yes	Yes	No

In the quality assessment (Table 4) the shortcomings OR strengths of the studies were assessed. To ensure the included study adds something new to clinical practices and fulfils a void that cannot be assessed by conventional statistical methods, the “limits in current non-AI approach” column is present in Table-4. If the AI model even outperformed the conventional statistical method the cell is filled with a ‘yes’. Except for two studies (59, 45), all 22 other included studies either outperformed conventional methods or accomplished the result which is not attainable through non-AI frameworks. This proves that AI algorithms are superior to conventional methods, and they might be suitable in clinical practices to aid physicians. The feature generation of the AI algorithm(s) before undergoing training is also mentioned. Feature engineering is particularly important as they are essential to select and transform raw data from clinical variables to essential features. As well as, handling missing data or imbalanced data is accounted for in the studies. From the included studies, a total of studies 19 conducted feature engineering in the raw data to make the algorithm more feasible. Four studies (39, 49, 58, 60) did not use feature engineering. Also, before the training of AI algorithms, some studies chose to tune hyperparameters to yield better outcomes. Among 24 studies only four (50, 54, 55, 58) studies did not include hyperparameters in the study. Hyperparameter tuning is necessary for ML models as the optimal combination reduces the predefined lost function as low as possible, which leads to a better performance of the model.

The model's higher predictive accuracy makes it more reliable for clinical use. In addition to the need for high accuracy in the model, validation of the model in different settings outside of the research's framework is critical. The vast majority of models have not been externally validated. External validation is required to determine the model's reproducibility and generalizability to unique patients. Aside from that, it is critical to ensure that the model is tested on a multiethnic cohort before implementing the AI framework globally. Only seven (40, 43, 47, 52, 53, 57, 58) of the 24 included articles had the predictive AI model externally validated (29 percent).

4. Discussions:

This is the first systematic review that reviews current applications of AI in DKD. The heterogeneity in the types of AI algorithms used are shown in the results - SVM, DT, RF, ANN, Naive Bayes etc. The myriads of different feature vectors used in the predictive performances were observed. Studies in this review were performed in different parts of the globe. Some of the studies included multiethnic cohorts. In most of these studies, it is observed that there persist limitations in conventional statistical models. However, ML and DL models proved to be far superior in many instances. AI methods applied in different aspects of DN had met the substantial needs, which conventional methods lack, in the proposed clinical applications. The only aspect of shortcoming in AI algorithms is the generalizability, which can be met with the growing number of EHR datasets in clinics.

It should be pointed out that most selected papers are published within the last four years, which indicates the application of AI in DN is an emerging technology. Despite being not so perfect, AI still prevails a large opportunity to improve the current uses. The predictive models can be improved a lot if expert knowledge is integrated with the data-driven approach. The physicians with adequate knowledge in DN can craft safety constraints as a guide during the training process (16). The performance of AI frameworks in clinical practices of DKD can adequately be assessed from prospective studies (63). For future work, prospective studies should be conducted in studies focused on category four. Therefore, the study would be able to both assess the performance of the AI algorithm along the acceptance and satisfaction of patients with automated approaches.

For the generalizability of the study, the best performing AI algorithm in internal validation should be tested on multiple different ethnic cohorts before adopting its use in clinical practices. On the other hand, further AI devices can be developed that can help the patients of DN to maintain the risk of ESRD at a low rate. However, caution must be adopted before adopting AI algorithms. Such as, whether AI-enabled care is safe and effective enough from the results of robust studies. As well as, the machine-generated results need to be confirmed by a physician or not etc (38).

The review has a few limitations. At the outset, the review only searched open-access databases. Thus, ignoring registers and webpages and other databases such as Scopus and Embase etc. However, this study aimed to only include articles published in peer-reviewed journals to avoid any existing biases. On the other hand, to ensure the reproducibility of the review the manuscript only included open-access databases.

In this study, a systematic effort was made by the author to identify every possible peer-reviewed article available and review AI application in DN. DN is one of the most dangerous complications caused by DM. However, with the proper caution and early detection of DN, a lot of the harmful outcomes can be avoided. The advent of computational biology, with the vast amount of EHR datasets generated every day, enables the further in-depth application of AI in the diagnosis and treatment of DN. Despite its recent emergence in healthcare, they have already outperformed conventional statistical methods. It is high time that physicians adopt AI-based algorithms in clinical approaches of DN.

