A systematic literature review of predicting diabetic retinopathy, nephropathy and neuropathy in patients with type 1 diabetes using machine learning

Qingqing Xu[1], Liye Wang[2], Sujit S. Sansgiry[1]

[1]Department of Pharmaceutical Health Outcomes and Policy, College of Pharmacy, University of Houston, Houston, TX, USA; [2]Department of Pharmacological and Pharmaceutical Sciences, College of Pharmacy, University of Houston, Houston, TX, USA

Contributions: (I) Conception and design: All authors; (II) Administrative support: SS Sansgiry; (III) Provision of study material or patients: None; (IV) Collection and assembly of data: Q Xu; (V) Data analysis and interpretation: Q Xu, SS Sansgiry; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Sujit S. Sansgiry, MS, PhD. Professor, Health Building 2, 4849 Calhoun Road, Room 4050, Houston, TX 77204-5047, USA. Email: sansgiry@central.uh.edu.

Background: Diabetic retinopathy, nephropathy and neuropathy in patients with type 1 diabetes (T1D) are microvascular complications that can adversely impact disease prognosis and incur greater healthcare costs. Early identification of patients at risk of these microvascular complications using predictive models through machine learning (ML) can be helpful in T1D management. The objective of current review was to systematically identify and summarize published predictive models that used ML to assess the risk of diabetic nephropathy, retinopathy and neuropathy in T1D patients.

Methods: A targeted review of English literature was undertaken in PubMed (http://www.ncbi.nlm.nih.gov/pubmed) and Google Scholar (http://scholar.google.com/) from January 1, 2016 to May 31, 2019. Eligible articles were also identified from cross-references. Following concepts were used in combination to conduct the search queries: diabetes, retinopathy, nephropathy, neuropathy, microvascular complication, risk/predictive model, and ML/artificial intelligence/data mining.

Results: A total of 3,769 hits were found from all sources combined, duplicates were removed, titles and abstracts were screened, 61 studies underwent full-text review and a total of six studies met the eligibility criteria. Among them, four studies had developed risk models using data obtained from T1D patients alone, whereas two used data from both T1D and type 2 diabetes (T2D) patients. There was only one study that evaluated all three types of microvascular complications while the other five focused on one individual complication, i.e., either diabetic retinopathy, nephropathy or neuropathy. Only two studies evaluated time to developing a complication. The other four studies assessed complications as either binary (yes/no) or categorical (multiple levels). Prediction models were built using cross-sectional data from survey questionnaire (n=1, Iran) and longitudinal data (n=5) which were further classified as sources of electronic medical records (EMR) (n=3, US: 1, Europe: 2), clinical trial (n=1, US) and prospective study (n=1, Europe). Common predictors across studies as well as across types of microvascular complications included age, gender, diabetes duration, BMI, blood pressure, lipid level, and mean or a single HbA1C value. Commonly used ML algorithms included classification and regression tree (CART) and random forest (RF) (CART/RF, n=3), support vector machines (SVMs, n=2), logistic regression (LR, n=2) and neural networks (NNs, n=1). Model performance was evaluated using area under curve (AUC, n=4) and accuracy (n=2). Only half (n=3) of the included studies tested their developed models in an external dataset of patients with T1D.

Conclusions: Overall, very few studies reported predictive models for diabetic retinopathy, nephropathy and neuropathy using ML specifically for T1D patients. Future research that utilizes contemporary clinical data from T1D patients to predict the three types of microvascular complications is needed.

Keywords: Machine learning (ML); predictive model; diabetic retinopathy; nephropathy; neuropathy; type 1 diabetes (T1D)

Introduction

Type 1 diabetes (T1D) is a chronic disease caused by the autoimmune destruction of the insulin-producing beta cells in the islets of Langerhans (1). It affects around 23 million individuals worldwide (2,3) and accounts for about 5% of all types of diabetes mellitus (4). In the United States (US), over 1.5 million individuals are living with T1D and 40,000 incident cases are diagnosed annually (5). T1D is characterized by elevated blood glucose level, which is associated with pathological changes of the blood vessels in the eyes, kidneys, and nerves and can lead to microvascular complications including diabetic retinopathy, nephropathy, and neuropathy (1,6-8).

Diabetic retinopathy is the most common microvascular complication among the three microvascular complications (6). It is associated with over 10,000 new cases of blindness annually in the US (9) and is a risk factor for other microvascular and macrovascular complications of diabetes (10). It was estimated that diabetic patients with retinopathy were more likely to have four or more health care visits than those without retinopathy (11). A systematic literature review (SLR) in PubMed and MEDLINE indicated that diabetic macular edema can adversely impact patients' health-related quality of life (HRQoL) and incur significantly higher healthcare expenditures compared to diabetic patients without retinal complications (12).

