Review: Machine learning in precision pharmacotherapy of type 2 diabetes—A promising future or a glimpse of hope?

Xiantong Zou* (iD), Yingning Liu* and Linong Ji

Abstract

Precision pharmacotherapy of diabetes requires judicious selection of the optimal therapeutic agent for individual patients. Artificial intelligence (AI), a swiftly expanding discipline, holds substantial potential to transform current practices in diabetes diagnosis and management. This manuscript provides a comprehensive review of contemporary research investigating drug responses in patient subgroups, stratified via either supervised or unsupervised machine learning approaches. The prevalent algorithmic workflow for investigating drug responses using machine learning involves cohort selection, data processing, predictor selection, development and validation of machine learning methods, subgroup allocation, and subsequent analysis of drug response. Despite the promising feature, current research does not yet provide sufficient evidence to implement machine learning algorithms into routine clinical practice, due to a lack of simplicity, validation, or demonstrated efficacy. Nevertheless, we anticipate that the evolving evidence base will increasingly substantiate the role of machine learning in molding precision pharmacotherapy for diabetes.

Introduction

Diabetes is a highly heterogeneous disease. The rationale of precision medicine is to find the right therapy for the right patient at the right time. The concept of implementing individualized therapy in diabetes patients is not novel; for instance, insulin therapy has been considered based on patients' endogenous insulin secretion levels since approximately half a century ago. The main treatment for patients with obvious insulin deficiency, such as those with type 1 diabetes, is exogenous insulin administration. Recent advances in disease etiology and mechanism, encompassing big data, biomarkers, genetics, epigenetics, high-throughput sequencing, proteomics, metabolomics, and gut microbiota, have catalyzed a paradigm shift in diabetes management. In 2020, the European Association for the Study of Diabetes (EASD) and American Diabetes Association (ADA) published their consensus report on precision medicine in diabetes, subdivided into components such as diagnosis, precision therapeutic, precision prevention, precision treatment, precision prognostics, and precision monitoring.[1] A number of studies have been conducted to bridge the evidence gap in the clinical application of precision medicine in diabetes care. Notably, precision therapeutics focuses on refining the classification of patients and selecting the most suitable diabetes management regimes. A variety of technologies have been employed to this end, among which artificial intelligence (AI) has garnered substantial attention.

Department of Endocrinology and Metabolism, Peking University People's Hospital, Beijing, China

*Equal contribution

Corresponding authors:
Xiantong Zou, Department of Endocrinology and Metabolism, Peking University People's Hospital, Beijing, 100044, China.
Email: eva2172@163.com

Linong Ji, Department of Endocrinology and Metabolism, Peking University People's Hospital, Beijing, 100044, China.
Email: jiln@bjmu.edu.cn

AI is being heralded as the catalyst for the fourth industrial revolution. Machine learning, a subset of AI, is utilized in the creation of automated systems that learn from experience. The basic process of machine learning involves learning and application.[2] Its commercial success in areas such as computer vision, speech recognition, and natural language processing has stimulated the application of machine learning to many other fields. Within healthcare, machine learning has become a focal point for physicians and clinicians aiming to automate and streamline medical procedures.[3] Machine learning has the potential to enhance predictive accuracy compared to traditional methods using identical variables and cohorts,[4] thus providing a correct estimation of diabetes incidence and progression. However, robust evidence is necessary before clinicians can confidently adopt these techniques in making clinical decisions, especially on individualized drug treatment regimes. The objective of this review is to evaluate whether the current evidence sufficiently supports the integration of machine learning to reshape precision pharmacotherapy for diabetes.

Machine learning

Machine learning is roughly divided into supervised learning and unsupervised learning.[5] Using supervised learning, a model is trained by learning the characteristics related to labeled outcomes, and unknown outcomes can be predicted using the trained model. Specifically, a classification algorithm can be used to predict categorized outcomes, while a regression algorithm can be adopted to predict continuous outcomes. Typical supervised machine learning algorithms include linear regression, random forest, gradient boosting, support vector machines, and artificial neural networks (ANN).

Unlike supervised learning, unsupervised learning does not have a predetermined outcome. The models divided the data automatically according to their similarity in density, structure, distance, or other features. Clustering is the most commonly used unsupervised learning method, including K means or K medium clustering, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), and deep belief networks (Figure 1)

Recently, deep learning, as an advanced form of machine learning was applied to medical studies. Using deep learning, researchers can solve difficult problems that shallow architectures were unable to address due to their dimensionality limitations. In deep learning, multiple layers are trained unsupervised, and then all layers are fine-tuned under supervision. This allows the discovery of robust features and precision prediction of outcomes.[6]

Machine learning and its application in diabetes

Machine learning algorithms could be applied to precision medicine in diabetes in many ways. Supervised learning can be trained to predict a specific outcome, e.g., incident diabetes,[7–9] glycemic control,[10] hypoglycemia,[11–13] development of complications,[14,15] and the glucose-lowering effect of an intervention.[16] Unsupervised machine learning
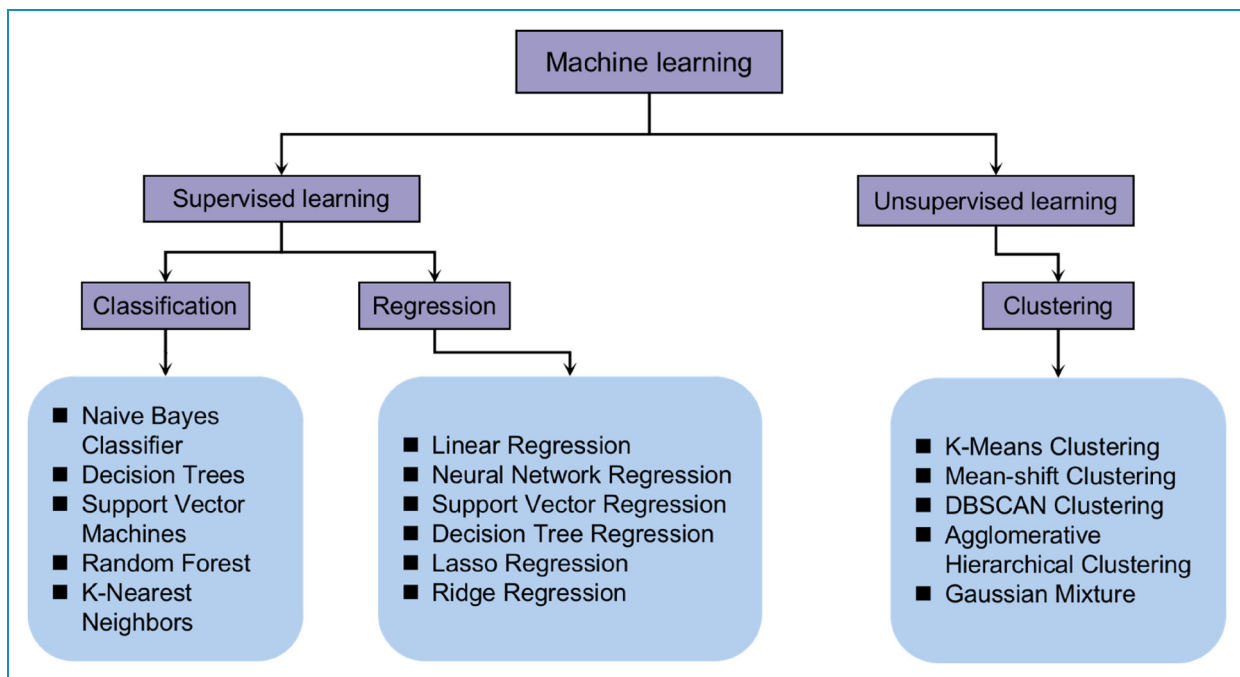

