




Machine Learning Models for Prediction of Diabetic Microvascular Complications

Journal of Diabetes Science and Technology
2024, Vol. 18(2) 273–286
© 2024 Diabetes Technology Society
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/19322968231223726
journals.sagepub.com/home/dst



Sarah Kanbour, MD¹, Catharine Harris, BS²,
Benjamin Lalani², Risa M. Wolf, MD³,
Hugo Fitipaldi, PhD⁴, Maria F. Gomez, PhD⁴,
and Nestoras Mathioudakis, MD, MHS²

Abstract

Importance and Aims: Diabetic microvascular complications significantly impact morbidity and mortality. This review focuses on machine learning/artificial intelligence (ML/AI) in predicting diabetic retinopathy (DR), diabetic kidney disease (DKD), and diabetic neuropathy (DN).

Methods: A comprehensive PubMed search from 1990 to 2023 identified studies on ML/AI models for diabetic microvascular complications. The review analyzed study design, cohorts, predictors, ML techniques, prediction horizon, and performance metrics.

Results: Among the 74 identified studies, 256 featured internally validated ML models and 124 had externally validated models, with about half being retrospective. Since 2010, there has been a rise in the use of ML for predicting microvascular complications, mainly driven by DKD research across 27 countries. A more modest increase in ML research on DR and DN was observed, with publications from fewer countries. For all microvascular complications, predictive models achieved a mean (standard deviation) c-statistic of 0.79 (0.09) on internal validation and 0.72 (0.12) on external validation. Diabetic kidney disease models had the highest discrimination, with c-statistics of 0.81 (0.09) on internal validation and 0.74 (0.13) on external validation, respectively. Few studies externally validated prediction of DN. The prediction horizon, outcome definitions, number and type of predictors, and ML technique significantly influenced model performance.

Conclusions and Relevance: There is growing global interest in using ML for predicting diabetic microvascular complications. Research on DKD is the most advanced in terms of publication volume and overall prediction performance. Both DR and DN require more research. External validation and adherence to recommended guidelines are crucial.

Keywords

diabetes mellitus, machine learning, microvascular complications, risk prediction

Introduction

Diabetes, particularly when accompanied by conditions such as hypertension and hyperlipidemia, increases the risk of the microvascular complications of diabetic retinopathy (DR), diabetic kidney disease (DKD), and diabetic neuropathy (DN). These complications are major contributors to the morbidity and mortality associated with diabetes, leading to severe outcomes such as amputations, end-stage renal disease (ESRD), and vision loss.^{1,2} Given the significant clinical and economic impact of these conditions, the development and utilization of predictive models, particularly those leveraging machine learning and artificial intelligence (ML/AI), have become crucial. These models are

instrumental in healthcare for several reasons: They predict disease progression allowing for earlier interventions,

¹Diabetes Centre at AMAN Hospital, Doha, Qatar

²Division of Endocrinology, Diabetes, & Metabolism, Johns Hopkins University School of Medicine, Baltimore, MD, USA

³Division of Pediatric Endocrinology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

⁴Department of Clinical Sciences, Lund University Diabetes Centre, Lund University, Malmö, Sweden

Corresponding Author:

Nestoras Mathioudakis, MD, MHS, Division of Endocrinology, Diabetes, & Metabolism, Johns Hopkins University School of Medicine, 1830 East Monument Street, Suite 333, Baltimore, MD 21287, USA.
Email: nmathio1@jhmi.edu

enable personalized treatment strategies, and assist in judicious resource allocation.³ ML/AI models, by analyzing extensive electronic health record (EHR) data, can identify complex patterns more effectively than traditional methods, offering rapid and insightful disease predictions.³

Our study addresses a gap in current research of ML/AI prediction models, which often isolates specific microvascular complications or lacks in-depth evaluation of high-performing model features.^{4,5} We present a comprehensive review that encompasses all three major complications (DR, DKD, and DN), focusing on patients with type 2 diabetes (T2D), where multiple risk factors play a significant role in complication development. This review aims to (1) summarize current research on ML models for predicting microvascular complications, (2) compare and contrast various ML approaches, and (3) highlight features of ML models that offer the highest predictive value. This approach not only provides a comparative analysis of ML algorithms across different complications but also deepens our understanding of the factors that may enhance the predictive utility of ML/AI models.

Methods

Study Design and Article Selection

We conducted a literature search on PubMed using the following MeSH terms: “Machine Learning/Artificial intelligence” AND (“diabetic retinopathy” OR “diabetic neuropathy” OR “diabetic nephropathy” OR “diabetic foot”). Details of our search strategy are in the Supplemental Document. We selected articles meeting eligibility criteria published from January 1990 to July 2023. We adhered to the SANRA scale in both the study design and the reporting of results.⁶

Eligibility Criteria

We included longitudinal studies, such as randomized controlled trials, prospective or retrospective cohort studies, and registries, evaluating ML models for predicting microvascular complications (DR, DKD, and DN). As our focus was on predicting the onset of microvascular complications, we excluded studies centered on microvascular complication detection (ie, cross-sectional analysis) and image-based screening (eg, AI-based retinal detection). We also excluded case-control and genetic studies. We included studies with unspecified diabetes types, assuming a majority had T2D; however, we excluded those solely involving type 1 or gestational diabetes patients. We aimed to review and evaluate prediction models for microvascular complications in ambulatory settings, and therefore, excluded studies on short-term outcomes during hospitalization, such as imminent amputations in diabetic foot ulcer patients. We excluded studies lacking predictive performance measures such as c-statistic, accuracy, or positive predictive values (PPVs)/negative

predictive values (NPVs). We included meta-analyses that provided external validation and assisted in cross-referencing to identify additional eligible studies.

Data Extraction

For eligible studies, we collected details such as author, country, publication year, design, and type of microvascular complication(s). We also noted each model’s validation method (internal/external), cohort sizes, cohort descriptions, predictor variables, the ML algorithm used, prediction horizon, and performance measures, including c-statistic, accuracy, PPV, NPV, positive likelihood ratio (PLR), negative likelihood ratio (NLR), and the F1-Score. As most evaluated studies tended to report c-statistics and omitted other measures of model performance, we focused the analysis on that metric. When the prediction horizon was not explicitly stated, we extracted the mean or median follow-up time reported in the study. Internal validation was defined as both development and validation within the same cohort, while external validation involved a distinct cohort from a different institution, population, or country. Studies using temporal validation (ie, evaluating individuals from the same cohort but in a different time frame) were categorized as internal validation, regardless of how the studies labeled them.⁷

The methods for internal validation were classified into one or more of the following: split-sample (hold-out method), bootstrapping (bagging), cross-validation (k-fold, nested), temporal validation, or not specified. We categorized ML/AI algorithms as linear regression (LinReg), including generalized additive models; logistic regression (LogReg), including multinomial logistic regression; tree-based methods (Random Forest [RF], Decision Trees [DT], Gradient Boosting Machines [GBM], XGBoost, LightGBM); Support Vector Machines (SVM); K-Nearest Neighbors (K-NN); Neural Networks (NN); ensemble methods (ie, AdaBoost); Survival Analysis (SA), including Cox, accelerated failure, competing risk, Fine and Gray, and Weibull models; and probabilistic models (Naïve Bayes [NB] and Markov Models). We collected and managed study data using REDCap electronic data capture tools hosted by Johns Hopkins University.^{8,9}

Predictor Variables

When studies reported results from several ML models using clinically available predictors, we extracted data from all these models. For studies evaluating novel biomarkers alongside clinically available predictors, we focused only on the most fully adjusted models. We extracted predictor variables as specified in the studies, either as exact variables (eg, A1C, blood pressure) or broad categories (eg, diagnoses, labs, medications). We reclassified variables representing similar concepts for analysis (eg, weight and waist circumference), as detailed in Supplemental Table 9.

Outcome Definition

We collected information on the precise outcomes used to define each microvascular complication. Diabetic retinopathy was defined by one or more of the following: retinopathy detected via dilated funduscopy, blindness, vitrectomy, retinal photocoagulation, and others (typically a diagnostic code for retinopathy). Diabetic kidney disease definitions included microalbuminuria, macroalbuminuria, albuminuria not otherwise specified (NOS), chronic kidney disease (CKD) diagnosis, CKD progression markers (serum creatinine doubling, estimated glomerular filtration rate [eGFR] drop, CKD stage decline), ESRD (dialysis or renal therapy), and renal disease-related death. Diabetic neuropathy was defined through the Michigan Neuropathy Screening Instrument Score (NSIS), physical examination findings (such as vibratory sensory loss, absent ankle jerk reflex, loss of protective sensation), electromyography/nerve conduction studies (EMG/NCS) results, lower extremity amputation, or neuropathy diagnostic codes. Many ML models used a composite of these individual outcomes.

Statistical Analysis

Data normality was assessed using skewness and kurtosis tests. Non-normally distributed data were summarized with medians and interquartile ranges, while normally distributed data were reported using means and standard deviations. For each ML algorithm, we calculated the mean, minimum, and maximum c-statistic across model types and microvascular outcomes. Analyses were conducted separately for internal and external validation results, with subgroup analyses focusing on individual outcome definitions, the number of predictor variables, and prediction horizons. Statistical analyses were performed using STATA (Release 18), RStudio (2020), and Python.