Diabetic nephropathy or kidney disease is reflected through albuminuria and declining glomerular filtration rate (GFR) (13). Its prevalence in T1D patients is around 15–40% (14). Certain ethnic groups including South Asians, Hispanic Americans, and African Americans are more likely to develop macroalbuminuria and African Americans and South Asians are at a higher risk of progressing rapidly to more advanced stages of chronic kidney disease (CKD) (15). Diabetic nephropathy is the leading cause of end-stage renal disease (ESRD)/renal failure and associated with a higher risk of cardiovascular diseases (6,14). Having microalbuminuria or macroalbuminuria is associated with higher costs of between $3,580–$12,830 and significantly more healthcare resource utilization (HRU) compared with normo-albuminuria (16).

Last but not the least, diabetic neuropathy is a group of disorders that mainly affects peripheral nerves and can also damage autonomic nerves (17). Diabetic peripheral neuropathy (DPN) is characterized by numbness and/or burning and tingling pain in extremities, although up to 50% of patients can be symptomless (17). The prevalence of DPN in diabetic patients can be as high as over 30% (18). DPN increases a patient's risk of diabetic foot ulceration and lower extremity amputation, which represent major causes of morbidity and mortality in diabetic patients (19). The annual costs of DPN and its complications in T1D patients in the US were between $0.3 and $1.0 billion in 2001 (20). Depending on the severity of DPN, per-patient per-year (PPPY) direct medical costs for diabetic patients with DPN ranged between $12,492 and $30,755 in 2015, which were significantly higher than those patients with diabetes only ($6,632) (21). DPN can substantially impair patients' HRQoL and work productivity (18). On the other hand, although less common, diabetic autonomic neuropathy (DAN) can affect various organ systems and manifest as gastroparesis, constipation or diarrhea, bladder dysfunction, erectile impotence, and cardiovascular autonomic dysfunction (CAD) (22). CAD independently increases a patient's chances of developing silent ischemia and even sudden cardiac death (23). Among all diabetes-related complications, diabetic neuropathy was reported to be among the factors that have the greatest impact on patient-reported HRQoL, the other two are dialysis and stroke (24).

These three types of microvascular complications are often synergistic and if not well managed, can adversely impact disease prognosis and greatly drive healthcare costs (25,26). Treating T1D and its complications is expensive: approximately $15 billion were spent on T1D management in the US every year (27). Hence, the American Diabetes Association (AMA) clinical guidelines stress on the importance of early screening and prevention of diabetic microvascular complications (28). Predictive models can be a way to assist in early identification of patient at risk. Predictive models can output the likelihood of occurrence of certain health outcomes using existing patient data. Recently, machine learning (ML) has also been used for prediction (29-31). ML is a sub-domain of artificial intelligence. It includes a variety of algorithms that can enable data learning (32). Two SLRs revealed that

most existing prediction models in diabetes research were focusing more on longer term macrovascular outcomes such as cardiovascular diseases or mortality rather than microvascular complications (33,34). Moreover, the data used for prediction in these studies were from patients with type 2 diabetes (T2D) or a mixture of T2D (majority) and T1D (33,34). There is a gap in knowledge regarding prediction of microvascular complications specifically for T1D patients. Hence, this review aims to identify and summarize published predictive models using ML for diabetic nephropathy, retinopathy and neuropathy in T1D patients.

Methods

Based on a previous review of predictive models in management of diabetes and its complication that was published in 2016 (33), we conducted a targeted review of English literature in PubMed (http://www.ncbi.nlm.nih.gov/pubmed) and Google Scholar (http://scholar.google.com/) from Jan 1, 2016 to May 31, 2019. PubMed is the most-often used database with millions of publications in the field of medical research. Google scholar covers literature across diverse disciplines including computer science where ML originated. These two search engines were selected as the data source in order to retrieve publications from both fields of research, namely the health care and computer science. Articles were also identified from cross-references. Following concepts were used in combination to conduct the search queries: diabetes, retinopathy, nephropathy, neuropathy, microvascular complication, risk model and ML. Studies analyzing image data, not focusing on predictive models, not focusing on any microvascular complication, not specifying T1D patients, and letters, opinions or abstracts were excluded.

In PubMed, titles and abstracts were screened first to discard duplicate and irrelevant publications before full-text review was conducted. In Google Scholar, identified searches were sorted by relevance. Duplicates were removed and title and keywords were screened for inclusion. Publications passed the first-round screening would undergo full-text review. Characteristics of selected publications, including outcomes predicted, operational definition of outcomes, information related to the model development dataset and if applicable, the external test set such as data source, study design and setting, data collection period, and patient sample, prediction horizon, predictors, modeling methods, and model performance were summarized.