Funding:

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors

References

1. Alberti KG, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabet Med.* 1998;15(7):539–53.
2. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet.* 2012;13(6):395–405.

3. Shih KC, Kwong ASK, Wang JHL, Wong JKW, Ko WWK, Lai JSM, et al. Diabetic retinopathy screening during the coronavirus disease 2019 pandemic. *EYE*. 2020;34(7):1246–7.
4. Samsu N. Diabetic nephropathy: Challenges in pathogenesis, diagnosis, and treatment. *Biomed Res Int*. 2021;2021:1497449.
5. USRDS 2015 Annual Data Report. Chapter 1: Incidence, Prevalence, Patient Characteristics, and Treatment Modalities. <https://www.usrds.org/2015/view>.
6. Narres M, Claessen H, Droste S, Kvitkina T, Koch M, Kuss O, et al. The incidence of end-stage renal disease in the diabetic (compared to the non-diabetic) population: A systematic review. *PLoS One*. 2016;11(1):e0147329.
7. Pippias M, Kramer A, Noordzij M, Afentakis N, Alonso de la Torre R, Ambühl PM, et al. The European Renal Association - European Dialysis and transplant association Registry annual report 2014: A summary. *Clin Kidney J*. 2017;10(2):154–69.
8. Orasanu G, Plutzky J. The pathologic continuum of diabetic vascular disease. *J Am Coll Cardiol*. 2009;53(5 Suppl):S35-42.
9. Gross JL, de Azevedo MJ, Silveiro SP, Canani LH, Caramori ML, Zelmanovitz T. Diabetic nephropathy: diagnosis, prevention, and treatment. *Diabetes Care*. 2005;28(1):164–76.

10. Gheith O, Othman N, Nampoory N, Halimb M, Al-Otaibi T. Diabetic kidney disease: difference in the prevalence and risk factors worldwide. *J Egypt Soc Nephrol Transplant*. 2016;16(3):65.
11. Krizhevsky A, Sutskever I, Hinton G. E. In Curran Associates, Nevada; 2012.
12. Kim SS, Kim JH, Kim IJ. Current challenges in diabetic nephropathy: Early diagnosis and ways to improve outcomes. *Endocrinol Metab (Seoul)*. 2016;31(2):245–53.
13. Verhave JC, Gansevoort RT, Hillege HL, Bakker SJL, De Zeeuw D, de Jong PE, et al. An elevated urinary albumin excretion predicts de novo development of renal function impairment in the general population. *Kidney Int Suppl*. 2004;66(92):S18-21.
14. Atkins RC, Briganti EM, Lewis JB, Hunsicker LG, Braden G, Champion de Crespigny PJ, et al. Proteinuria reduction and progression to renal failure in patients with type 2 diabetes mellitus and overt nephropathy. *Am J Kidney Dis*. 2005;45(2):281–7.
15. M. C. *Pattern Recognition and Machine Learning*. 1st ed. New York, NY: Springer; 2006.

16. Zhu T, Li K, Herrero P, Georgiou P. Deep learning for diabetes: A systematic review. *IEEE J Biomed Health Inform.* 2021;25(7):2744–57.
17. Chaki J, Thillai Ganesh S, Cidham SK, Ananda Theertan S. Machine learning and artificial intelligence based Diabetes Mellitus detection and self-management: A systematic review. *J King Saud Univ - Comput Inf Sci [Internet].* 2020; Available from: <http://dx.doi.org/10.1016/j.jksuci.2020.06.013>
18. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine learning and data mining methods in diabetes research. *Comput Struct Biotechnol J.* 2017;15:104–16.
19. Nielsen KB, Lautrup ML, Andersen JKH, Savarimuthu TR, Grauslund J. Deep learning-based algorithms in screening of diabetic retinopathy: A systematic review of diagnostic performance. *Ophthalmol Retina.* 2019;3(4):294–304.
20. Tejedor M, Woldaregay AZ, Godtliebsen F. Reinforcement learning application in diabetes blood glucose control: A systematic review. *Artif Intell Med.* 2020;104(101836):101836.
21. Islam M, Poly TN, Yang HC, Atique S, Li Y-CJ. Deep learning for accurate diagnosis of glaucomatous optic neuropathy using digital fundus image: A meta-analysis. *Stud Health Technol Inform.* 2020;270:153–7.
22. Bellemo V, Lim G, Rim TH, Tan GSW, Cheung CY, Sadda S, et al. Artificial intelligence screening for diabetic retinopathy: The real-world emerging application. *Curr Diab Rep.* 2019;19(9):72.