Results

We identified 74 studies that met our eligibility criteria. Of these, 66 studies presented the original version of the ML model,¹⁰⁻⁷⁵ and an additional eight studies externally validated these models.⁷⁶⁻⁸³ This yielded results for 256 and 124 internally and externally validated ML models, respectively. Study designs, outcomes, and countries of origin are detailed in Supplemental Table 1, with ML model details in Supplemental Tables 2 to 7. The most common study designs were retrospective cohorts (49.2%), prospective cohorts (28.4%), randomized controlled trials (13.4%), and registries (9.0%). The frequency of study types by microvascular complication is shown in Supplemental Figure 3, and the individual and composite outcome definitions used in the ML models are in Supplemental Figures 5 and 6.

Global Trends in ML Prediction Models for Microvascular Complications

Figure 1 shows an increasing trend in using ML-based models for predicting microvascular complications, with a notable surge in DKD studies since 2010. This rise aligns with the general growth in ML applications in endocrinology.⁸⁴ While AI-based retinal image detection using ML has grown significantly, the trend for ML-based predictions of incident retinopathy and neuropathy has been stable.

Publications on ML models for predicting microvascular complications show global involvement, with 31 countries represented in the studies (Figure 2). The total numbers of contributing countries for DR, DKD, and DN are 16, 27, and 6, respectively. The United Kingdom leads in DR research with 41.3% of publications, followed by the United States (15.3%) and Italy (15.0%). The DKD research landscape is more varied, with the United States (18.3%) and China (17.7%) leading, followed by Australia, the United Kingdom, Japan, and Taiwan. Diabetic neuropathy research is dominated by Italy, the United States, and the United Kingdom, contributing over 80% of publications.

Model Characteristics and Predictive Performance

Table 1 summarizes the key findings of the ML models. Of the included studies, 256 models underwent internal validation, and 124 had external validation. The c-statistic was the main performance metric, reported for 47 DR, 181 DKD, and 28 DN models in internal validation, and for 39 DR, 81 DKD, and four DN models in external validation. For internal validation, 19.1% of DR, 41.9% of DKD, and 23.6% of DN models omitted the 95% confidence interval; for external validation, the absence of 95% confidence interval (CI) was less frequent, missing in 10.3% of DR, 17.2% of DKD, and none of the DN models. In addition, some studies mislabeled internal validation as external validation. Validation methods are detailed in Supplemental Figure 4.

On average, models had 12 predictors for internal validation and eight for external validation, with prediction horizons of 5 and 6.4 years, respectively. The mean [standard deviation] c-statistics were 0.79 (0.09) for internal and 0.72 (0.12) for external validation. Models with shorter prediction horizons and fewer predictors typically demonstrated better discrimination. Outcome definitions are detailed in Supplemental Figures 5 and 6.

In internal validation, DKD models showed the highest discrimination with a mean c-statistic of 0.81 (0.09), followed by DR at 0.74 (0.10) and DN at 0.71 (0.09). This trend persisted in external validation. Specifically, 181 DKD models were internally validated (c-statistic 0.81 [0.09]) and 81 externally (c-statistic 0.74 [0.13]), with models including ESRD as an outcome achieving the highest internal predictive

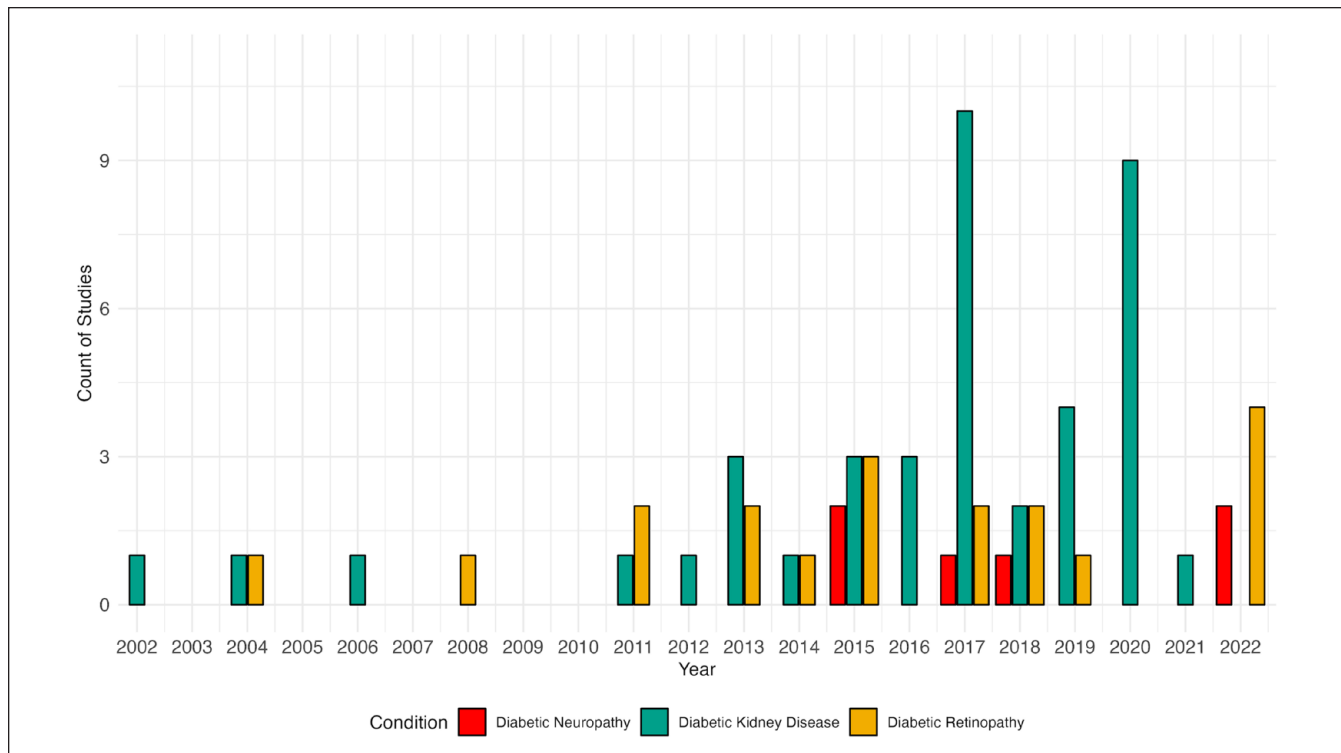


Figure 1. Trend in publications of machine learning models by microvascular complications from 2002 to 2023. Symbols indicate the number of publications per year for diabetic retinopathy, diabetic kidney disease, and diabetic neuropathy. For 2023, the number of publications corresponds to the first seven months of the year.

performance. For DR, 47 models underwent internal validation and 39 external validation, resulting in mean c-statistics of 0.74 (0.10) and 0.71 (0.11), respectively. Models incorporating retinopathy diagnosed via dilated fundoscopy outperformed others. Diabetic neuropathy was represented by 28 internal models and four external ones, achieving c-statistics of 0.71 (0.09) and 0.67 (0.15), respectively.

We investigated the impact of prediction horizon on model performance. In internal validation, the median prediction horizons for DR, DKD, DN, and all complications were 3.1, 5.0, 5.0, and 5.0 years, respectively, and for external validation, they were 10.0, 5.0, 10.0, and 6.4 years. Diabetic retinopathy models with shorter horizons showed higher discrimination, with c-statistics of 0.79 (0.10) for models below the five-year median versus 0.70 (0.10) for those at or above it. Diabetic kidney disease model performance was not significantly affected by the prediction horizon, but for DN, shorter horizons led to higher discrimination, with c-statistics of 0.73 (0.10) versus 0.67 (0.08).

We assessed whether the number of predictor variables influences model performance, stratifying results by complication, and whether predictor counts were below or at/above the median. In internal validation, the median numbers of predictors for DR, DKD, DN, and all microvascular complications were 7, 13, 11, and 12, respectively; for external validation, they were 6, 12, 12, and 8. Diabetic retinopathy models with fewer predictors showed better performance

(0.79 vs 0.70). The number of predictors did not significantly affect DKD model performance. For DN, models with fewer predictors performed slightly better, with scores of 0.72 (0.12) versus 0.70 (0.06).

Predictor Variables

Figure 3 displays a heatmap of the predictor variables used in the ML models. Common predictors across all complications include age, sex, smoking, body mass index (BMI), blood pressure, A1C, and creatinine. For DR, unique predictors include diabetes duration and retinopathy history. Diabetic kidney disease models frequently use variables like urine albumin-creatinine ratio (UACR), HDL (high-density lipoprotein) cholesterol, estimated glomerular filtration rate (eGFR), cardiovascular diseases (CVD), triglycerides, LDL (low-density lipoprotein) cholesterol, and uric acid. Diabetic neuropathy models typically incorporate race/ethnicity, socioeconomic factors, hypertension, diabetes duration, and hemoglobin.