Results

A total of 3,769 hits were found from all sources combined (n=240 from PubMed, n=59 from cross-references and n=3,470 from Google Scholar). After removing duplicates, screening titles and abstracts and reviewing full-text articles, a total of six studies met the eligibility criteria and were included in this review (35-40). The detailed description of the publication selection process is presented in *Figure 1*.

Table 1 summarized general characteristics of the selected studies, including which outcome(s) were predicted, information of the model development data set, how missing values in the data were handled, which modelling methods were used and how models were evaluated. Among the selected studies, four had developed risk models using data obtained from T1D patients alone (35,37,39,40), whereas two used data from both T1D and T2D (majority) patients and incorporated type of diabetes (T1D or T2D) as a predictor (36,38). There was only one study that evaluated all three types of microvascular complications (37), while the other five studies focused on one individual complication, i.e., either diabetic retinopathy, nephropathy or neuropathy. Prediction models were built on both cross-sectional data from survey questionnaire (n=1, Iran) and longitudinal data from sources of electronic medical records (EMR) (n=3, US:1, Europe:2), clinical trial (n=1, US) and prospective study (n=1, Europe) that have an average follow-up between 4 and 7 years. Missing values of baseline patient characteristics were imputed in four studies.

Commonly used ML algorithms included classification and regression tree (CART) and random forest (RF, n=3), support vector machines (SVMs, n=2), logistic regression (LR, n=2) and neural networks (NNs, n=1). Both bootstrap and oversampling as well as cross-validation techniques were used for handling small sample size and imbalanced data (36,40). Furthermore, cross-validation was used in four studies for evaluating average model performance (35-38).

Table 2 provided more detailed information on the operational definition of each complication, predictors used, whether the model was tested in an external test set and if so, characteristics of the external test set, and best model performance in the development and external test data sets. Only two studies evaluated time to developing a complication (37,38), whereas the other four assessed complications as either binary (yes/no) or categorical (multiple levels). Common predictors across studies as well as across types of microvascular complications included age, gender, diabetes duration, BMI, blood pressure, lipid
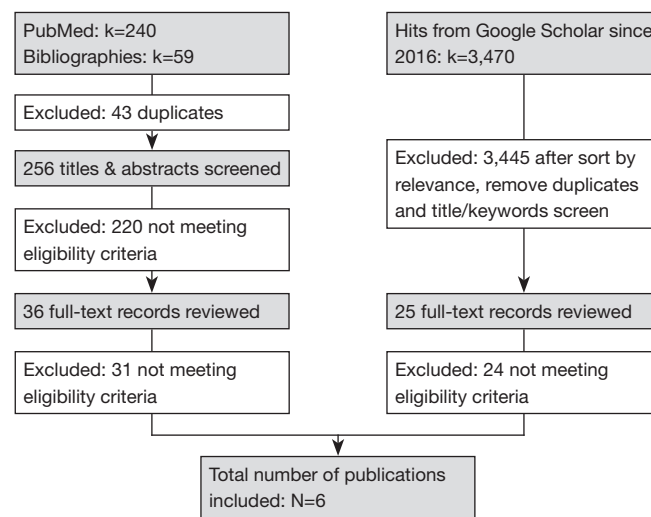
Figure 1 Flow chart of the selection process of predictive models for diabetic retinopathy, nephropathy and neuropathy in T1D patients using machine learning. T1D, type 1 diabetes.

level, and mean or a single HbA1C value. Only half (n=3) of the included studies tested their developed models in an external dataset of patients with T1D (37-39). Model performance was measured in terms of area under curve (AUC, or c-statistic, n=4) and accuracy (n=2). The average AUC ranged between 0.66–0.83.

Discussion

In this review, we found only six studies that have investigated ML models for predicting microvascular complications specifically in T1D patients, with the earliest one published in 2010 using a hybrid of ML algorithms. Only one study evaluated all three types of microvascular complications and the other five studies focused on one type of complication. Diabetic neuropathy was the least investigated (n=2). Because of the variation in diagnosis and definitions of each microvascular complication in different patient populations and data sources, it is hard to directly compare predictive models for the same microvascular complication across studies. The scarcity of predictive models for T1D patients can be partly due to lack of research interest in T1D, as it was not as prevalent as T2D. It can also be caused by a paucity of large contemporary longitudinal high-quality data dedicated to T1D patients for each microvascular complication. Furthermore, only three studies had tested developed models in an external dataset of T1D patients, which again suggests an inadequacy of data.