23. Marx V. Biology: The big challenges of big data: Biology. *Nature*. 2013;498(7453):255–60.
24. Mattmann CA. A vision for data science. *Nature*. 2013;493(7433):473–5.
25. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. *J Big Data [Internet]*. 2019;6(1). Available from: <http://dx.doi.org/10.1186/s40537-019-0217-0>
26. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst*. 2014;2(1):3.
27. Cichosz SL, Johansen MD, Hejlesen O. Toward big data analytics: Review of predictive models in the management of diabetes and its complications: Review of predictive models in management of diabetes and its complications. *J Diabetes Sci Technol*. 2015;10(1):27–34.
28. Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism*. 2017;69S:S36–40.
29. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*. 2017;2(4):230–43.

30. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. *Science*. 2015;349(6245):255–60.
31. Ognjanovic I. Artificial intelligence in healthcare. *Stud Health Technol Inform*. 2020;274:189–205.
32. Senanayake S, White N, Graves N, Healy H, Baboolal K, Kularatna S. Machine learning in predicting graft failure following kidney transplantation: A systematic review of published predictive models. *Int J Med Inform*. 2019;130(103957):103957.
33. Pham Q, Gamble A, Hearn J, Cafazzo JA. The need for ethnoracial equity in Artificial Intelligence for Diabetes Management: Review and recommendations. *J Med Internet Res*. 2021;23(2):e22320.
34. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng*. 2018;2(3):158–64.
35. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature*. 1986;323(6088):533–6.
36. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform*. 2018;19(6):1236–46.
37. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Syst Rev*. 2021;10(1):89.

38. Fogel AL, Kvedar JC. Artificial intelligence powers digital medicine. *NPJ Digit Med.* 2018;1(1):5.

39. Chan L, Nadkarni GN, Fleming F, McCullough JR, Connolly P, Mosoyan G, et al. Derivation and validation of a machine learning risk score using biomarker and electronic patient data to predict progression of diabetic kidney disease. *Diabetologia.* 2021;64(7):1504–15.

40. Ahn H-S, Kim JH, Jeong H, Yu J, Yeom J, Song SH, et al. Differential urinary proteome analysis for predicting prognosis in type 2 diabetes patients with and without renal dysfunction. *Int J Mol Sci.* 2020;21(12):4236.

41. Belur Nagaraj S, Pena MJ, Ju W, Heerspink HL, The BEAt-DKD Consortium. Machine-learning–based early prediction of end-stage renal disease in patients with diabetic kidney disease using clinical trials data. *Diabetes Obes Metab.* 2020;22(12):2479–86.

42. Kitamura S, Takahashi K, Sang Y, Fukushima K, Tsuji K, Wada J. Deep learning could diagnose diabetic nephropathy with renal pathological immunofluorescent images. *Diagnostics (Basel).* 2020;10(7):466.

43. Chen J, Tang H, Huang H, Lv L, Wang Y, Liu X, et al. Development and validation of new glomerular filtration rate predicting models for Chinese patients with type 2 diabetes. *J Transl Med.* 2015;13(1):317.

44. Bao X, Xie S, Zhang K, Song K, Yang Y. Machine learning based information extraction for diabetic nephropathy in clinical text documents. In: 2019 6th International Conference on Systems and Informatics (ICSAI). IEEE; 2019.
45. Hayashi Y. Detection of lower albuminuria levels and early development of diabetic kidney disease using an artificial intelligence-based rule extraction approach. *Diagnostics (Basel)*. 2019;9(4):133.
46. Cho BH, Yu H, Kim K-W, Kim TH, Kim IY, Kim SI. Application of irregular and unbalanced data to predict diabetic nephropathy using visualization and feature selection methods. *Artif Intell Med*. 2008;42(1):37–53.
47. Ginley B, Lutnick B, Jen K-Y, Fogo AB, Jain S, Rosenberg A, et al. Computational segmentation and classification of diabetic glomerulosclerosis. *J Am Soc Nephrol*. 2019;30(10):1953–67.
48. Huang J, Huth C, Covic M, Troll M, Adam J, Zukunft S, et al. Machine learning approaches reveal metabolic signatures of incident chronic kidney disease in individuals with prediabetes and type 2 diabetes. *Diabetes*. 2020;69(12):2756–65.
49. Kim Y-M, Kathuria P, Delen D. Machine learning to compare frequent medical problems of African American and Caucasian diabetic kidney patients. *Healthc Inform Res*. 2017;23(4):241.