Figure 1. Machine learning algorithms (adapted from Alpaydin et al.[5]).

methods were widely used to categorize and stratify patients. For example, K means data-driven clusters divided diabetes into five different categories and each subgroup had a distinct glucose trajectory and complication development.[17] Alternatively, Bayesian nonnegative matrix factorization (bNMF) clustering was used to identify five clusters of type 2 diabetes mellitus (T2DM), with each cluster displaying differently in clinical outcomes including coronary artery disease (CAD) and stroke.[18] Data-driven clustering can also be used to define subgroups with different cardiovascular risks in participants with T2DM with established atherosclerotic cardiovascular disease (ASCVD).[19] Because of its high accuracy, deep learning has been applied to guide the insulin pump dosage system in type 1 diabetes mellitus (T1DM)[20] and guide insulin dosage and glycemic response in T2DM,[21] which has been validated in clinical trials and implemented in clinics.[22] Additionally, deep learning was used to read fundus photos not only for diabetes retinopathy[23,24] but also for diabetes kidney disease.[25]

Machine learning and deep learning algorithms have facilitated the process of protein structure analysis and design of novel antidiabetic drugs[26,27] or screening for chemicals of novel drug development targets.[28] In addition, novel biomarkers of diabetes and other metabolic diseases are being identified by machine learning and deep learning. These biomarkers also involve multi-omic signatures, e.g., functional connectome on magnetic resonance imaging (MRI) image,[29,30] metabolomics,[31] and epigenetics and genetics.[32] The application of deep learning in diabetes-related tasks was properly summarized and reviewed.[33–35]

Predicting diabetes and its cardiovascular risks using machine learning

Numerous studies have utilized machine learning to predict the incidence or presence of diabetes, largely due to the high diagnostic accuracy of these models. Decision trees, logistic regression, and random forest were commonly employed algorithms for diabetes prediction. Two meta-analyses suggested the average receiver operating characteristic area under the curve (ROCAUC) of these models to be between 0.81 (95% confidence interval (CI) of 0.79 to 0.83) and 0.86 (0.82 to 0.89).[36,37] Predictive variables incorporate a range of clinical anthropometric measurements, such as age, gender, and body mass index (BMI), laboratory test results, lifestyle factors, and high-dimensional variables like physical activity tracker data,[38] electrocardiograms (ECGs),[39] and chest radiograph.[40] Deep learning typically performs well when high-dimensional variables are included.[40,41] The number of machine learning-based diabetes prediction models is steadily increasing since chatbot-based AI tools now permit clinicians to generate models using various attributes via a simple user interface.[42]

A central topic in diabetes management is the micro- and macro-vascular complications of diabetes patients. Cardiovascular disease remains the primary cause of mortality in this population, yet robust tools for estimating cardiovascular risks are lacking. General cardiovascular risk estimation models, e.g., the Framingham score, may be not applicable to participants with diabetes.[43] Current conventional cardiovascular risk scoring systems, such as the Action in Diabetes and Vascular Disease: Preterax and Diamicron-MR Controlled Evaluation (ADVANCE)[44] and SCORE2-Diabetes,[45] performed well within their development cohorts; however, their external validity is less satisfactory[46] or not yet tested in the global area.[45] Machine learning algorithms can be potentially a robust tool to estimate cardiovascular complications. A recent systemic review demonstrated that the ROC AUC for derivation cohorts varied from 0.69 to 0.77. AI models achieved better performance than conventional models in some specific scenarios (ROC AUC 0.75 for AI models and 0.69 for conventional risk scores). However, only one out of the 176 AI models underwent an external validation study.[46] Further studies are warranted to enhance the predictive accuracy of these models and expand the external validation. This will facilitate the implementation of machine learning-based algorithms in clinical settings.

Evaluating machine learning methods in diabetes pharmacotherapy

Even if machine learning-based algorithms had achieved high performance in diabetes and relative complication estimation, there was a missing link in a very critical question: can machine learning shape current strategies for pharmacotherapy in diabetes patients?