Among the three types of microvascular complications, diabetic retinopathy was predicted most frequently (n=3). However, each study has certain limitations. The study by Aspelund *et al*. illustrated the recommended screening intervals for T1D and T2D patients, respectively and demonstrated that T1D and T2D patients should have different eye screening intervals based on patient risks (38). The other model developed by Skevofilakas *et al*. was only internally validated through bootstrap and cross-validation based on a very small sample size (40). The exceptionally high accuracy (98%) reported in this development dataset would be hard to achieve given an external test set collected in a different patient population. Lastly, although the study by Lagani *et al*. was carefully designed and based on data from the largest clinical trial of T1D patients (37), it still had several opportunities for improvement. First, the Diabetes Control and Complications Trial (DCCT) only included patients who were relatively younger in age (between 13 and 39 years old) and had lived with T1D for a period of 1 to 15 years (41). Because these patients were under strict clinical monitoring during the trial period, they may have had better adherence to treatment and hence had a slower progression to adverse outcomes such as diabetic retinopathy compared to patients in real-world settings who were more likely to have poorer treatment adherence. Hence, the generalizability of this model to T1D patients not within the age range as well as patients in real-world is questionable. Second, the external test data had a rather small sample size (n=36) and missing

Table 1 Summary of reviewed studies addressing microvascular complication prediction in T1D patients and model development data set

Study	Outcome(s)	Study design	Model development data set						Modelling method(s)	Validation method
			Data source/setting	Data collection period	Prediction horizon (years)	Patient sample	Mean [SD] age (years)	Handling of missing values		
Ravizza et al. 2019 (35)	Nephropathy	Retrospective	EMR data extracted from the IBM Explorys database, US	Not reported	5	417,912 T1D & T2D patients	60 [10]	Not reported	Logistic regression	Bootstrapping
Kazemi et al. 2016 (36)	Neuropathy	Cross-sectional	Survey data collected at the Diabetes Center of Hamadan, Iran	2011	n/a	49 T1D & 551 T2D patients	53 [14]	n/a	MSVM	Synthetic minority class oversampling & 10-fold cross-validation
Lagani et al. 2015 (37)	Retinopathy, nephropathy, neuropathy	Prospective	DCCT clinical trial data and followed by EDIC data, US	1983–1989	6.5	1,391 T1D patients	27 [7]	Imputation by mean/mode values of respective predictors from the model development data set	• Cox regression • Ridge Cox regression • AFT models • RSF • SVCR	Nested-cross validation
Aspelund et al. 2011 (38)	Retinopathy	Retrospective	Clinical data from a large diabetic eye clinic, Iceland	1989–1994	4	175 T1D patients	32 [12]	Imputation by pre-determined single values	Weibull proportional hazards model	Not reported
Vergouwe et al. 2010 (39)	Nephropathy	Prospective	Data from the EURODIAB PCS from 31 centers in 16 European countries	1990–1991	7	1,115 T1D patients	33 [10]	Linear regression with single imputation	Logistic regression	Bootstrapping
Skevofilakas et al. 2010 (40)	Retinopathy	Retrospective	EMR data collected at the Hippokration Athens Hospital, Greece	Not reported	5	55 T1D patients	29 [13]	n/a	A hybrid voting algorithm combining FNN, CART and iHWNN	Bootstrapping, resampling and 10-fold cross-validation

EMR, electronic medical records; DCCT, The Diabetes Control and Complications Trial; EDIC, Epidemiology of Diabetes Interventions and Complications; EURODIAB PCS, European Diabetes Prospective Complications Study; T1D, type 1 diabetes; T2D, type 2 diabetes; MSVM, multicategory support vector machine; AFT, accelerated failure time; RSF, random survival forest; SVCR, SVM censored regression; FNN, feed forward neural network; CART, classification and regression tree; iHWNN, improved hybrid wavelet neural network.

Table 2 Summary of characteristics of predictive models by outcomes of diabetic retinopathy, nephropathy and neuropathy

| Study | Operational definition of outcome(s) | Predictors | External test | | | | | Model Performance | | |
			Y/N	Patient sample/mean [SD] age	Study design/setting	Follow-up/data collection period	Measure	Development set (95% CI)	External test set (95% CI)
Retinopathy									
Lagani et al. 2015 (37)	Time to event; any worsening in the retina condition that lasted at least six months	5 predictors: • HbA1C • post pubescent diabetes duration • BMI • Marital status • Retinopathy level	Y	36 T1D patients/40 [19] years	Retrospective cohort at a single location, UK	3–6 years/ 2004–2014	AUC	AFT model: 0.66 (0.61–0.67)	0.72 (0.59–0.87)
Aspelund et al. 2011 (38)	Time to event; diabetic macular edema and proliferative diabetic retinopathy	3 predictors: • HbA1C • SBP • Presence of NPDR	Y	1,275 T1D patients/not specified	Prospective cohort at a hospital, Denmark	Up to 13 years/ 1994–2007	AUC	Not reported	0.76 (0.74–0.78)
Skevofilakas et al. 2010 (40)	Binary (yes/no); not specified	7 predictors: • Age • Duration of T1D • Total cholesterol • Triglyceride level • Hypertension • Treatment duration	N	n/a	n/a	n/a	Accuracy iHWNN: 97%		n/a