50. Kim JE, Park SJ, Kim YC, Min S-I, Ha J, Kim YS, et al. Deep learning-based quantification of visceral fat volumes predicts post-transplant diabetes mellitus in kidney transplant recipients. *Front Med (Lausanne)*. 2021;8:632097.

51. Song X, Waitman LR, Yu AS, Robbins DC, Hu Y, Liu M. Longitudinal risk prediction of chronic kidney disease in diabetic patients using a temporal-enhanced gradient boosting machine: Retrospective cohort study. *JMIR Med Inform*. 2020;8(1):e15510.

52. Song X, Waitman LR, Hu Y, Yu ASL, Robbins DC, Liu M. Robust clinical marker identification for diabetic kidney disease with ensemble feature selection. *J Am Med Inform Assoc*. 2019;26(3):242–53.

53. Heinzl A, Kammer M, Mayer G, Reindl-Schwaighofer R, Hu K, Perco P, et al. Validation of plasma biomarker candidates for the prediction of eGFR decline in patients with type 2 diabetes. *Diabetes Care*. 2018;dc180532.

54. Makino M, Yoshimoto R, Ono M, Itoko T, Katsuki T, Koseki A, et al. Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning. *Sci Rep*. 2019;9(1):11862.

55. Rodriguez-Romero V, Bergstrom RF, Decker BS, Lahu G, Vakilynejad M, Bies RR. Prediction of nephropathy in type 2 diabetes: An analysis of the ACCORD trial applying machine learning techniques. *Clin Transl Sci*. 2019;12(5):519–28.

56. Lin S, Guo Y. Value analysis of using urinary microalbumin in artificial intelligence medical institutions to detect early renal damage in diabetes. *J Healthc Eng.* 2021;2021:6678454.
57. Zhang K, Liu X, Xu J, Yuan J, Cai W, Chen T, et al. Deep-learning models for the detection and incidence prediction of chronic kidney disease and type 2 diabetes from retinal fundus images. *Nat Biomed Eng.* 2021;5(6):533–45.
58. Zhao L, Ren H, Zhang J, Cao Y, Wang Y, Meng D, et al. Diabetic retinopathy, classified using the Lesion-aware Deep Learning System, predicts diabetic end-stage renal disease in Chinese patients. *Endocr Pract.* 2020;26(4):429–43.
59. Xu H, Kong Y, Tan S. Predictive modeling of diabetic kidney disease using random forest algorithm along with features selection. In: *Proceedings of the 2020 International Symposium on Artificial Intelligence in Medical Sciences*. New York, NY, USA: ACM; 2020.
60. Parab J, Sequeira M, Lanjewar M, Pinto C, Naik G. Backpropagation neural network-based machine learning model for prediction of blood urea and glucose in CKD patients. *IEEE J Transl Eng Health Med.* 2021;9:4900608.
61. Zhang J, Kowsari K, Boukhechba M, Harrison J, Lobo J, Barnes L. Sparse longitudinal representations of electronic health record data for the early detection of chronic kidney disease in diabetic patients. In: *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE; 2020.

62. Chauhan K, Nadkarni GN, Fleming F, McCullough J, He CJ, Quackenbush J, et al. Initial validation of a machine learning-derived prognostic test (KidneyIntelX) integrating biomarkers and electronic health record data to predict longitudinal kidney outcomes. *Kidney360*. 2020;1(8):731–9.

63. Abràmoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, et al. Improved automated detection of Diabetic Retinopathy on a publicly available dataset through integration of deep learning. *Invest Ophthalmol Vis Sci*. 2016;57(13):5200–6.