Performance of Predictor Variables

Determining the independent predictive value of specific variables for microvascular complications would require studies to compare models with different combinations of predictors, while keeping other factors constant (eg, validation cohort, prediction horizon, outcome definition). Due to the variability



Figure 2. Country of affiliation for study authors by microvascular complication. Results reflect the proportion of studies reporting on the outcomes of diabetic retinopathy, diabetic kidney disease, and diabetic neuropathy. The data utilized for generating this visualization were acquired from PubMed and PubMed Central. We employed a combination of manual curation and the application of text mining functions that were developed using R software version 4.1.2. To arrive at the ultimate proportions of ancestries, we calculated them individually for each distinct study and subsequently aggregated them.

in prediction models across studies, it was beyond the scope of this review to quantitatively assess the relative influence of specific predictors. However, insights were gleaned from previous meta-analyses that externally validated various prediction models for DR and DKD, using different predictor combinations while maintaining other parameters constant. Specifically, for DR, in validation studies with the Hoorn Diabetes Care System prospective cohort using Cox models,⁸⁵ a set of predictors for DR outcomes showed superior performance: gender, diabetes duration, A1C, systolic blood pressure, albuminuria, creatinine clearance, and DR.¹⁵ For DKD, validated with the same cohort and modeling technique and focusing on microalbuminuria (UACR ≥ 30 mg/g) over two-, five-, or 10-year horizons, a distinct combination of predictors proved more effective⁸¹: ethnicity, A1C, systolic blood pressure, UACR, eGFR, DR, baseline anti-hypertensive medication, and waist circumference.¹⁷

Predictive Performance by ML Technique

Figure 4 presents the internal validation performance of ML techniques for each microvascular complication, with external

validation results in Supplemental Figure 2. The frequency of ML techniques and specific performance metrics are in Supplemental Figure 1 and Table 8, respectively.

The number of models per technique is indicated by “N” values. For DR, top-performing ML techniques were the Markov model (N = 3), XGBoost (N = 2), and RF (N = 5), while K-NN (N = 1) and Naïve Bayes (N = 3) had the lowest performance. In DKD, LightGBM (N = 2), XGBoost (N = 13), and Neural Network (N = 5) were top performers, with Gradient Boosting Machine (N = 5) and AdaBoost (N = 2) showing the lowest, yet still relatively high, discrimination. For DN, XGBoost (N = 1), RF (N = 4), and Support Vector Machine (N = 4) performed best, whereas Survival Analysis (N = 7) and Naïve Bayes (N = 1) had lower discriminative ability. Commonly, XGBoost, RF, and Logistic Regression consistently showed high discrimination across all complications, while Survival Analysis and Naïve Bayes underperformed.

Similar observations were made in studies that kept other model characteristics constant while only varying their machine learning technique.^{32,44,60,62,67,69,71,73} For example, in the case of DR, the study by Zhao et al⁶² compared five ML

Table 1. Summary of Key Findings of Machine Learning Models by Microvascular Complication.

Outcome/subgroup	Internal validation			External validation				
	No. of models	No. of predictors ^a	Prediction horizon, years ^a	C-statistic ^b	No. of models	No. of predictors ^a	Prediction horizon, years ^a	C-statistic ^b
Any microvascular complication	256	12 (7, 18)	5.0 (3.0, 10.0)	0.79 [0.09]	124	8 (6, 14)	6.4 (5.0, 10.0)	0.72 [0.12]
Prediction horizon								
Below median	108	13 (6, 30)	3.0 (3.0, 3.1)	0.79 [0.10]	60	7 (6, 13)	5.0 (2.0, 5.0)	0.75 [0.12]
At or above median	148	11 (7, 14)	5.0 (5.0, 10.0)	0.78 [0.10]	64	9 (7, 14)	10.0 (10.0, 10.0)	0.70 [0.11]
No. of predictors								
Below median	127	7 (5, 7)	5.0 (3.0, 5.0)	0.79 [0.11]	65	6 (6, 7)	5.0 (3.0, 10.0)	0.72 [0.14]
At or above median	129	18 (14, 46)	5.0 (3.0, 5.0)	0.78 [0.09]	59	14 (12, 14)	10.0 (5.0, 10.0)	0.73 [0.10]
Diabetic retinopathy	47	7 (6, 30)	3.1 (3.0, 6.5)	0.74 [0.10]	39	6 (6, 9)	10.0 (5.0, 10.0)	0.71 [0.11]
Component of outcome definition								
Retinal photocoagulation	3	11 (11, 19)	10.0 (5.0, 10.0)	0.67 [0.05]	7	7 (6, 11)	10.0 (10.0, 10.0)	0.73 [0.15]
Vitrectomy	3	11 (11, 19)	10.0 (5.0, 10.0)	0.57 [0.05]	1	11	10.0	0.57
Blindness	5	11 (9, 12)	10.0 (5.0, 10.0)	0.68 [0.05]	7	6 (2, 9)	6.5 (5.0, 10.9)	0.62 [0.08]
Retinopathy (fundus)	34	7 (6, 11)	3.1 (3.0, 5.0)	0.77 [0.10]	31	6 (6, 9)	10.0 (3.0, 10.0)	0.73 [0.10]
Other	14	49 (12, 49)	5.0 (3.0, 5.0)	0.69 [0.06]	1	5	6.5	0.55
Prediction horizon								
Below median	24	7 (6, 40)	3.0 (3.0, 3.0)	0.79 [0.10]	15	6 (4, 6)	3.0 (3.0, 5.0)	0.69 [0.10]
At or above median	23	7 (7, 12)	6.5 (5.0, 10.0)	0.70 [0.09]	24	7 (6, 10)	10.0 (10.0, 10.0)	0.71 [0.11]
No. of predictors								
Below median	12	5 (4, 6)	3.0 (1.5, 5.0)	0.81 [0.10]	7	4 (2, 5)	3.0 (3.0, 6.4)	0.65 [0.08]
At or above median	35	11 (7, 30)	5.0 (3.0, 7.0)	0.72 [0.10]	32	7 (6, 9)	10.0 (8.2, 10.0)	0.72 [0.10]
Diabetic kidney disease	181	13 (6, 8)	5.0 (3.0, 5.0)	0.81 [0.09]	81	12 (7, 14)	5.0 (5.0, 10.0)	0.74 [0.13]
Component of outcome definition								
Microalbuminuria	40	18 (7, 46)	5.0 (3.0, 7.0)	0.76 [0.08]	32	8 (7, 14)	5.0 (2.5, 10.0)	0.66 [0.10]
Macroalbuminuria	19	18 (14, 46)	5.0 (3.0, 7.0)	0.79 [0.09]	25	14 (14, 14)	5.0 (5.0, 10.0)	0.70 [0.10]
Albuminuria NOS	2	7 (6, 8)	20.0 (20.0, 20.0)	0.82 [0.04]	—	—	—	—
CKD Diagnosis	74	13 (9, 26)	5.0 (3.1, 5.0)	0.78 [0.07]	23	7 (6, 12)	5.0 (2.0, 10.0)	0.74 [0.14]
Progression of CKD Stage	51	15 (6, 18)	4.0 (3.0, 6.0)	0.77 [0.06]	21	14 (7, 14)	5.0 (2.0, 10.0)	0.69 [0.05]

(Continued)

Table 1. (Continued)

Outcome/subgroup	Internal validation			External validation				
	No. of models	No. of predictors ^a	Prediction horizon, years ^a	C-statistic ^b	No. of models	No. of predictors ^a	Prediction horizon, years ^a	C-statistic ^b
ESRD	85	13 (6, 18)	5.0 (3.0, 5.0)	0.84 [0.09]	36	14 (7, 14)	5.0 (5.0, 10.0)	0.73 [0.12]
Death from renal disease	17	13 (8, 17)	5.0 (5.0, 10.0)	0.83 [0.08]	11	12 (5, 12)	5.1 (5.0, 10.0)	0.74 [0.18]
Prediction horizon								
Below median	74	13 (6, 18)	3.0 (3.0, 3.1)	0.81 [0.09]	17	8 (7, 14)	2.0 (2.0, 2.0)	0.73 [0.13]
At or above median	95	13 (7, 18)	5.0 (5.0, 10.0)	0.81 [0.08]	64	12 (7, 14)	6.5 (5.0, 10.0)	0.73 [0.13]
No. of predictors								
Below median	88	6 (5, 8)	5.0 (3.0, 5.0)	0.81 [0.10]	40	7 (6, 8)	5.0 (4.0, 8.3)	0.74 [0.14]
At or above median	93	18 (14, 23)	5.0 (3.0, 5.0)	0.80 [0.08]	41	14 (14, 14)	10.0 (5.0, 10.0)	0.73 [0.11]
Diabetic neuropathy	28	11 (7, 49)	5.0 (3.3, 7.0)	0.71 [0.09]	4	12 (8, 13)	10.0 (8.3, 10.0)	0.67 [0.15]
Component of outcome definition								
Michigan NSIS	1	14	10.0	0.60	—	—	—	—
Examination findings	3	14 (14, 14)	10.0 (10.0, 10.0)	0.61 [0.04]	1	12	10.0	0.69
EMG/NCS	13	7 (7, 7)	5.0 (3.0, 7.0)	0.71 [0.12]	—	—	—	—
LEA	2	12 (11, 13)	10.0 (10.0, 10.0)	0.73 [0.04]	2	12 (11, 13)	10.0 (10.0, 10.0)	0.77 [0.01]
Diagnostic code	9	49 (49, 49)	3.3 (3.3, 5.0)	0.75 [0.04]	1	5	6.5	0.44
Prediction horizon								
Below median	17	10 (7, 49)	3.3 (3.3, 5.0)	0.73 [0.10]	—	—	—	—
Above median	11	11 (7, 14)	10.0 (7.0, 10.0)	0.67 [0.08]	—	—	—	—
No. of predictors								
Below median	14	7 (7, 7)	5.0 (3.0, 7.0)	0.72 [0.12]	—	—	—	—
At or above median	14	49 (14, 49)	5.0 (3.3, 10.0)	0.70 [0.06]	—	—	—	—

When single studies are reported, SD and IQR are not provided. Dark green= highest c-statistic for outcome comparing internal to external validation; light green= highest c-statistic within subgroup category.