Table 2 (continued)

Table 2 (*continued*)

Study	Operational definition of outcome(s)	Predictors	External test				Model Performance		
			Y/N	Patient sample/mean [SD] age	Study design/setting	Follow-up/data collection period	Measure	Development set (95% CI)	External test set (95% CI)
Nephropathy									
Ravizza et al. 2019 (35)	Binary (yes/no); diagnosis of CKD, hypertensive CKD, diabetes mellitus with other diabetic kidney complication, hypertensive heart and kidney disease or other relevant codes using ICD 9 or 10 codes	7 predictors: • HbA1C • Age • BMI • GFR • Creatinine level • Albumin level • Glucose level	N	n/a	n/a	n/a	AUC	0.79 (0.77–0.81)	n/a
Lagani et al. 2015 (37)	Time to event; microalbuminuria: ACR ≥2.5 mg/mmol for men or 3.5 mg/mmol for women or albumin concentration ≥20 mg/L	7 predictors: • HbA1C • Marital status • Weight • Diabetes duration • Albumin level • Insulin regimen • Retinopathy level	Y	36 T1D patients/40 [19] years	Retrospective cohort at a single location, UK	3–6 years/ 2004–2014	AUC	RSF: 0.74 (0.67–0.78)	0.82 (0.66–0.97)
	Time to event; proteinuria: ACR ≥30 mg/mmol or albumin concentration ≥200 mg/L	6 predictors: • HbA1C • Diabetes duration • Albumin level • Insulin regimen • Presence of neuropathy • HDL	N	n/a	n/a	n/a		Ridge Cox regression: 0.83 (0.54–0.96)	n/a

Table 2 (*continued*)

Table 2 (*continued*)

Study	Operational definition of outcome(s)	Predictors	External test — Y/N	Patient sample/mean [SD] age	Study design/setting	Follow-up/data collection period	Model Performance — Measure	Development set (95% CI)	External test set (95% CI)
Vergouwe et al. 2010 (39)	Binary (yes/no); microalbuminuria: an AER of 20–200 µg/min	5 predictors: • HbA1C • AER • Waist-to-hip ratio • BMI • Ever smoking	Y	• EDC: 190 T1D patients/26 [8] years • FinnDiane: 1,248 T1D patients/36 [12] years • CACTI: 331 T1D patients/37 [9] years	• EDC: retrospective cohort at a single location, USA • FinnDiane: a nationwide multicenter study, Finland • CACTI: prospective cohort, USA	7 years • EDC: 1986–1988; • FinnDiane: 1998 • CACTI: 2000–2002	AUC	0.76 (0.74–0.78)	• EDC: 0.71 • FinnDiane: 0.79 • CACTI: 0.79

Neuropathy

Study	Operational definition of outcome(s)	Predictors	External test — Y/N	Patient sample/mean [SD] age	Study design/setting	Follow-up/data collection period	Model Performance — Measure	Development set (95% CI)	External test set (95% CI)
Kazemi et al. 2016 (36)	4-level category; severity of DPN: categorized as no DPN, mild, moderate and severe DPN diagnosed by clinicians through examinations and two neuropathy diagnosis scales	13 predictors: • Glucose level • Age • T1D duration • Type of diabetes • Education level • BMI • Height • Weight • SBP • History of hypertension • History of foot ulcer • History of laser photocoagulation • Medications	N	n/a	n/a	n/a	Accuracy	76%	n/a

Table 2 (*continued*)

Table 2 (*continued*)

| Study | Operational definition of outcome(s) | External test | | | | Model Performance | | |
		Y/N	Patient sample/mean [SD] age	Study design/setting	Follow-up/data collection period	Measure	Development set (95% CI)	External test set (95% CI)
Lagani et al. 2015 (37)	Time to event: • Development set: presence of abnormality of nerve conductions or autonomic nerve system • External test set: dysfunction in bowel/bladder or erectile dysfunction • Post-pubescent diabetes duration	Y	37 T1D patients/40 [19] years	Retrospective cohort at a single location, UK	3–6/2004–2014	AUC	AFT model: 0.67 (0.55–0.74)	0.74 (0.55–0.91)

Predictors (from separate column):
5 predictors:
• HbA1C
• Albumin level
• Age
• Retinopathy level