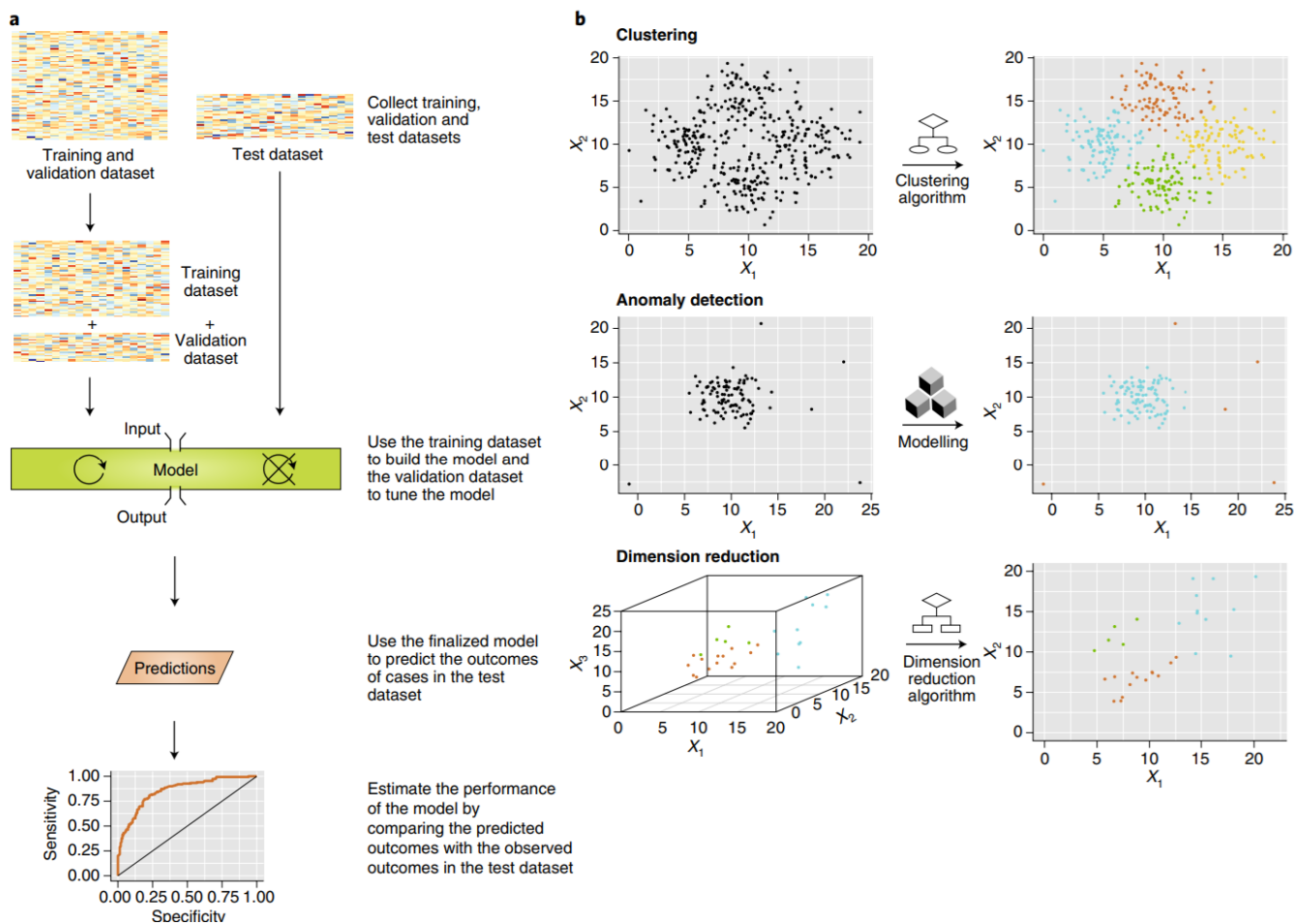


Figure 1: | Supervised and unsupervised machine learning. **a**, General workflow of supervised machine-learning approaches. First, training and test datasets are collected. Next, part of the training set is used to build the prediction model, and the other part to tune and validate the model (circular arrow). After the machine-learning model is finalized (crossed-out circular arrow), the established model is used to generate predictions on the test dataset and the model’s performance is estimated by comparing the predicted outcomes with the observed outcomes for the test dataset. **b**, Unsupervised machine learning includes clustering, anomaly detection and dimensionality reduction. Clustering algorithms group data points with similar measurements into clusters. Anomaly detection identifies outliers in the dataset. Dimensionality reduction reduces the number of random variables used to describe the data; for example, by representing an image with thousands of parameters as a smaller vector of summary features. The resulting summary vector preserves the important information in the raw data; for example, summary vectors from similar images will bear more resemblance than those obtained from irrelevant images

Note: Descriptive phrase that serves as title and description. Reprinted from “Artificial Intelligence in Healthcare” by Ognjanovic I, Year 2020, Stud Health Technol Inform, *Volume(issue)*, doi: 10.3233/SHTI200677. PMID: 32990674 (7).