Research efforts geared towards seeking empirical support for the application of machine learning in diabetes treatment predominantly follow two distinct strategies, as delineated in Figure 2. Typically, cohort data are processed, and predicting variables are selected. For supervised learning, a specific endpoint was chosen, and the cohort was divided into development cohort and internal validation cohorts. Ideally, an external cohort should be used to test the model performance. Subgroups of patients with different endpoint risks can be stratified using the model. For unsupervised data-driven machine learning, data were automatically subdivided into groups, and clusters' characteristics, disease trajectory, and drug responses were analyzed. Given the absence of internal validation for unsupervised machine learning approaches, the external validation of identified subgroups assures heightened significance. The researchers evaluated the drug response in these subgroups generated by either supervised machine learning or unsupervised machine learning by assessing the treatment-by-group interaction. (Figure 2) The main studies evaluating the drug
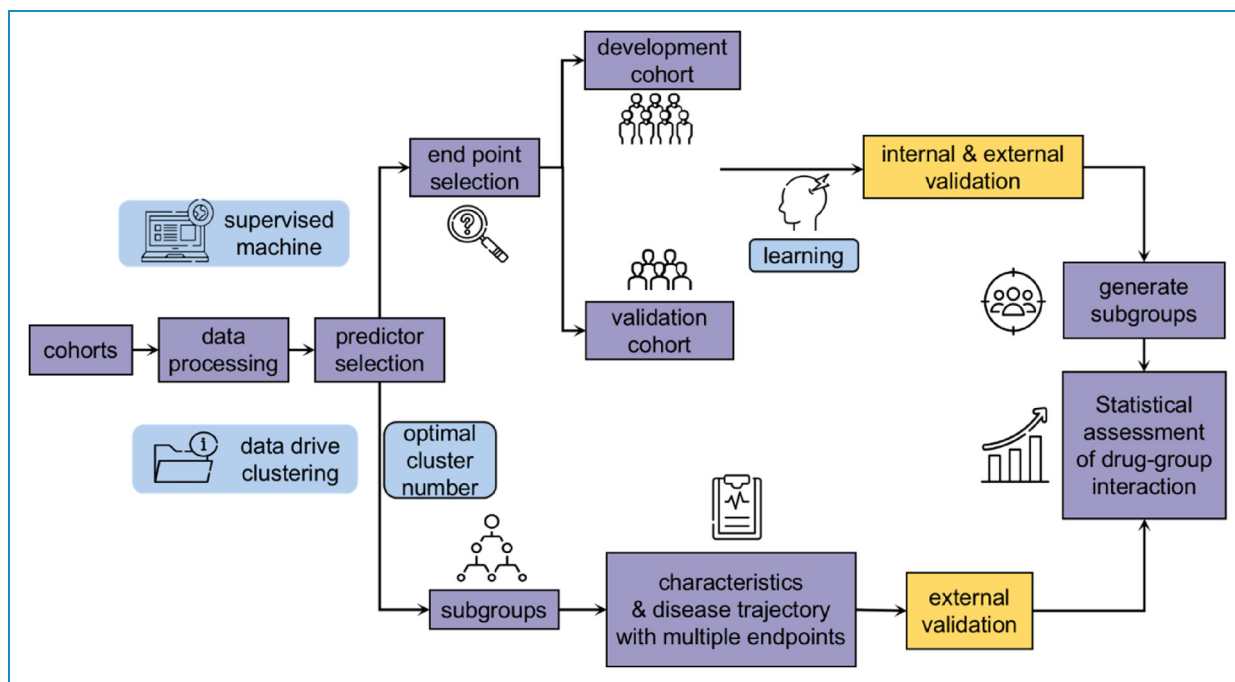
Figure 2. Algorithm of machine learning-based algorithms to predict drug responses.

responses in subgroups derived using unsupervised learning algorithms and supervised learning algorithms were summarized in Tables 1 and 2, respectively.

Cohorts

There are no specific requirements for cohort sample size; conventionally, the preference leans towards larger sample sizes. To validate a clinical outcome, most studies utilized prospective or retrospective cohorts rather than cross-sectional studies. The follow-up periods in these studies range widely from as short as 7 days to as long as 20 years, depending on the selected endpoints. The types of studies involve registry-based studies,[17] hospital-based cohorts,[54] epidemiological surveys,[69] electronic health records (EHRs), medical insurance dataset,[68] and clinical trials.[50,56]

Data processing

Raw data from cohorts can be voluminous and unstructured, so data preprocessing, including cleaning, normalization, and standardization of these heterogeneous data, is essential. Usually, normalization and standardization of the data are necessary to fit the data for machine learning algorithms or statistical testing. Additionally, dimensionality reduction techniques such as principal component analysis (PCA) or *t*-distributed stochastic neighbor embedding (*t*-SNE) can be employed to handle high-dimensional data.[70]

Dealing with missing data is an imperative issue in machine learning analysis. Missing data can be handled using simple deletion, multiple imputation, full information maximum likelihood, and expectation-maximization algorithm[71] or within the machine learning algorithm, e.g., decision trees.[72] Usually, multiple imputation is only applied if less than 30% of the variables are missing.[69] The nature of missing data is important for choosing methods to handle missing data, since whether the data are missing partly/completely or not at fandom could affect the model. Usually, sensitivity analysis regarding various data processing techniques is required[47] for model development and validation.

Selecting predictors

Machine learning and deep learning allow clinicians to use multi-nominal data as inputs. In spite of traditional clinical data including medical history, physical examination, and laboratory measurements, metabolites,[54] fundus photos,[25] radiographic images,[40] continuous glucose monitoring (CGM) data[73] and genetic information[74] can be candidate predictors. Regardless of the architecture of machine learning models, the selection of predictors is of critical importance in model development. The accuracy of the model hinges largely on the strength of the association between these predictors and the outcomes. Some studies manually chose common clinical variables, such as age, BMI, hemoglobin A1c (HbA1c), homeostasis model assessment (HOMA2IR), and HOMA2B.[17] Most studies initially choose as many parameters as they can and then select suitable predictors using LASSO regression algorithms[9] or

Table 1. Different responses to pharmacotherapy in subgroups stratified using unsupervised machine learning.

Author	Development cohorts	Validation cohorts	Clustering methods	Predictors	Clusters	Follow-up time	Outcomes	Drug response
Ahlqvist[17]	8980 registry-based prospective cohort	NHANES, CDMDS, CANVAS, ORIGIN, ADOPT, and RECORD	K means	Age-of-onset, BMI, HOMA2IR, HOMA2B, HbA1c, GADab	Five: SIRD, SIDD, MOD, MARD, and SAID	HbA1c variation in and complications in 15 years	Increased risk for cardiorenal disease in SIRD and increased retinopathy in SIDD and SAID	MARD: DPPIV or SU, SIRD: TZD MOD: SGLT2i SIDD: SU for the short-term and early requirement for insulin[47–49]
Mariam[50]	4946 participants treated with intensified glucose lowering in the ACCORD trial	N/A	Modified dynamic time-warping approach	HbA1c trajectories	Four	Follow-up time for 7 years	MACE risk varied in different clusters	A group benefited from intensive glycemia treatment in reducing CVD risk
Nourizadeh-Sedaghat[51]	71 old T2DM patients	N/A	K means or the K medoids	Age, BMI, eGFR, TG, duration of diabetes, HbA1c	Here	5 days observation	N/A	Insulin dosage difference among the three clusters
Segar[52]	ACCORD N = 6466	Look AHEAD (n = 4211) BARI 2D (n = 1495)	Gaussian mixture models, latent class analysis, finite mixture models (FMMs), and principal component analysis (PCA)	Demographics, medical and social history, laboratory values, and diabetes complications	Three phenotype groups	9.1 follow-up years	Difference in the risk of early coronary revascularization among subgroups	difference in glucose levels in response to intensive glycemic control among subgroups
Nair[53]	Scottish Care Information-Diabetes (SCI-Diabetes) N = 23,137	UK Biobank (n = 7, 332) and a diabetes outcome progression trial (ADOPT, N = 4150)	DDRTree algorithm	11 phenotypes including age of diagnosis, sex, HbA1c, BMI, HDL-C, triglycerides, total cholesterol, ALT, creatinine, and SBP and DBP at diagnosis	A tree structure was used to visualize diabetes	5-year risk for all endpoints was estimated	Uneven distribution of MACE, CKD, and diabetes retinopathy on the tree	Uneven distribution of risks of insulin initiation, SU, and TZD failure on the tree