CKD, chronic kidney disease; ICD, international classification of diseases; GFR, glomerular filtration rate; ACR, albumin/creatinine ratio; AER, albumin excretion ration; DPN, diabetic peripheral neuropathy; HbA1C, hemoglobin A1C; BMI, body mass index; SBP, systolic blood pressure; HDL, high-density lipoprotein; NPDR, non-proliferative diabetic retinopathy; T1D, type 1 diabetes; EDC, Epidemiology of Diabetes Complications; FinnDiane, Finnish diabetic nephropathy; CACTI, coronary artery calcification in type 1 diabetes; SD, standard deviation; CI, confidence interval; AUC, area under curve; AFT, accelerated failure time; RSF, random survival forest; iHWNN, improved hybrid wavelet neural network.

values were imputed by data obtained from the DCCT data itself. This might explain why the average AUC in the external data set (0.72) was higher than in the development dataset (0.66). Had the model been tested in another larger external dataset with different baseline patient attributes, the model's external performance would likely be lower than 0.66 and far from optimal. Third, the DCCT data was collected in the 1980s and can be obsolete as it was before many of newer treatment modalities for T1D patients such as newer generations of insulin analogs and continuous glucose monitors became available (42). Predictive models using more recent data are more likely to reflect T1D progression in patients receiving contemporary treatments. These drawbacks also applied to the prediction of diabetic nephropathy and neuropathy in this study as well.

Though a total of three studies assessed nephropathy in T1D patients, in the Ravizza *et al.* study the outcome was not specific to diabetic nephropathy, but also comprised other diagnoses such as hypertensive CKD and hypertension heart and kidney disease that may be caused by conditions other than T1D (35). Hence, when using their model for prediction, clinicians should take caution in interpreting the resulted risks to patients. In addition, even though this study utilized large real-world data extracted from IBM Explorys database (n=417,912), the specific number of T1D patients was not disclosed. Moreover, only logistic regression was used in this study while other newer and more advanced ML algorithms such as NNs were not attempted. The study by Vergouwe *et al.* also used logistic regression only (39). One main constraint of statistical methods such as logistic regression lies in their restricted ability in handling correlated data. Future studies predicting diabetic nephropathy may consider using algorithms like NNs that can incorporate correlated predictors and see whether it would improve prediction.

Diabetic neuropathy was least evaluated (n=2), with one predicted DPN and the other predicted DAN. Neuropathy was defined as bowel/bladder or erectile dysfunction in the test set of the study led by Lagani *et al.* (37). Whether erectile dysfunction is caused by diabetic neuropathy or other peripheral vascular conditions of a patient is arguable. Moreover, this definition is different from the definition of neuropathy in the development set. Hence, cautions need to be taken when interpreting their external model performance relative to the performance in the development set. The study by Kazemi *et al.* reported an acceptable accuracy of 76% (36). However, the data used for model building in this study was from a cross-sectional survey

collected in a single location in Iran, within which only 49 patients with T1D. To evaluate the progression of T1D, patients will usually need to be followed for at least several years to observe incidences of adverse outcomes such as any microvascular complication. Hence, considering either the representativeness of the patient sample in the study or the necessary follow-up period of patients, the predictive model by Kazemi *et al.* was unlikely to be useful for future application to other patient populations.

Candidate predictors across complications and across studies were selected based on literature review and clinical expertise. The final models mostly included from 3 to 7 predictors considering parsimony and applicability. Common risk factors include older age, certain races, longer duration of T1D, dyslipidemia, hypertension, overweight and obesity, smoking, and physical inactivity (28,43). Retinopathy may put patients at higher risk of developing the other two types of microvascular complications (28). For kidney diseases, the level of albumin excretion rate (AER) is an additional strong predictor (39). Similarly, past or current ulcer is a specific risk parameter for neuropathy (44). On the other hand, the use of ACE inhibitors was reported to reduce the chance of progression to microvascular complications in T1D patients (44). It's worth noting that researchers found that variability of HbA1C (or long-term variability) was adversely associated with both micro- and macro- vascular complications and mortality independently of mean HbA1C value (45,46). However, none of the predictive models have taken variability of HbA1C into account. As new ML algorithms do not assume independence of predictors, they can easily fit HbA1C variability into the predictive models. Future research is needed to apply HbA1C variability in clinical risk assessment.

Methodology-wise, only two studies by Lagani *et al.* and Skevofilakas *et al.* attempted more than one method to develop the predictive model (37,40). Other four studies only tried one modeling method and within them, three still resorted to conventional statistical methods such as logistic regression and Cox regression (36,39,40). The overall performances of all models were moderate, with AUCs below 0.8 and accuracy around 80% (except for the hybrid model by Skevofilakas *et al.* with an accuracy of 98%). However, we do notice some intriguing findings. It was counter-intuitive that models performed better in the external test dataset compared to in the model development dataset as in the study of Lagani and colleagues (37). This could be due to several reasons: first, the sample size of

the test dataset was so small that model performance was evaluated by nested-cross validation using this small sample size; second, missing values were replaced by attributes calculated from the development dataset. Hence, the performance in the external test set may have been over-estimated.