Abbreviations: NOS, not otherwise specified; CKD, chronic kidney disease; ESRD, end-stage renal disease; NSIS, Neuropathy Screening Instrument Score; EMG/NCS, electromyogram/nerve conduction study; LEA, lower extremity amputation; IQR, interquartile range.

^aMedian (interquartile range).

^bMean [standard deviation].

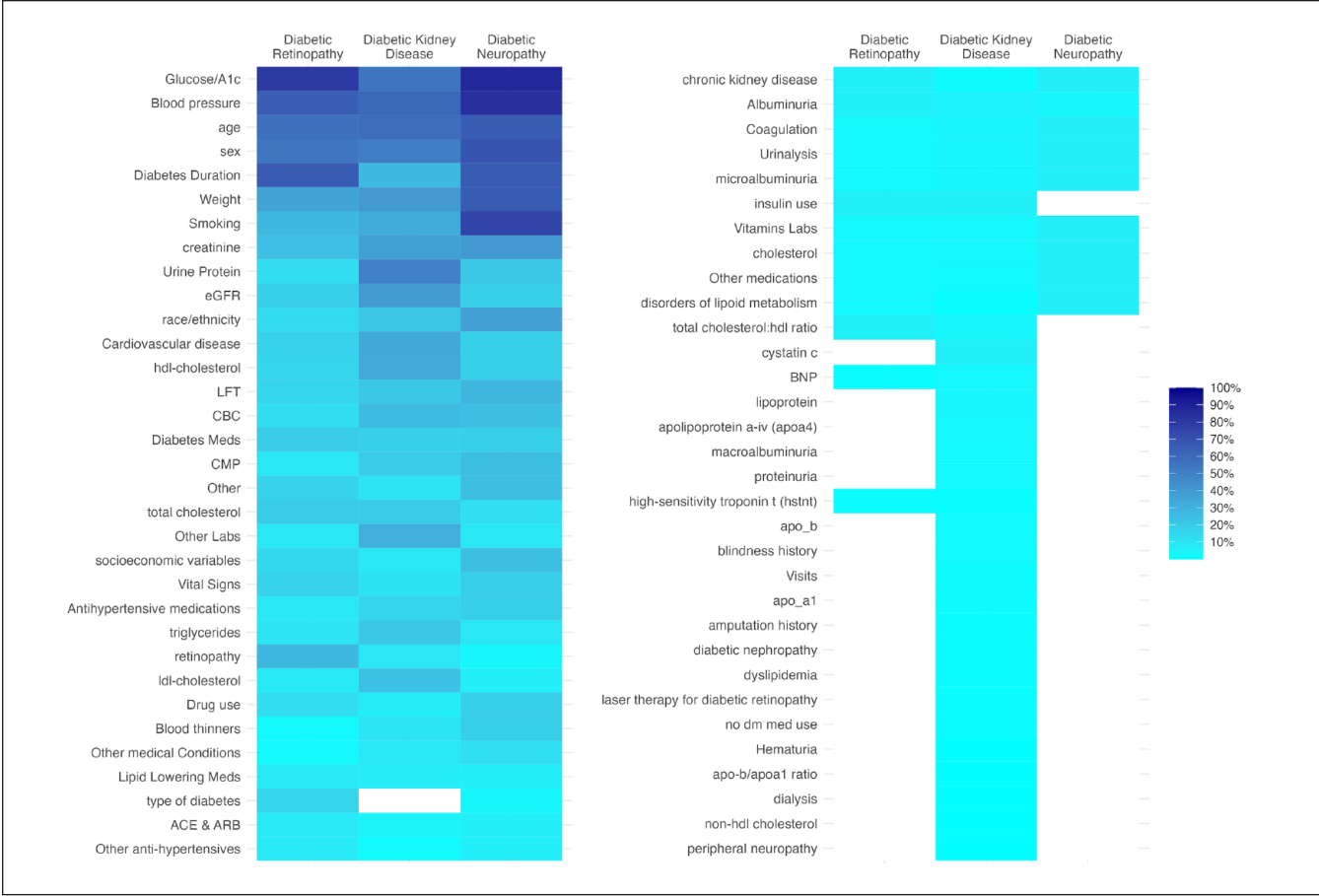


Figure 3. Heat map of predictor variables included in machine learning models by microvascular complication. Proportion of models using the category of predictor variable. Abbreviations: eGFR, estimated glomerular filtration rate; LFT, liver function test; CBC, complete blood count; CMP, comprehensive metabolic panel; HDL, high-density lipoprotein; apo_a1, apolipoprotein A1; DM, diabetes mellitus; apo-b, apolipoprotein B.

techniques: XGBoost (area under the curve [AUC]: 0.91 [0.9–0.93]), RF (AUC: 0.87 [0.86–0.89]), Logistic Regression (AUC: 0.81 [0.79–0.83]), SVM (AUC: 0.80 [0.78–0.82]), and K-NN (AUC: 0.63 [0.6–0.66]). In the case of DKD, the study by Dong et al⁶⁰ compared seven ML techniques: LightGBM (AUC: 0.82 [0.75–0.88]), AdaBoost (AUC: 0.81 [0.74–0.87]), Neural Network (AUC: 0.80 [0.73–0.87]), Logistic Regression (AUC: 0.80 [0.73–0.87]), XGBoost (AUC: 0.78 [0.71–0.85]), Support Vector Machine (AUC: 0.79 [0.72–0.86]), and DT (AUC: 0.58 [0.5–0.67]).

Discussion

In this comprehensive comparative review of ML-based prediction models for microvascular complications, we identified key trends and methodological findings from 74 longitudinal studies. The global reach of these studies, spanning 31 countries, underscores the international interest in this burgeoning field of research. Among the microvascular complications, the greatest volume of validated models and

overall prediction performance were observed for DKD. Predictive performance was invariably better on internal validation compared with external validation, as expected due to fundamental issues related to homogeneity of development cohorts, data set similarities, and overfitting. Model performance was influenced by the type and number of predictor variables, outcome definitions, prediction horizon, and the ML techniques used. Notably, XGBoost, RF, and Logistic Regression emerged as top-performing techniques across microvascular complications, although other techniques—such as LightGBM, and NN—also excelled for individual microvascular complications. In contrast, DT, Survival Analysis, K-NN, and Naïve Bayes had poorer predictive performance. Many studies offered only minimal performance metrics, such as the c-statistic (and often without confidence intervals), limiting the scope for detailed meta-analyses.

We speculate that the superior performance of DKD prediction models relative to DR and DN might stem from several reasons. One key factor is the use of large, prospective,