NHANES, The National Health and Nutrition Examination Survey; CDMDS, China National Diabetes and Metabolic Disorders Survey; CANVAS, Canagliflozin Cardiovascular Assessment Study; ORIGIN, Outcome Reduction With Initial Glargine Intervention; ADOPT, A Diabetes Outcome Progression Trial; RECORD, Rosiglitazone Evaluated for Cardiovascular Outcomes in Oral Agent Combination Therapy for Type 2 Diabetes; BMI, body mass index; HOMA, homeostasis model assessment; HbA1c, hemoglobin A1c; GADab, glutamic acid decarboxylase antibody; SIRD, severe insulin-resistant diabetes; SIDD, severe insulin-deficient diabetes; MOD, mild obesity-related diabetes; MARD, mild age-related diabetes; SAID, severe autoimmune diabetes; SGLT2i, sodium-glucose cotransporter 2 inhibitors; DPP4i, dipeptidyl peptidase 4 inhibitors; SU, sulfonylureas; TZD, thiazolidinedione; ACCORD, Action to Control Cardiovascular Risks in Diabetes Study; MACE, major adverse cardiovascular events; CVD, cardiovascular disease; T2DM, type 2 diabetes mellitus; eGFR, estimated glomerular filtration rate; TG, triglyceride; AHEAD, Action for Health in Diabetes; BARI2D, Bypass Angioplasty Revascularization Investigation in Type 2 Diabetes; DDRTree, Discriminative Dimensionality Reduction via Learning a Tree; HDL-C, high-density lipoprotein cholesterol; ALT, alanine transaminase; SBP, systolic blood pressure; DBP, diastolic blood pressure; CKD, chronic kidney disease.

Table 2. Predicting responses to pharmacotherapy in subgroups stratified using supervised machine learning.

Author	Type of dataset	Sample size	Methods	Predictors	Followup time	Model performance	Outcomes
Glucose-lowering effect							
Huang[54]	Perspective cohort	Development dataset: N = 90 Validation dataset: N = 26	A novel method: differential metabolic network construction (DMNC),	Metabolites panel plus laboratory measurements	16 weeks of treatment	Roc AUC: 0.893 to 1.000	The HbA1c lowering effect of gliclazide modified release tablets
Fujihara[55]	Cross-sectional registry-based cohort	4860	logistic regression (LR) versus neural network (NN)	Age, sex, BMI, duration of diabetes, HbA1c, hypertension, eGFR	NA	Roc AUC of 0.80, for LR and 0.70 for NN	Predicting the insulin initiation
Del Parigi[56]	RCT	1363	RF	age, sex, race, ethnicity, background treatment BMI smoking, eGFR, HbA1c SBP and FPG	52 weeks	Out-of-bag estimates of the prediction error rate: 28.4–22.5%	Predicting the response to linagliptin and empagliflozin or combination therapy
Eby[16]	Nationally representative insurance claims database	15, 331	XgBoost	Demographic and clinical data	8.7 years	Average ROC 0.79	Model predicted the patients on target, maintained the target, and never met the target
Berchialla[57]	RCT	n = 385 for Prologue and n = 103 in SAIS1	GBM, GLM, RF, CART, BART, SVM, and a super learner by combining all these methods:	Demographic and clinical data	6 months	ROC AUC: 0.9205	Predicting HbA1c decline of sitagliptin versus placebo
Murphree[58]	Retrospective cohort of commercially insured adults and Medicare Advantage beneficiaries with prediabetes or diabetes	12,147	avNNet, gcvEarth and bagEarthGCV, bayesglm, earth, evtree, fda, mStepAIC	Comorbidities, baseline HbA1c level, baseline metformin dosage, and demographic variables	12 months	Roc AUC 0.58 to 0.75	Predicting HbA1c on-target rate of metformin
Wang[59]	cross-sectional data of insulin-treated patients in multiple centers	2787	RF, SVM, BP-ANN with EN	Demographic and clinical data	NA	0.61–0.73 with RFISVM and BPANN, 0.72–0.75 with EN	Predicting glycemic on-target rate of insulin treatment

(continued)

Table 2. Continued.

Author	Type of dataset	Sample size	Methods	Predictors	Followup time	Model performance	Outcomes
Safety endpoints							
Pettus[60]	The Optum Humedica EHR database	157, 573	LASSO	Manually created covariates and covariates automatically created from all available data	188–264 days	ROC AUC 0.75–0.84	Predicting the hypoglycemia episodes and severe events of basal insulin
Yang[61]	EHR	29,843	XgBoost	37 predictive variables and their weights were selected from 176 variables by XGboost	More than 24 h	ROC AUC 0.82	Predicting hypoglycemia responses to insulin, sulfonylureas, or nateglinide
Yang[62]	5% random sample of Fee-for-Service Medicare beneficiaries	17,694	RF, LASSO, and EN	65 predictor candidates	1.5 year follow up	C statistics of RF : 0.72	Predicting incident AKI event after index date of SGLT2i
Elhadd[63]	Prospective cohort	13	XgBoost, LR, RF, SVM, and DNN	Clinical data plus pedometer and CGM data	2 weeks before and 2 weeks during Ramadan	XgBoost predicted R2 of 0.836 and MAE of 17.47.	Predicting glucose level and hypoglycemic episodes of antidiabetic medications
Mortality and cardio-renal outcomes							
Basu, S.[64]	ACCORD trial	10,251	Gradient forest + decision tree	Conventional clinical measurements at baseline and hemoglobin glycosylation index	7 years	C statistics: 0.62–0.66	Differences in mortality with intensified therapy versus standard therapy
Yamada[65]	Milliman Consolidated Health Cost Guidelines Sources Database from 2011 to 2016)	199,116	Deep neural network-based machine learning	Conventional clinical measurements at baseline and drug use information	Medium observation period 16.5–18.7	ROC AUC 0.76	Predicting differed risks in myocardial infarction in patients treated with DPP4i, SGLT2i; versus GLP1RA
Yang[66]	Medicare beneficiaries	13 904	Elastic net, LASSO, gradient boosting machine, and random forests	16 variables	1.5 years	C-statistic of 0.81	Predicting lower extremity amputations of canagliflozin

Table 2. Continued.