Conclusions

Early prediction of diabetic retinopathy, nephropathy and neuropathy specifically for T1D patients is important for risk stratification and T1D management. We found limited studies that developed predictive models using ML to assess diabetic retinopathy, nephropathy and neuropathy specifically for T1D patients. Definition of each microvascular complication varied between studies. Hence the output of patient risk from each predictive model should be interpreted carefully by clinicians. A standardized way to measure and operationalize each microvascular complication is needed to facilitate application of risk models in clinical settings. More research is needed to predict each microvascular complication using contemporary real-world data of T1D patients as well as more advanced ML algorithms such as NNs. Predictors such as variability of HbA1C should also be incorporated into predictive models.

Acknowledgments

Footnote

Conflict of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/jmai.2019.10.04). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Chiang JL, Kirkman MS, Laffel LMB, Peters AL. Type 1 diabetes through the life span: A position statement of the American Diabetes Association. Diabetes Care 2014;37:2034-54.
2. Cho NH, Shaw JE, Karuranga S, et al. IDF diabetes atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes Res Clin Pract 2018;138:271-81.
3. Global report on diabetes. World Health Organization. Geneva, Switzerland: WHO Library Cataloguing-in-Publication Data; 2016. Available online: https://www.who.int/diabetes/global-report/en/
4. About diabetes. Centers for Disease Control and Prevention (CDC). Available online: https://www.cdc.gov/diabetes/basics/diabetes.html. Accessed 06/16/2018.
5. Type 1 diabetes. American Diabetes Association. Available online: http://www.diabetes.org/diabetes-basics/type-1/?loc=util-header_type1,%20accessed%2001/29/2019. Accessed 09/18/2019
6. Fowler MJ. Microvascular and macrovascular complications of diabetes. Clinical Diabetes. 2008;26:77-82.
7. Todd JA. Etiology of type 1 diabetes. Immunity 2010;32:457-67.
8. Bluestone JA, Herold K, Eisenbarth G. Genetics, pathogenesis and clinical interventions in type 1 diabetes. Nature 2010;464:1293-300.
9. Fong DS, Aiello LP, Ferris FL, Klein R. Diabetic retinopathy. Diabetes Care 2004;27:2540-53.
10. Pearce I, Simó R, Lövestam-Adrian M, et al. Association between diabetic eye disease and other complications of diabetes: Implications for care. A systematic review. Diabetes Obes Metab 2019;21:467-78.
11. Candrilli SD, Davis KL, Kan HJ, et al. Prevalence and the associated burden of illness of symptoms of diabetic peripheral neuropathy and diabetic retinopathy. J Diabetes Complications 2007;21:306-14.
12. Chen E, Looman M, Laouri M, et al. Burden of illness of diabetic macular edema: Literature review. Curr Med Res Opin 2010;26:1587-97.
13. Bjornstad P, Cherney D, Maahs DM. Early diabetic nephropathy in type 1 diabetes: New insights. Curr Opin

Endocrinol Diabetes Obes 2014;21:279-86.

14. Viswanathan V. Preventing microvascular complications in type 1 diabetes mellitus. Indian J Endocrinol Metab 2015;19:S36-8.

15. Ameh OI, Okpechi IG, Agyemang C, et al. Global, regional, and ethnic differences in diabetic nephropathy. In: Roelofs JJ, Vogt L, editors. Diabetic nephropathy: Pathophysiology and clinical aspects. Springer, Cham, 2019:33-44.

16. Zhou Z, Chaudhari P, Yang H, et al. Healthcare resource use, costs, and disease progression associated with diabetic nephropathy in adults with type 2 diabetes: A retrospective observational study. Diabetes Ther 2017;8:555-71.

17. American Diabetes Association. 10. Microvascular Complications and Foot Care: Standards of Medical Care in Diabetes-2018. Diabetes Care 2018;41:S105-18.

18. Alleman CJ, Westerhout KY, Hensen M, et al. Humanistic and economic burden of painful diabetic peripheral neuropathy in Europe: A review of the literature. Diabetes Res Clin Pract 2015;109:215-25.

19. Pop-Busui R, Boulton AJ, Feldman EL, et al. Diabetic neuropathy: A position statement by the American Diabetes Association. Diabetes Care 2017;40:136-54.

20. Gordois A, Scuffham P, Shearer A, et al. The health care costs of diabetic peripheral neuropathy in the U.S. Diabetes Care 2003;26:1790-95.

21. Sadosky A, Mardekian J, Parsons B, et al. Healthcare utilization and costs in diabetes relative to the clinical spectrum of painful diabetic peripheral neuropathy. J Diabetes Complications 2015;29:212-7.