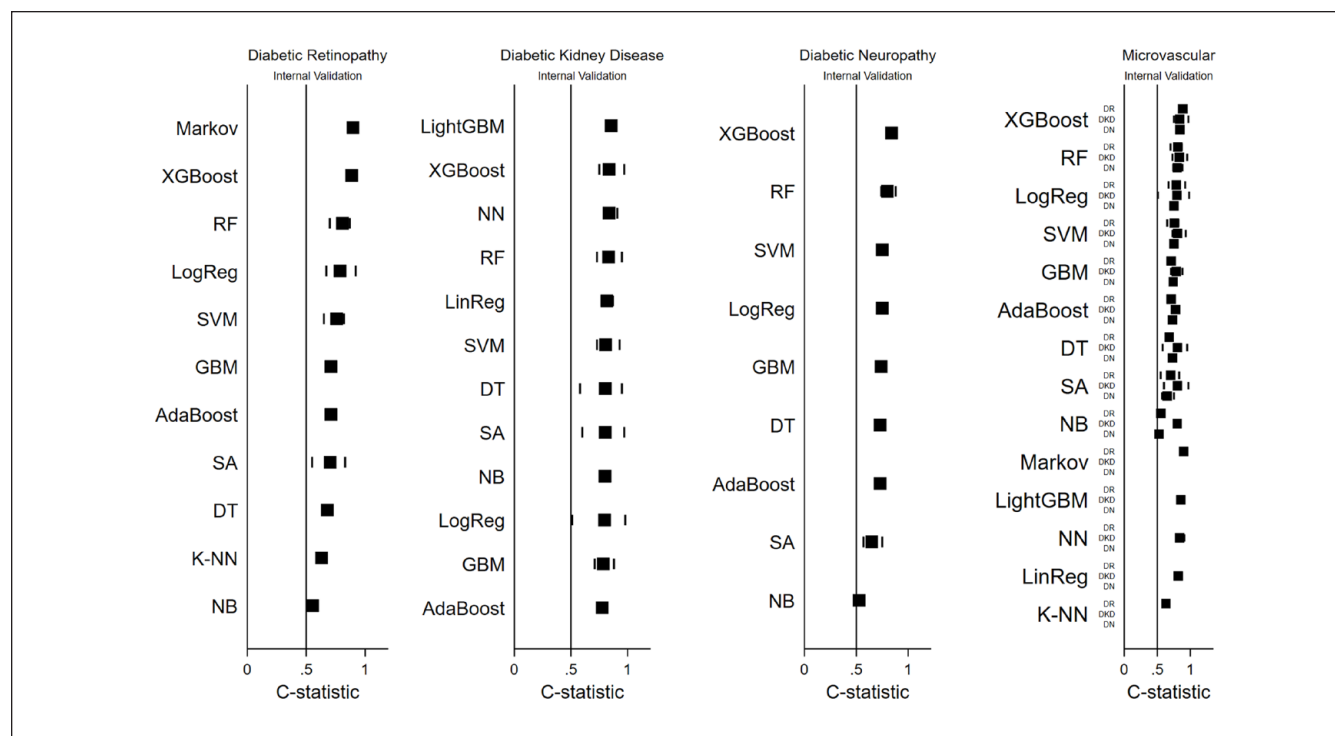


Figure 4. Model performance by ML technique for microvascular complications. Black squares indicate mean c-statistic; lines indicate minimum and maximum c-statistic for each ML model.

Abbreviations: ML, machine learning; RF, Random Forest; LogReg, logistic regression; SVM, Support Vector Machines; GBM, Gradient Boosting Machines; SA, survival analysis; DT, Decision Trees; K-NN, K-Nearest Neighbors; NB, Naïve Bayes; LinReg, Linear Regression.

multinational data sets with diverse type 2 diabetes patients.³ Diabetic kidney disease models benefit from a global development perspective, whereas DR models are mainly concentrated in North America, Europe, and East Asia. Despite advancements in ML-driven AI for retinal image recognition, DR prediction using EHR data has not seen similar growth. Diabetic neuropathy prediction models are even less widespread. Diabetic kidney disease models also stand out for their detailed staging of outcomes, unlike DR models that focus on later disease stages. The availability of clinical biomarkers (such as creatinine, eGFR, and urine albumin) in DKD may have enabled clearer differentiation between stages, improving predictive accuracy of ML models for DKD compared with DR and DN.³

Furthermore, model performance may be influenced by the type of predictors used.³ For DKD, typical predictors included age, sex, blood pressure, A1C, weight, smoking, creatinine, urine albumin-to-creatinine ratio, eGFR, HDL, LDL, triglycerides, and cardiovascular risk. Diabetic retinopathy models incorporated predictors such as age, sex, blood pressure, A1C, duration of diabetes, history of retinopathy, and diabetes type. Additional predictors for DKD are ethnicity, duration of diabetes, and DR,⁵² while DR unique predictors include history of pregnancy and cataract surgery.⁸⁶ No study in this review concurrently incorporated all these predictors for either DKD or DR, suggesting that

including more of these overlapping factors may improve model performance.

Beyond selecting clinically relevant predictors, it is also important to choose predictors that are stable over time, integrate seamlessly with the EHR as discrete data, are accurately measured, and are generalizable across diverse patient populations.³ For example, the presence of microvascular complications can provide insights into the duration of diabetes and the level of glycemic control. Such predictors tend to be more time-stable than baseline glucose or A1C levels, especially over long prediction durations. Due to the annual lab and imaging screenings recommended for monitoring DKD and DR, a wealth of EHR data is available for these microvascular complications.⁸⁷ Furthermore, with advances in AI-based image detection, enhanced DR screening and standardized reporting are on the horizon.⁸⁸ The relatively lower predictive performance for DN models is not unexpected given the lack of definitive biomarkers for DN, potential inconsistencies in recording neuropathy examination findings in the EHR (often not as discrete data), and the absence of a standardized system to monitor DN progression.

The prediction horizon is another fundamental factor for model performance. Longer horizons pose challenges in maintaining accuracy due to new data and secular trends. For instance, introduction of renal protective drugs like SGLT-2

inhibitors alters patient trajectories. An ideal prediction horizon should balance between being sufficiently long to observe the event of interest and short enough to keep baseline predictors relevant.³ For example, the five- and 10-year incidence rates for any DR in T2D are 20% and 75%, respectively, suggesting a five-year horizon for DR prediction; conversely, a 10-year horizon may be more suitable for diabetic macular edema, especially in newly diagnosed patients.⁸⁹

In our study, the prediction horizon's impact varied with the outcome. Shorter horizons generally yielded better performance, but DKD is an exception, likely because ESRD is a common outcome with a low incidence rate, potentially inflating the c-statistic performance due to high NPV.⁹⁰ Therefore, researchers should consider providing additional metrics such as PLR and NLR, which are less influenced by of outcome prevalence.⁹¹

The type of ML techniques may also play a role in model performance. Key differentiators among ML techniques include their ability to (1) capture complex interactions and handle nonlinearity, (2) manage missing data, (3) prevent overfitting and reduce misclassification errors, (4) optimize computational efficiency, and (5) maintain model interpretability. Our analysis aimed to assess how these techniques affect model performance, acknowledging that factors beyond modeling technique can influence outcomes. We found that XGBoost, RF, and Logistic Regression had the highest average c-statistics across all microvascular complications. XGBoost and RF, both ensemble ML algorithms, improve performance by integrating decisions from multiple models. These methods, in contrast to traditional logistic regression, which assumes a linear relationship between the outcome and predictor variables and struggles with missing data, demonstrate enhanced discrimination capabilities.^{92,93}

In addition, the Markov model demonstrated high predictive accuracy for DR, while LightGBM and NN excelled in DKD prediction. However, the Markov model's accuracy, primarily validated internally in three models, requires external validation confirmation.⁹⁴ LightGBM, akin to XGBoost, is efficient and effective with extensive EHR data.⁹⁵ Neural Networks, as unsupervised models, excel in identifying feature interactions and manage nonlinear relationships in large medical data sets.⁹⁶

Conversely, K-NN, Naïve Bayes, and Survival Analysis showed lower predictive accuracy for microvascular complications. K-NN struggles with numerous predictors, leading to misclassification and computational intensity.⁹⁷ Naïve Bayes' assumption of independent predictor variables is often unrealistic in microvascular complications.⁹⁸ Survival analysis, such as Cox models, assume constant hazard ratios, unsuitable for diabetic complications where risk factors vary over time. Alternative ML techniques might better capture complex patterns and nonlinear relationships in diabetic complications.

Our analysis identified a significant discrepancy in validation methods across studies. Many authors mislabeled their validation as "external" when it resembled split-sample validation within the same cohort used for training. This distinction is crucial for understanding model generalizability. Split-sample validation, dividing data into training and validation sets, assesses model reliability but lacks the robustness of true external validation. External validation tests the model on a completely different data set, often from another institution or region, highlighting its applicability across diverse patient groups and clinical scenarios.⁹⁹

Our findings build on a 2021 meta-analysis focused on microvascular complications, particularly DKD and DR.⁵ This analysis showed that only 25% of 71 studies underwent external validation. In the DR subset, only six studies used ML techniques, with internal validation c-statistics of 0.82 for logistic regression and 0.75 for Cox. For DKD, of 96 equations from 30 studies, only 17 used ML, with a c-statistic of 0.77. No ML techniques were used in 18 studies on ESRD, showing c-statistics of 0.87 internally and 0.86 externally. The meta-analysis highlighted a lack of external validation and a need for more ML studies. Our study revealed that many of these initially reported studies underwent external validation in subsequent meta-analyses.^{78,81}

In conclusion, ML shows promise for precision medicine in diabetes, aiding treatment decisions and early detection. Our study emphasizes the need for rigorous design, including using diverse cohorts for internal and independent ones for external validation, selecting efficient ML techniques for multidimensional EHR data, focusing on clinically relevant outcomes, setting appropriate time horizons, and choosing predictors for their ease of collection, bias minimization, stability, and predictive value. Future research should follow established guidelines and focus on integrating ML models into EHRs to assess impact on clinical outcomes, such as microvascular complications.