Author	Type of dataset	Sample size	Methods	Predictors	Followup time	Model performance	Outcomes
Oikonomou[67]	RCT	Development cohort: CANVAS (n = 4327) Validation cohort: CANVAS-R (n = 5828)	XgBoost	75 variable out of 146 variables	5 years	Internal cross-validation RMSE : 0.46	Identifying patients who can benefit from SGLT2i treatment versus placebo in preventing MACE progression
Zhou 2019[68]	Japanese commercial medical database	n = 990 on SGLT2i and 4257 on DDP4 inhibitors; splitted 7:3 to learning and validation datasets	Proprietary supervised learning algorithm (Q-Finder)	150 clinical features	15 months	The c-statistics ranged from 0.79 to 0.82 in the learning dataset and from 0.80 to 0.84 in the validation dataset	Responses to SGLT2i versus DPP4i in renal function preservation
Zou[47]	RCT	Development cohort: placebo arm of CANVAS-R (N = 2771), Validation cohort: CANVAS (N = 1043)	XgBoost	Demographic and clinical data	5 years	ROC 0.71	Stratify patients with high albuminuria risks who benefited from SGLT2i therapy

RMSE, root mean squared error; MAPE, mean absolute percentage error; GBM, gradient boosting machines; GLM, generalized linear model; CART, classification and regression tree; BART, Bayesian additive regression trees; RNN, recurrent neural network; GRU, gated recurrent unit; LSTM, long-short term memory; EN, elastic network; DNN, deep neural networks; LASSO, least absolute shrinkage and selection operator; RF, random forest; SVM, support vector machines; LR, logistic regression; BP-ANN, backpropagation–artificial neural network; eGFR, estimated glomerular filtration rate; ROC AUC, area under the curve of the receiver operating characteristic curve; EHR, electronic health record; RCT, randomized clinical trial; CANVAS, Canagliflozin Cardiovascular Assessment Study; NHANES, The National Health and Nutrition Examination Survey; CDMDS, China National Diabetes and Metabolic Disorders Survey; ORIGIN, Outcome Reduction with Initial Glargine Intervention; ADOPT, A Diabetes Outcome Progression Trial; RECORD, Rosiglitazone Evaluated for Cardiovascular Outcomes in Oral Agent Combination Therapy for Type 2 Diabetes; SGLT2i, sodium–glucose cotransporter 2 inhibitors; DPP4i, dipeptidyl peptidase 4 inhibitors

other algorithms. Most supervised machine learning algorithms facilitate the computation of variable importance during model derivation.[36] The importance assigned to a variable within a model underscores the correlation between that variable and the endpoints.

Endpoint selection

For supervised machine learning, a specific endpoint should be determined prior to model development. The most common selected endpoints for drug selection are HbA1c decline and HbA1c on-target rate.[56,59] Hypoglycemic episodes, one of the most common side effects of hypoglycemic therapies and the key consideration of insulin delivery systems, are usually selected as the safety endpoint.[75] Drug-specific safety endpoints, e.g., lower limb amputation for canagliflozin[66] and acute kidney injury for sodium-dependent glucose transporters 2(SGLT2i),[62] were chosen in specific cohorts. There is a particular focus on models predicting cardiovascular and renal endpoints, including major adverse cardiovascular events (MACE) and albuminuria progress.[47,65,68]

For unsupervised machine learning, multi-endpoints are evaluated among subgroups in most studies,[17,64] and subgroups may have different disease trajectories and cardiovascular outcomes. As an exception, a study used soft clustering methods such as Gaussian mixture models and finite mixture models (FMMs)[52] to predict a single outcome: the atherosclerotic cardiovascular disease risk in type 2 diabetes patients.[52] However, the subgroups were not replicated in other cohorts. There was also a study that identified subgroups with different risks of recurring CVD events; however, these subtypes were not associated with drug treatment decisions.[19]

Development and validation of supervised machine learning models

The basic paradigm for developing supervised machine learning algorithms encompasses three stages: derivation, internal validation, and external validation. Cohorts are typically partitioned into training and internal validation datasets. Models are trained using predictors and labeled outcomes. The internal validation datasets, which bear high similarity to the training set, serve to assess the algorithm's predictive capacity for outcomes. To avoid sampling bias, a five- or 10-fold cross-validation is always applied for model deviation and internal validation. The parameters are finely tuned to achieve the highest internal prediction accuracy. Usually, it would be ideal to assess the prediction accuracy both in the internal validation dataset and a spare external validation cohort to avoid model overfitting, which commonly happens in complex models of machine learning and deep learning. C-statistics or the area under the curve of the receiver

operator curve (ROC AUC) are usually used to assess diagnostic accuracy. Root mean square error (RMSE) and mean absolute percentage error (MAPE) are typically used to estimate the accuracy of the regression models. Other evaluations include error rates, F1 score, and decision curve analysis (DCA).[61]

Subgroups generated by unsupervised learning

Usually, supervised learning algorithms can be used to predict the presence of an outcome in patients treated with certain drugs[59] or stratify participants into groups according to their progression risks or a threshold of predicted outcomes.[47] Clusters are generated based on selected features. Optimal group number is critical, and there are a few methods to determine the optimal subgroup numbers. For example, Gap Statistic, Elbow Method, Silhouette Coefficient, and Bayesian information criterion (BIC) are used to determine the optimal K number for K means clustering.[76] Time-series data such as CGM data usually require specific clustering methods such as longitudinal finite mixture modeling (LFMM), which contains latent class growth analysis (LCGA), group-based trajectory models (GBTM), and growth mixture modeling (GMM).[77] Unsupervised algorithms offer numerous ways to generate subgroups, which necessitate extensive validation of these subgroups. Clinicians only adopt those subgroups that consistently demonstrate influence on key parameters, including glucose endpoints, micro- and macro-vascular complications, and drug responses.

Differences in drug responses

Previous reviews mostly assessed the prediction accuracy of machine learning models. However, even using models with nearly 100% accuracy, it is difficult to be accepted by clinicians unless there is a treatment-by-group interaction between a specific drug and subgroups generated by this model. Algorithms divided cohorts into subgroups and "p" for interaction between drug effect and subgroup was assessed. This process serves to underline the clinical utility of the algorithms.[47,50] For decision-making, randomized controlled trials (RCTs) offer high-level clinical evidence. Therefore, subgroup analysis or post hoc analysis can provide exploratory evidence for the algorithm's applicability in drug selection, but is not sufficient enough to bring changes to current clinical practice. In cohorts that have not been randomized, alternative comparisons may also uncover potential treatment differences among drugs. In a study using EHR, the patients on SGLT2i and dipeptidyl peptidase 4 inhibitors (DPP4i) were matched using propensity scoring, and the class effect of these drugs on renal function preservation was examined.[68] These methods could potentially be used to assess the

treatment-by-group interaction in machine learning-identified subgroups.