22. Boulton AJM, Vinik AI, Arezzo JC, et al. Diabetic neuropathies. Diabetes Care 2005;28:956-62.

23. Maser RE, Mitchell BD, Vinik AI, et al. The association between cardiovascular autonomic neuropathy and mortality in individuals with diabetes. Diabetes Care 2003;26:1895-901.

24. Zhang P, Brown MB, Bilik D, et al. Health utility scores for people with type 2 diabetes in U.S. managed care health plans: Results from Translating Research Into Action for Diabetes (TRIAD). Diabetes Care 2012;35:2250-6.

25. Atkinson MA, Eisenbarth GS, Michels AW. Type 1 diabetes. Lancet 2014;383:69-82.

26. Kähm K, Laxy M, Schneider U, Holle R. Exploring different strategies of assessing the economic impact of multiple diabetes-associated complications and their interactions: A large claims-based study in Germany. Pharmacoeconomics 2019;37:63-74.

27. Tao B, Pietropaolo M, Atkinson M, et al. Estimating the cost of type 1 diabetes in the U.S.: a propensity score matching method. PLoS One 2010;5:e11501.

28. 11. American Diabetes Association. 11. Microvascular Complications and Foot Care: Standards of Medical Care in Diabetes-2019. Diabetes Care 2019;42:S124-38.

29. Dagliati A, Marini S, Sacchi L, et al. Machine learning methods to predict diabetes complications. J Diabetes Sci Technol 2018;12:295-302.

30. Kavakiotis I, Tsave O, Salifoglou A, et al. Machine learning and data mining methods in diabetes research. Comput Struct Biotechnol J 2017;15:104-16.

31. Contreras I, Vehi J. Artificial intelligence for diabetes management and decision support: Literature review. J Med Internet Res 2018;20:e10775.

32. Geron A. Hands-on machine learning with Scikit-Learn and TensorFlow: Concepts, tools, and techniques to build intelligent systems. O'Reilly, Sebastopol, CA, First Edition 2017.

33. Cichosz SL, Johansen MD, Hejlesen O. Toward big data analytics: Review of predictive models in management of diabetes and its complications. J Diabetes Sci Technol 2015;10:27-34.

34. Lagani V, Koumakis L, Chiarugi F, et al. A systematic review of predictive risk models for diabetes complications based on large scale clinical studies. J Diabetes Complications 2013;27:407-13.

35. Ravizza S, Huschto T, Adamov A, et al. Predicting the early risk of chronic kidney disease in patients with diabetes using real-world data. Nature Medicine 2019;25:57-9.

36. Kazemi M, Moghimbeigi A, Kiani J, et al. Diabetic peripheral neuropathy class prediction by multicategory support vector machine model: a cross-sectional study. Epidemiol Health 2016;38:e2016011.

37. Lagani V, Chiarugi F, Thomson S, et al. Development and validation of risk assessment models for diabetes-related complications based on the DCCT/EDIC data. J Diabetes Complications 2015;29:479-87.

38. Aspelund T, Thornórisdóttir O, Olafsdottir E, et al. Individual risk assessment and information technology to optimise screening frequency for diabetic retinopathy. Diabetologia 2011;54:2525-32.

39. Vergouwe Y, Soedamah-Muthu SS, Zgibor J, et al. Progression to microalbuminuria in type 1 diabetes: Development and validation of a prediction rule. Diabetologia 2010;53:254-62.

40. Skevofilakas M, Zarkogianni K, Karamanos BG, et al. A hybrid decision support system for the risk assessment of

retinopathy development as a long term complication of type 1 diabetes mellitus. Conf Proc IEEE Eng Med Biol Soc 2010;2010:6713-16.

41. The Diabetes Control and Complications Trial Research Group. The relationship of glycemic exposure (HbA1c) to the risk of development and progression of retinopathy in the diabetes control and complications trial. Diabetes 1995;44:968-83.

42. Aathira R, Jain V. Advances in management of type 1 diabetes mellitus. World J Diabetes 2014;5:689-96.

43. Risk factors for complications. Centers for Disease Control and Prevention (CDC). Available online: https://www.cdc. gov/diabetes/data/statistics-report/risks-complications. html. Accessed 09/18/2019.

44. Donnelly R, Emslie-Smith AM, Gardner ID, et al. Vascular complications of diabetes. BMJ 2000;320:1062-66.

45. Gorst C, Kwok CS, Aslam S, et al. Long-term glycemic variability and risk of adverse outcomes: A systematic review and meta-analysis. Diabetes Care 2015;38:2354-69.

46. Nalysnyk L, Hernandez-Medina M, Krishnarajah G. Glycaemic variability and complications in patients with diabetes mellitus: Evidence from a systematic review of the literature. Diabetes Obes Metab 2010;12:288-98.