Abbreviations

AUC-ROC, area under the receiver operating curve; BMI, body mass index; CVD, cardiovascular diseases; CKD, chronic kidney disease; CI, confidence intervals; DT, Decision Trees; DKD, diabetic kidney disease; DN, diabetic neuropathy; DR, diabetic retinopathy; EMG/NCS, electromyography/nerve conduction studies; EHR, electronic health record; ESRD, end-stage renal disease; eGFR, estimated glomerular filtration rate; GBM, Gradient Boosting Machines; K-NN, K-Nearest Neighbors; LinReg, linear regression; LogReg, logistic regression; ML/AI, machine learning/artificial intelligence; NSIS, Michigan Neuropathy Screening Instrument Score; NB, Naïve Bayes; NLR, negative likelihood ratio; NPV, negative predictive value; NN, Neural Networks; NOS, not otherwise specified; PLR, positive likelihood ratio; PPV, positive predictive value; RF, Random Forest; RCT, randomized controlled trial; Stata, Stata Statistical Software; SA, survival analysis; T2D, type 2 diabetes; UACR, urine albumin-creatinine ratio.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: N.M. received funding from the NIDDK (K23DK111986). M.G. and H.F. received funding from the Swedish Heart-Lung Foundation (20190470), Swedish Research Council (EXODIAB, 2009-1039; 2018-02837), Swedish Foundation for Strategic Research (LUDC-IRC, 15-0067), and EU H2020-JTI-IMI2-2015-05 (Grant agreement number 115974—BEAT-DKD).

ORCID iDs

Sarah Kanbour  <https://orcid.org/0000-0002-3267-4419>

Benjamin Lalani  <https://orcid.org/0000-0002-7969-8629>

Nestoras Mathioudakis  <https://orcid.org/0000-0002-0210-655X>

Supplemental Material

Supplemental material for this article is available online.

References

- Collaborators GD. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet*. 2023;402(10397):203-234. doi:10.1016/S0140-6736(23)01301-6. (In English)
- Brownrigg JR, Hughes CO, Burleigh D, et al. Microvascular disease and risk of cardiovascular events among individuals with type 2 diabetes: a population-level cohort study. *Lancet Diabetes Endocrinol*. 2016;4(7):588-597. doi:10.1016/S2213-8587(16)30057-2. (In English)
- Pencina MJ, Goldstein BA, D'Agostino RB. Prediction models—development, evaluation, and clinical application. *NEJM*. 2020;382(17):1583-1586. doi:10.1056/NEJMp2000589.
- Huang J, Yeung AM, Armstrong DG, et al. Artificial intelligence for predicting and diagnosing complications of diabetes. *J Diabetes Sci Technol*. 2023;17(1):224-238. doi:10.1177/19322968221124583.
- Saputro SA, Pattanaprteep O, Pattanateepapon A, Karmacharya S, Thakkestian A. Prognostic models of diabetic microvascular complications: a systematic review and meta-analysis. *Syst Rev*. 2021;10(1):288. doi:10.1186/s13643-021-01841-z.
- Baethge C, Goldbeck-Wood S, Mertens S. SANRA—a scale for the quality assessment of narrative review articles. *Res Integr Peer Rev*. 2019;4:5. doi:10.1186/s41073-019-0064-8.
- Moons KG, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012;98(9):691-698. doi:10.1136/heartjnl-2011-301247.
- Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform*. 2019;95:103208. doi:10.1016/j.jbi.2019.103208.
- Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377-381. doi:10.1016/j.jbi.2008.08.010.
- Goldfarb- Rumyantzev AS, Pappas L. Prediction of renal insufficiency in Pima Indians with nephropathy of type 2 diabetes mellitus. *Am J Kidney Dis*. 2002;40(2):252-264. doi:10.1053/ajkd.2002.34503.
- Ng JS, Bearse MA Jr, Schneck ME, Barez S, Adams AJ. Local diabetic retinopathy prediction by multifocal ERG delays over 3 years. *Invest Ophthalmol Vis Sci*. 2008;49(4):1622-1628. doi:10.1167/iov.07-1157.
- Clarke PM, Gray AM, Briggs A, et al. A model to estimate the lifetime health outcomes of patients with type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68). *Diabetologia*. 2004;47(10):1747-1759. doi:10.1007/s00125-004-1527-z.
- Yang XL, So WY, Kong AP, et al. End-stage renal disease risk equations for Hong Kong Chinese patients with type 2 diabetes: Hong Kong Diabetes Registry. *Diabetologia*. 2006;49(10):2299-2308. doi:10.1007/s00125-006-0376-3.
- Afghahi H, Cederholm J, Eliasson B, et al. Risk factors for the development of albuminuria and renal impairment in type 2 diabetes—the Swedish National Diabetes Register (NDR). *Nephrol Dial Transplant*. 2011;26(4):1236-1243. doi:10.1093/ndt/gfq535.
- Semeraro F, Parrinello G, Cancarini A, et al. Predicting the risk of diabetic retinopathy in type 2 diabetic patients. *J Diabetes Complications*. 2011;25(5):292-297. doi:10.1016/j.jdia-comp.2010.12.002.
- Aspelund T, Thormorisdottir O, Olafsdottir E, et al. Individual risk assessment and information technology to optimise screening frequency for diabetic retinopathy. *Diabetologia*. 2011;54(10):2525-2532. doi:10.1007/s00125-011-2257-7.
- Jardine MJ, Hata J, Woodward M, et al. Prediction of kidney-related outcomes in patients with type 2 diabetes. *Am J Kidney Dis*. 2012;60(5):770-778. doi:10.1053/j.ajkd.2012.04.025.
- Tanaka S, Tanaka S, Iimuro S, et al. Predicting macro- and microvascular complications in type 2 diabetes: the Japan Diabetes Complications Study/the Japanese Elderly Diabetes Intervention Trial risk engine. *Diabetes Care*. 2013;36(5):1193-1199. doi:10.2337/dc12-0958.
- Hayes AJ, Leal J, Gray AM, Holman RR, Clarke PM. UKPDS outcomes model 2: a new version of a model to simulate lifetime health outcomes of patients with type 2 diabetes mellitus using data from the 30 year United Kingdom Prospective Diabetes Study: UKPDS 82. *Diabetologia*. 2013;56(9):1925-1933. doi:10.1007/s00125-013-2940-y.
- Elley CR, Robinson T, Moyes SA, et al. Derivation and validation of a renal risk score for people with type 2 diabetes. *Diabetes Care*. 2013;36(10):3113-3120. doi:10.2337/dc13-0190.
- Welsh P, Woodward M, Hillis GS, et al. Do cardiac biomarkers NT-proBNP and hsTnT predict microvascular events in patients with type 2 diabetes? results from the ADVANCE trial. *Diabetes Care*. 2014;37(8):2202-2210. doi:10.2337/dc13-2625.
- Riphagen IJ, Kleefstra N, Drion I, et al. Comparison of methods for renal risk prediction in patients with type 2 diabetes (ZODIAC-36). *PLoS ONE*. 2015;10(3):e0120477. doi:10.1371/journal.pone.0120477.
- Lagani V, Chiarugi F, Thomson S, et al. Development and validation of risk assessment models for diabetes-related complica-