Current clinical evidence

Current supervised learning algorithms have acceptable diagnostic accuracy, and they may help to guide the use of insulin, oral hypoglycemic drugs, and glucagon-like peptide-1 receptor agonist (GLP-1RA) with regard to their HbA1c-lowering effects, HbA1c on-target rates, hypoglycemic episodes, renal function preservation, and cardiovascular outcomes. Some studies were able to identify the class effect of two active drugs[68] or drug versus placebo[67] effect on a specific outcome. Some studies predicted the glycemic response of a single therapy such as insulin[59] and metformin.[58] Despite the promising potential, two primary obstacles hinder the clinical implementation of these algorithms. Firstly, the complexity of some machine learning models remains a significant challenge for clinicians. Certain studies even employed more than a hundred variables as inputs, so the algorithm became too time-consuming to be applicated in routine clinical practice. As an improvement, a study developed online tools with nine inputs to facilitate the use of their algorithm in clinics[67] and another study used only four variables and used easy-to-use cutoff values to define different subgroups to predict the mortality risk of intensive hypoglycemic therapy.[64] For clinical practicality, the models with fewer and simpler predictors are generally more acceptable. However, there might be a trade-off between model simplicity and accuracy. Secondly, another issue is the external validation of the algorithms. Few algorithms have undergone extensive validation in diverse cohorts. Precision medicine is an intricate process. For example, the models specifically designed for canagliflozin may not be suitable for other SGLT2 inhibitors, thus constraining the broader applicability of these models. Therefore, external validation is imperative to ensure the generalizability of these models.

For data-driven clusters, the external validity was much more accepted than supervised machine learning. The All New Diabetics In Scania (ANDIS) study was used to generate five clusters using simple variables and the model was stable in many ethnic groups and clinical trials. To date, external validation of ANDIS clusters was conducted in more than 20 cohorts, although a specific ethnic cluster may exist in India.[78] Evidence was built on different responses of clusters to insulin,[48] sulfonylureas (SU), thiazolidinedione (TZD), metformin,[49] SGLT2i,[47] and metabolic surgery[79] using data from clinical trials and retrospective cohorts. Generally, SU may be used for severe insulin-deficient diabetes (SIDD) to control short-term hyperglycemia; however, the sustainability of blood glucose control was not optimal. DPP4i can be used in mild age-related diabetes (MARD) for the high glycemic on-target rate in this group and low incidence of hypoglycemia. It was found that severe insulin-resistant diabetes (SIRD) may respond better to TZD for better glycemic control[49] and mild obesity-related diabetes (MOD) achieved the highest glycemic decline using SGLT2i compared to DPP4i and SU.[47] However, the evidence for whether data-driven clusters responded differently to GLP-1RA was missing, although disease progression was described in GLP-1-RA cohorts.[80] There was a lack of validation for other subgroups derived from unsupervised learning to predict drug responses.[19,50] Before data-driven clusters can be used in clinics, there are still a few things that need to be addressed. (1) Inconsistencies have been observed in the progression of complications across different cohorts, which might be attributed to cluster transitions that occur in some patients.[81] This suggests that the use of simple baseline predictors to create subgroups may not adequately refine responses to drugs, given that both glycemic control and cardiorenal risks are dynamic processes. (2) Although clusters may theoretically respond differently to drug therapies in terms of complication development,[82] few studies have observed the effect of pharmacotherapy can alter the cardio-renal in a specific subgroup.[47] (3) Studies suggested that data-driven clusters were not as effective as simple clinical measurements, e.g., HbA1c, age, and BMI, in distinguishing treatment effects.[49] This may limit the application of this algorithm. In summary, substantial work remains before this method can guide clinical decision-making in pharmacotherapy effectively.

Conclusions

In current practice, machine learning methods are robust to predict clinical outcomes and even drug responses; however, they are not widely accepted to guide clinical decisions in precision diabetes pharmacotherapy. We hope machine learning can help clinicians precisely identify who may achieve the largest benefit from a certain drug.