- tions based on the DCCT/EDIC data. *J Diabetes Complications*. 2015;29(4):479-487. doi:10.1016/j.jdiacomp.2015.03.001.
24. Dunkler D, Gao P, Lee SF, et al. Risk prediction for early CKD in type 2 diabetes. *Clin J Am Soc Nephrol*. 2015;10(8):1371-1379. doi:10.2215/CJN.10321014.
 25. Scanlon PH, Aldington SJ, Leal J, et al. Development of a cost-effectiveness model for optimisation of the screening interval in diabetic retinopathy screening. *Health Technol Assess*. 2015;19(74):1-116. doi:10.3310/hta19740.
 26. Woodward M, Hirakawa Y, Kengne AP, et al. Prediction of 10-year vascular risk in patients with diabetes: the AD-ON risk score. *Diabetes Obes Metab*. 2016;18(3):289-294. doi:10.1111/dom.12614.
 27. Li HY, Lin HA, Nien FJ, et al. Serum vascular adhesion protein-1 predicts end-stage renal disease in patients with type 2 diabetes. *PLoS ONE*. 2016;11(2):e0147981. doi:10.1371/journal.pone.0147981.
 28. Parrinello CM, Matsushita K, Woodward M, Wagenknecht LE, Coresh J, Selvin E. Risk prediction of major complications in individuals with diabetes: the Atherosclerosis Risk in Communities Study. *Diabetes Obes Metab*. 2016;18(9):899-906. doi:10.1111/dom.12686.
 29. Low S, Lim SC, Zhang X, et al. Development and validation of a predictive model for chronic kidney disease progression in type 2 diabetes mellitus based on a 13-year study in Singapore. *Diabetes Res Clin Pract*. 2017;123:49-54. doi:10.1016/j.diabres.2016.11.008.
 30. Saulnier PJ, Gand E, Velho G, et al. Association of Circulating Biomarkers (Adrenomedullin, TNFR1, and NT-proBNP) with renal function decline in patients with type 2 diabetes: a French prospective cohort. *Diabetes Care*. 2017;40(3):367-374. doi:10.2337/dc16-1571.
 31. Miao DD, Pan EC, Zhang Q, Sun ZM, Qin Y, Wu M. Development and validation of a model for predicting diabetic nephropathy in Chinese people. *Biomed Environ Sci*. 2017;30(2):106-112. doi:10.3967/bes2017.014.
 32. Dagliati A, Marini S, Sacchi L, et al. Machine learning methods to predict diabetes complications. *J Diabetes Sci Technol*. 2018;12(2):295-302. doi:10.1177/1932296817706375.
 33. Jenks SJ, Conway BR, McLachlan S, et al. Cardiovascular disease biomarkers are associated with declining renal function in type 2 diabetes. *Diabetologia*. 2017;60(8):1400-1408. doi:10.1007/s00125-017-4297-0.
 34. Zobel EH, von Scholten BJ, Reinhard H, et al. Symmetric and asymmetric dimethylarginine as risk markers of cardiovascular disease, all-cause mortality and deterioration in kidney function in persons with type 2 diabetes and microalbuminuria. *Cardiovasc Diabetol*. 2017;16(1):88. doi:10.1186/s12933-017-0569-8.
 35. Wan EYF, Fong DYT, Fung CSC, et al. Prediction of new onset of end stage renal disease in Chinese patients with type 2 diabetes mellitus—a population-based retrospective cohort study. *BMC Nephrol*. 2017;18(1):257. doi:10.1186/s12882-017-0671-x.
 36. Basu S, Sussman JB, Berkowitz SA, Hayward RA, Yudkin JS. Development and validation of Risk Equations for Complications Of type 2 Diabetes (RECODe) using individual participant data from randomised trials. *Lancet Diabetes Endocrinol*. 2017;5(10):788-798. doi:10.1016/S2213-8587(17)30221-8.
 37. Eleuteri A, Fisher AC, Broadbent DM, et al. Individualised variable-interval risk-based screening for sight-threatening diabetic retinopathy: the Liverpool Risk Calculation Engine. *Diabetologia*. 2017;60(11):2174-2182. doi:10.1007/s00125-017-4386-0.
 38. Peters KE, Davis WA, Ito J, et al. Identification of novel circulating biomarkers predicting rapid decline in renal function in type 2 diabetes: the Fremantle Diabetes Study Phase II. *Diabetes Care*. 2017;40(11):1548-1555. doi:10.2337/dc17-0911.
 39. Lin CC, Li CI, Liu CS, et al. Development and validation of a risk prediction model for end-stage renal disease in patients with type 2 diabetes. *Sci Rep*. 2017;7(1):10177. doi:10.1038/s41598-017-09243-9.
 40. Kim JH, Oh SY, Kim EH, et al. Addition of nonalbumin proteinuria to albuminuria improves prediction of type 2 diabetic nephropathy progression. *Diabetol Metab Syndr*. 2017;9:68. doi:10.1186/s13098-017-0267-4.
 41. Nowak N, Skupien J, Smiles AM, et al. Markers of early progressive renal decline in type 2 diabetes suggest different implications for etiological studies and prognostic tests development. *Kidney Int*. 2018;93(5):1198-1206. doi:10.1016/j.kint.2017.11.024.
 42. García-Fiñana M, Hughes DM, Cheyne CP, et al. Personalized risk-based screening for diabetic retinopathy: a multivariate approach versus the use of stratification rules. *Diabetes Obes Metab*. 2019;21(3):560-568. doi:10.1111/dom.13552.
 43. 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):242-253. doi:10.1093/jamia/ocy165.
 44. 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):519-528. doi:10.1111/cts.12647.
 45. Ochs A, McGurnaghan S, Black MW, et al. Use of personalised risk-based screening schedules to optimise workload and sojourn time in screening programmes for diabetic retinopathy: a retrospective cohort study. *PLoS Med*. 2019;16(10):e1002945. doi:10.1371/journal.pmed.1002945.
 46. Wysham CH, Gauthier-Loiselle M, Bailey RA, et al. Development of risk models for major adverse chronic renal outcomes among patients with type 2 diabetes mellitus using insurance claims: a retrospective observational study. *Curr Med Res Opin*. 2020;36(2):219-227. doi:10.1080/03007995.2019.1682981.
 47. Peters KE, Davis WA, Ito J, Bringans SD, Lipscombe RJ, Davis TME. Validation of a protein biomarker test for predicting renal decline in type 2 diabetes: the Fremantle Diabetes Study Phase II. *J Diabetes Complications*. 2019;33(12):107406. doi:10.1016/j.jdiacomp.2019.07.003.
 48. Nelson RG, Grams ME, Ballew SH, et al. Development of risk prediction equations for incident chronic kidney disease. *JAMA*. 2019;322(21):2104-2114. doi:10.1001/jama.2019.17379.
 49. Jiang S, Fang J, Yu T, et al. Novel model predicts diabetic nephropathy in type 2 diabetes. *Am J Nephrol*. 2020;51(2):130-138. doi:10.1159/000505145.
 50. Aminian A, Zajichek A, Arterburn DE, et al. Predicting 10-year risk of end-organ complications of type 2 diabetes with and without metabolic surgery: a machine learning approach. *Diabetes Care*. 2020;43(4):852-859. doi:10.2337/dc19-2057.

51. Sun L, Shang J, Xiao J, Zhao Z. Development and validation of a predictive model for end-stage renal disease risk in patients with diabetic nephropathy confirmed by renal biopsy. *PeerJ*. 2020;8:e8499. doi:10.7717/peerj.8499.
52. Jiang W, Wang J, Shen X, et al. Establishment and validation of a risk prediction model for early diabetic kidney disease based on a systematic review and meta-analysis of 20 cohorts. *Diabetes Care*. 2020;43(4):925-933. doi:10.2337/dc19-1897.
53. Fernandez-Fernandez B, Mahillo I, Sanchez-Rodriguez J, et al. Gender, albuminuria and chronic kidney disease progression in treated diabetic kidney disease. *J Clin Med*. 2020;9(6):1611. doi:10.3390/jcm9061611.
54. Hu Y, Shi R, Mo R, Hu F. Nomogram for the prediction of diabetic nephropathy risk among patients with type 2 diabetes mellitus based on a questionnaire and biochemical indicators: a retrospective study. *Aging (Albany, NY)*. 2020;12(11):10317-10336. doi:10.18632/aging.103259.
55. Cheng Y, Shang J, Liu D, Xiao J, Zhao Z. Development and validation of a predictive model for the progression of diabetic kidney disease to kidney failure. *Ren Fail*. 2020;42(1):550-559. doi:10.1080/0886022X.2020.1772294.
56. Belur Nagaraj S, Pena MJ, Ju W, Heerspink HL, 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):2479-2486. doi:10.1111/dom.14178.
57. Dong W, Wan EYF, Fong DYT, et al. Prediction models and nomograms for 10-year risk of end-stage renal disease in Chinese type 2 diabetes mellitus patients in primary care. *Diabetes Obes Metab*. 2021;23(4):897-909. doi:10.1111/dom.14292.
58. Allen A, Iqbal Z, Green-Saxena A, et al. Prediction of diabetic kidney disease with machine learning algorithms, upon the initial diagnosis of type 2 diabetes mellitus. *BMJ Open Diabetes Res Care*. 2022;10(1):e002560. doi:10.1136/bmj-drc-2021-002560.
59. Gao YM, Feng ST, Yang Y, et al. Development and external validation of a nomogram and a risk table for prediction of type 2 diabetic kidney disease progression based on a retrospective cohort study in China. *Diabetes Metab Syndr Obes*. 2022;15:799-811. doi:10.2147/DMSO.S352154.
60. Dong Z, Wang Q, Ke Y, et al. Prediction of 3-year risk of diabetic kidney disease using machine learning based on electronic medical records. *J Transl Med*. 2022;20(1):143. doi:10.1186/s12967-022-03339-1.
61. Zou Y, Zhao L, Zhang J, et al. Development and internal validation of machine learning algorithms for end-stage renal disease risk prediction model of people with type 2 diabetes mellitus and diabetic kidney disease. *Ren Fail*. 2022;44(1):562-570. doi:10.1080/0886022X.2022.2056053.
62. Zhao Y, Li X, Li S, et al. Using machine learning techniques to develop risk prediction models for the risk of incident diabetic retinopathy among patients with type 2 diabetes mellitus: a cohort study. *Front Endocrinol (Lausanne)*. 2022;13:876559. doi:10.3389/fendo.2022.876559.
63. Wang S, Han J, Jung SY, et al. Development and implementation of patient-level prediction models of end-stage renal disease for type 2 diabetes patients using fast health-care interoperability resources. *Sci Rep*. 2022;12(1):11232. doi:10.1038/s41598-022-15036-6.
64. Inoguchi T, Okui T, Nojiri C, et al. A simplified prediction model for end-stage kidney disease in patients with diabetes. *Sci Rep*. 2022;12(1):12482. doi:10.1038/s41598-022-16451-5.
65. Nicolucci A, Romeo L, Bernardini M, et al. Prediction of complications of type 2 diabetes: a machine learning approach. *Diabetes Res Clin Pract*. 2022;190:110013. doi:10.1016/j.diabres.2022.110013.
66. Nugawela MD, Gurudas S, Prevost AT, et al. Development and validation of predictive risk models for sight threatening diabetic retinopathy in patients with type 2 diabetes to be applied as triage tools in resource limited settings. *EClinicalMedicine*. 2022;51:101578. doi:10.1016/j.eclinm.2022.101578.
67. Momenzadeh A, Shamsa A, Meyer JG. Bias or biology? importance of model interpretation in machine learning studies from electronic health records. *JAMIA Open*. 2022;5(3):ooac063. doi:10.1093/jamiaopen/ooac063.
68. Sun L, Wu Y, Hua RX, Zou LX. Prediction models for risk of diabetic kidney disease in Chinese patients with type 2 diabetes mellitus. *Ren Fail*. 2022;44(1):1454-1461. doi:10.1080/0886022X.2022.2113797.
69. Hosseini Sarkhosh SM, Hemmatabadi M, Esteghamati A. Development and validation of a risk score for diabetic kidney disease prediction in type 2 diabetes patients: a machine learning approach. *J Endocrinol Invest*. 2023;46(2):415-423. doi:10.1007/s40618-022-01919-y.
70. Østergaard HB, Read SH, Sattar N, et al. Development and validation of a lifetime risk model for kidney failure and treatment benefit in type 2 diabetes: 10-year and lifetime risk prediction models. *Clin J Am Soc Nephrol*. 2022;17(12):1783-1791. doi:10.2215/CJN.05020422.
71. Hosseini Sarkhosh SM, Esteghamati A, Hemmatabadi M, Daraei M. Predicting diabetic nephropathy in type 2 diabetic patients using machine learning algorithms. *J Diabetes Metab Disord*. 2022;21(2):1433-1441. doi:10.1007/s40200-022-01076-2.
72. Kanda E, Suzuki A, Makino M, et al. Machine learning models for prediction of HF and CKD development in early-stage type 2 diabetes patients. *Sci Rep*. 2022;12(1):20012. doi:10.1038/s41598-022-24562-2.
73. Schallmoser S, Zueger T, Kraus M, Saar-Tsechansky M, Stettler C, Feuerriegel S. Machine learning for predicting micro- and macrovascular complications in individuals with prediabetes or diabetes: retrospective cohort study. *J Med Internet Res*. 2023;25:e42181. doi:10.2196/42181.
74. Sim R, Chong CW, Loganadan NK, Adam NL, Hussein Z, Lee SWH. Comparison of a chronic kidney disease predictive model for type 2 diabetes mellitus in Malaysia using Cox regression versus machine learning approach. *Clin Kidney J*. 2023;16(3):549-559. doi:10.1093/ckj/sfac252.
75. Ou SM, Tsai MT, Lee KH, et al. Prediction of the risk of developing end-stage renal diseases in newly diagnosed type 2 diabetes mellitus using artificial intelligence algorithms. *BioData Min*. 2023;16(1):8. doi:10.1186/s13040-023-00324-2.
76. van der Heijden AA, Walraven I, van 't Riet E, et al. Validation of a model to estimate personalised screening frequency to monitor diabetic retinopathy. *Diabetologia*. 2014;57(7):1332-1338. doi:10.1007/s00125-014-3246-4.
77. Soto-Pedre E, Pinies JA, Hernaez-Ortega MC. External validation of a risk assessment model to adjust the frequency