ORCID iD: Xiantong Zou (iD) https://orcid.org/0000-0002-3262-2168

References

1. Chung WK, Erion K, Florez JC, et al. Precision medicine in diabetes: a consensus report from the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care* 2020; 43: 1617–1635.
2. Jordan MI and Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science* 2015; 349: 255–260.
3. Rajkomar A, Dean J and Kohane I. Machine learning in medicine. *N Engl J Med* 2019; 380: 1347–1358.
4. Razavian N, Blecker S, Schmidt AM, et al. Population-level prediction of type 2 diabetes from claims data and analysis of risk factors. *Big Data* 2015; 3: 277–287.
5. Alpaydin E. *Introduction to machine learning*. Cambridge, MA: MIT Press Ltd; 2020.
6. Lauzon FQ (ed) An introduction to deep learning. 2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA), 2–5 July 2012.
7. Fregoso-Aparicio L, Noguez J, Montesinos L, et al. Machine learning and deep learning predictive models for type 2 diabetes: A systematic review. *Diabetol Metab Syndr* 2021; 13: 148.
8. De Silva K, Enticott J, Barton C, et al. Use and performance of machine learning models for type 2 diabetes prediction in clinical and community care settings: protocol for a systematic review and meta-analysis of predictive modeling studies. *Digit Health* 2021; 7: 20552076211047390.
9. Liu Q, Zhang M, He Y, et al. Predicting the risk of incident type 2 diabetes mellitus in Chinese elderly using machine learning techniques. *J Pers Med* 2022; 12: 905.
10. Hertroijs DFL, Elissen AMJ, Brouwers M, et al. A risk score including body mass index, glycated haemoglobin and triglycerides predicts future glycaemic control in people with type 2 diabetes. *Diabetes Obes Metab* 2018; 20: 681–688.
11. Felizardo V, Garcia NM, Pombo N, et al. Data-based algorithms and models using diabetics real data for blood glucose and hypoglycaemia prediction - A systematic literature review. *Artif Intell Med* 2021; 118: 102120.
12. Sudharsan B, Peeples M and Shomali M. Hypoglycemia prediction using machine learning models for patients with type 2 diabetes. *J Diabetes Sci Technol* 2015; 9: 86–90.
13. Kodama S, Fujihara K, Shiozaki H, et al. Ability of current machine learning algorithms to predict and detect hypoglycemia in patients with diabetes mellitus: meta-analysis. *JMIR Diabetes* 2021; 6: e22458.
14. Zomer E, Liew D, Owen A, et al. Cardiovascular risk prediction in a population with the metabolic syndrome: Framingham vs. UKPDS algorithms. *Eur J Prev Cardiol* 2014; 21: 384–390.
15. 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.
16. Eby EL, Kelly NR, Hertzberg JK, et al. Predicting response to bolus insulin therapy in patients with type 2 diabetes. *J Diabetes Sci Technol* 2022; May 20: 19322968221098057.
17. Ahlqvist E, Storm P, Käräjämäki A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: A data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol* 2018; 6: 361–369.
18. Udler MS, Kim J, von Grotthuss M, et al. Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. *PLoS Med* 2018; 15: e1002654.
19. Sharma A, Zheng Y, Ezekowitz JA, et al. Cluster analysis of cardiovascular phenotypes in patients with type 2 diabetes and established atherosclerotic cardiovascular disease: A potential approach to precision medicine. *Diabetes Care* 2022; 45: 204–212.
20. Zhu T, Li K, Herrero P, et al. Basal glucose control in type 1 diabetes using deep reinforcement learning: An in silico validation. *IEEE J Biomed Health Inform* 2021; 25: 1223–1232.
21. Kim DY, Choi DS, Kim J, et al. Developing an individual glucose prediction model using recurrent neural network. *Sensors (Basel)* 2020; 20: 6460.
22. Breton MD, Kanapka LG, Beck RW, et al. A randomized trial of closed-loop control in children with type 1 diabetes. *N Engl J Med* 2020; 383: 836–845.
23. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama* 2016; 316: 2402–2410.
24. Raman R, Srinivasan S, Virmani S, et al. Fundus photograph-based deep learning algorithms in detecting diabetic retinopathy. *Eye (Lond)* 2019; 33: 97–109.
25. Zhang K, Liu X, Xu J, et al. Deep-learning models for the detection and incidence prediction of chronic kidney disease and type 2 diabetes from retinal fundus images. *Nat Biomed Eng* 2021; 5: 533–545.
26. Chang S, Chen JY, Chuang YJ, et al. Systems approach to pathogenic mechanism of type 2 diabetes and drug discovery design based on deep learning and drug design specifications. *Int J Mol Sci* 2020; 22: 166.
27. Zhao J, Xu P, Liu X, et al. Application of machine learning methods for the development of antidiabetic drugs. *Curr Pharm Des* 2022; 28: 260–271.
28. Srisongkram T, Waithong S, Thitimetharoch T, et al. Machine learning and in vitro chemical screening of potential α-amylase and α-glucosidase inhibitors from Thai Indigenous plants. *Nutrients* 2022; 14: 267.
29. Jiang R, Calhoun VD, Noble S, et al. A functional connectome signature of blood pressure in >30000 participants from the UK biobank. *Cardiovasc Res* 2023; 119: 1427–1440.
30. Avvisato R, Forzano I, Varzideh F, et al. A machine learning model identifies a functional connectome signature that predicts blood pressure levels: imaging insights from a large population of 35882 patients. *Cardiovasc Res* 2023; 119: 1458–1460.
31. Huang J, Huth C, Covic M, et al. Machine learning approaches reveal metabolic signatures of incident chronic kidney disease in individuals with prediabetes and type 2 diabetes. *Diabetes* 2020; 69: 2756–2765.

32. Hathaway QA, Roth SM, Pinti MV, et al. Machine-learning to stratify diabetic patients using novel cardiac biomarkers and integrative genomics. *Cardiovasc Diabetol* 2019; 18: 78.

33. Gautier T, Ziegler LB, Gerber MS, et al. Artificial intelligence and diabetes technology: A review. *Metabolism* 2021; 124: 154872.

34. Zhu T, Li K, Herrero P, et al. Deep learning for diabetes: A systematic review. *IEEE J Biomed Health Inform* 2021; 25: 2744–2757.

35. Contreras I and Vehi J. Artificial intelligence for diabetes management and decision support: literature review. *J Med Internet Res* 2018; 20: e10775.

36. Silva K, Lee WK, Forbes A, et al. Use and performance of machine learning models for type 2 diabetes prediction in community settings: A systematic review and meta-analysis. *Int J Med Inform* 2020; 143: 104268.

37. Olusanya MO, Ogunsakin RE, Ghai M, et al. Accuracy of machine learning classification models for the prediction of type 2 diabetes mellitus: A systematic survey and meta-analysis approach. *Int J Environ Res Public Health* 2022; 19: 14280.

38. Lam B, Catt M, Cassidy S, et al. Using wearable activity trackers to predict type 2 diabetes: machine learning-based cross-sectional study of the UK biobank accelerometer cohort. *JMIR Diabetes* 2021; 6: e23364.

39. Anoop RK, Ashwini AP, Kanchan VP, et al. Machine-learning algorithm to non-invasively detect diabetes and pre-diabetes from electrocardiogram. *BMJ Innov* 2023; 9: 32.

40. Pyrros A, Borstelmann SM, Mantravadi R, et al. Opportunistic detection of type 2 diabetes using deep learning from frontal chest radiographs. *Nat Commun* 2023; 14: 4039.

41. Wang L, Mu Y, Zhao J, et al. IGRNet: A deep learning model for non-invasive, real-time diagnosis of prediabetes through electrocardiograms. *Sensors (Basel)* 2020; 20: 2556.

42. Kumar JNVRS, Kumar KH and Haleem A (ed.) IBM auto AI bot: diabetes mellitus prediction using machine learning algorithms. 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 9–11 May 2022.

43. Kengne AP, Patel A, Colagiuri S, et al. The Framingham and UK prospective diabetes study (UKPDS) risk equations do not reliably estimate the probability of cardiovascular events in a large ethnically diverse sample of patients with diabetes: the action in diabetes and vascular disease: Preterax and Diamicron-MR controlled evaluation (ADVANCE) study. *Diabetologia* 2010; 53: 821–831.

44. Kengne AP, Patel A, Marre M, et al. Contemporary model for cardiovascular risk prediction in people with type 2 diabetes. *Eur J Cardiovasc Prev Rehabil* 2011; 18: 393–398.

45. SCORE2-Diabetes: 10-year cardiovascular risk estimation in type 2 diabetes in Europe. *Eur Heart J.* 2023; 44: 2544–2556.

46. Wang M, Francis F, Kunz H, et al. Artificial intelligence models for predicting cardiovascular diseases in people with type 2 diabetes: a systematic review. *Intelligence-Based Medicine* 2022; 6: 100072.