- of eye-screening visits in patients with diabetes mellitus. *J Diabetes Complications*. 2015;29(4):508-511. doi:10.1016/j.jdiacomp.2014.12.020.
78. van der Heijden AA, Nijpels G, Badloe F, et al. Prediction models for development of retinopathy in people with type 2 diabetes: systematic review and external validation in a Dutch primary care setting. *Diabetologia*. 2020;63(6):1110-1119. doi:10.1007/s00125-020-05134-3. (In English)
 79. McEwan P, Bennett H, Ward T, Bergenheim K. Refitting of the UKPDS 68 risk equations to contemporary routine clinical practice data in the UK. *Pharmacoeconomics*. 2015;33(2):149-161. doi:10.1007/s40273-014-0225-z.
 80. Basu S, Sussman JB, Berkowitz SA, et al. Validation of risk equations for complications of type 2 diabetes (RECODE) using individual participant data from diverse longitudinal cohorts in the U.S. *Diabetes Care*. 2018;41(3):586-595. doi:10.2337/dc17-2002.
 81. Slieker RC, van der Heijden A, Siddiqui MK, et al. Performance of prediction models for nephropathy in people with type 2 diabetes: systematic review and external validation study. *BMJ*. 2021;374:n2134. doi:10.1136/bmj.n2134.
 82. Sun Z, Wang K, Miller JD, Yuan X, Lee YJ, Lou Q. External validation of the risk prediction model for early diabetic kidney disease in Taiwan population: a retrospective cohort study. *BMJ Open*. 2022;12(12):e059139. doi:10.1136/bmjopen-2021-059139.
 83. Kress S, Bramlage P, Holl RW, et al. Validation of a risk prediction model for early chronic kidney disease in patients with type 2 diabetes: data from the German/Austrian diabetes prospective follow-up registry. *Diabetes Obes Metab*. 2023;25(3):776-784. doi:10.1111/dom.14925.
 84. Hong N, Park H, Rhee Y. Machine learning applications in endocrinology and metabolism research: an overview. *Endocrinol Metab (Seoul)*. 2020;35(1):71-84. doi:10.3803/EnM.2020.35.1.71.
 85. Mader JK, Neubauer KM, Schaupp L, et al. Efficacy, usability and sequence of operations of a workflow-integrated algorithm for basal-bolus insulin therapy in hospitalized type 2 diabetes patients. *Diabetes Obes Metab*. 2014;16(2):137-146. doi:10.1111/dom.12186.
 86. Collins GS, Omar O, Shanyinde M, Yu LM. A systematic review finds prediction models for chronic kidney disease were poorly reported and often developed using inappropriate methods. *J Clin Epidemiol*. 2013;66(3):268-277. doi:10.1016/j.jclinepi.2012.06.020.
 87. ElSayed NA, Aleppo G, Aroda VR, et al. 1. Improving care and promoting health in populations: standards of care in diabetes-2023. *Diabetes Care*. 2023;46(suppl 1):S10-S18. doi:10.2337/dc23-S001.
 88. Arcadu F, Benmansour F, Maunz A, Willis J, Haskova Z, Prunotto M. Deep learning algorithm predicts diabetic retinopathy progression in individual patients. *NPJ Digit Med*. 2019;2:92. doi:10.1038/s41746-019-0172-3.
 89. Sabanayagam C, Banu R, Chee ML, et al. Incidence and progression of diabetic retinopathy: a systematic review. *Lancet Diabetes Endocrinol*. 2019;7(2):140-149. doi:10.1016/S2213-8587(18)30128-1.
 90. Romero-Brufau S, Huddleston JM, Escobar GJ, Liebow M. Why the C-statistic is not informative to evaluate early warning scores and what metrics to use. *Crit Care*. 2015;19(1):285. doi:10.1186/s13054-015-0999-1.
 91. Lu JH, Callahan A, Patel BS, et al. Assessment of adherence to reporting guidelines by commonly used clinical prediction models from a single vendor: a systematic review. *JAMA Netw Open*. 2022;5(8):e2227779. doi:10.1001/jamanetworkopen.2022.27779.
 92. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5-32. doi:10.1023/a:1010933404324.
 93. Liu J, Wu J, Liu S, Li M, Hu K, Li K. Predicting mortality of patients with acute kidney injury in the ICU using XGBoost model. *PLoS ONE*. 2021;16(2):e0246306. doi:10.1371/journal.pone.0246306.
 94. Mizutani D, Lethanh N, Adey BT, Kaito K. Improving the estimation of Markov transition probabilities using mechanistic-empirical models. *Front Built Environ*. 2017;3:58. doi:10.3389/fbuilt.2017.00058. (In English)
 95. Al Daoud E. Comparison between XGBoost, LightGBM and CatBoost using a home credit dataset. *Int J Inf Contr Comp Sci*. 2019;12(1):6-10. doi:10.5281/zenodo.3607805.
 96. Egger J, Gsaxner C, Pepe A, et al. Medical deep learning: a systematic meta-review. *Comput Methods Programs Biomed*. 2022;221:106874. doi:10.1016/j.cmpb.2022.106874. (In English)
 97. Abu Alfeilat HA, Hassanat ABA, Lasassmeh O, et al. Effects of distance measure choice on K-Nearest Neighbor classifier performance: a review. *Big Data*. 2019;7(4):221-248. doi:10.1089/big.2018.0175. (In English)
 98. Webb GI. Naïve Bayes. In: Sammut C, Webb GI, eds. *Encyclopedia of Machine Learning*. Boston, MA: Springer; 2010:713-714.
 99. Debray TP, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KG. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol*. 2015;68(3):279-289. doi:10.1016/j.jclinepi.2014.06.018. (In English)