47. Zou X, Huang Q, Luo Y, et al. The efficacy of canagliflozin in diabetes subgroups stratified by data-driven clustering or a supervised machine learning method: a post hoc analysis of canagliflozin clinical trial data. *Diabetologia* 2022; 65: 1424–1435.

48. Pigeyre M, Hess S, Gomez MF, et al. Validation of the classification for type 2 diabetes into five subgroups: a report from the ORIGIN trial. *Diabetologia* 2022; 65: 206–215.

49. Dennis JM, Shields BM, Henley WE, et al. Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: An analysis using clinical trial data. *Lancet Diabetes Endocrinol* 2019; 7: 442–451.

50. Mariam A, Miller-Atkins G, Pantalone KM, et al. A type 2 diabetes subtype responsive to ACCORD intensive glycemia treatment. *Diabetes Care* 2021; 44: 1410–1418.

51. Nourizadeh-Sedaghati A, Herbin M, Lukas-Croisier C, et al. Study of insulin requirement modeling in hospitalized elderly patients with type 2 diabetes at a late stage of stepwise escalation therapy. *Diabetes Technol Ther* 2016; 18: 308–315.

52. Segar MW, Patel KV, Vaduganathan M, et al. Development and validation of optimal phenomapping methods to estimate long-term atherosclerotic cardiovascular disease risk in patients with type 2 diabetes. *Diabetologia* 2021; 64: 1583–1594.

53. Nair ATN, Wesolowska-Andersen A, Brorsson C, et al. Heterogeneity in phenotype, disease progression and drug response in type 2 diabetes. *Nat Med* 2022; 28: 982–988.

54. Huang X, Zhou Y, Tang H, et al. Differential metabolic network construction for personalized medicine: study of type 2 diabetes mellitus patients' response to gliclazide-modified-release-treated. *J Biomed Inform* 2021; 118: 103796.

55. Fujihara K, Matsubayashi Y, Harada Yamada M, et al. Machine learning approach to decision making for insulin initiation in Japanese patients with type 2 diabetes (JDDM 58): model development and validation study. *JMIR Med Inform* 2021; 9: e22148.

56. Del Parigi A, Tang W, Liu D, et al. Machine learning to identify predictors of glycemic control in type 2 diabetes: an analysis of target HbA1c reduction using empagliflozin/linagliptin data. *Pharmaceut Med* 2019; 33: 209–217.

57. Berchialla P, Lanera C, Sciannameo V, et al. Prediction of treatment outcome in clinical trials under a personalized medicine perspective. *Sci Rep* 2022; 12: 4115.

58. Murphree DH, Arabmakki E, Ngufor C, et al. Stacked classifiers for individualized prediction of glycemic control following initiation of metformin therapy in type 2 diabetes. *Comput Biol Med* 2018; 103: 109–115.

59. Wang J, Wang MY, Wang H, et al. Status of glycosylated hemoglobin and prediction of glycemic control among patients with insulin-treated type 2 diabetes in North China: a multicenter observational study. *Chin Med J (Engl)* 2020; 133: 17–24.

60. Pettus J, Roussel R, Liz Zhou F, et al. Rates of hypoglycemia predicted in patients with type 2 diabetes on insulin glargine 300 U/ml versus first- and second-generation basal insulin analogs: the real-world LIGHTNING study. *Diabetes Ther* 2019; 10: 617–633.

61. Yang H, Li J, Liu S, et al. Predicting risk of hypoglycemia in patients with type 2 diabetes by electronic health record-based machine learning: development and validation. *JMIR Med Inform* 2022; 10: e36958.

62. Yang L, Gabriel N, Hernandez I, et al. Identifying patients at risk of acute kidney injury among medicare beneficiaries with

type 2 diabetes initiating SGLT2 inhibitors: A machine learning approach. *Front Pharmacol* 2022; 13: 834743.

63. Elhadd T, Mall R, Bashir M, et al. Artificial intelligence (AI) based machine learning models predict glucose variability and hypoglycaemia risk in patients with type 2 diabetes on a multiple drug regimen who fast during Ramadan (the PROFAST - IT Ramadan study). *Diabetes Res Clin Pract* 2020; 169: 108388.

64. Basu S, Raghavan S, Wexler DJ, et al. Characteristics associated with decreased or increased mortality risk from glycemic therapy among patients with type 2 diabetes and high cardiovascular risk: machine learning analysis of the ACCORD trial. *Diabetes Care* 2018; 41: 604–612.

65. Yamada T, Iwasaki K, Maedera S, et al. Myocardial infarction in type 2 diabetes using sodium-glucose co-transporter-2 inhibitors, dipeptidyl peptidase-4 inhibitors or glucagon-like peptide-1 receptor agonists: proportional hazards analysis by deep neural network based machine learning. *Curr Med Res Opin* 2020; 36: 403–409.

66. Yang L, Gabriel N, Hernandez I, et al. Using machine learning to identify diabetes patients with canagliflozin prescriptions at high-risk of lower extremity amputation using real-world data. *Pharmacoepidemiol Drug Saf* 2021; 30: 644–651.

67. Oikonomou EK, Suchard MA, McGuire DK, et al. Phenomapping-derived tool to individualize the effect of canagliflozin on cardiovascular risk in type 2 diabetes. *Diabetes Care* 2022; 45: 965–974.

68. Zhou FL, Watada H, Tajima Y, et al. Identification of subgroups of patients with type 2 diabetes with differences in renal function preservation, comparing patients receiving sodium-glucose co-transporter-2 inhibitors with those receiving dipeptidyl peptidase-4 inhibitors, using a supervised machine-learning algorithm (PROFILE study): A retrospective analysis of a Japanese commercial medical database. *Diabetes Obes Metab* 2019; 21: 1925–1934.

69. Zou X, Zhou X, Zhu Z, et al. Novel subgroups of patients with adult-onset diabetes in Chinese and US populations. *Lancet Diabetes Endocrinol* 2019; 7: 9–11.

70. Reddy GT, Reddy MPK, Lakshmanna K, et al. Analysis of dimensionality reduction techniques on big data. *IEEE Access* 2020; 8: 54776–54788.

71. Dong Y and Peng C-YJ. Principled missing data methods for researchers.

72. Emmanuel T, Maupong T, Mpoeleng D, et al. A survey on missing data in machine learning. *J Big Data* 2021; 8: 140.

73. Li L, Sun J, Ruan L, et al. Time-series analysis of continuous glucose monitoring data to predict treatment efficacy in patients with T2DM. *J Clin Endocrinol Metab* 2021; 106: 2187–2197.

74. Mordi IR, Trucco E, Syed MG, et al. Prediction of major adverse cardiovascular events from retinal, clinical, and genomic data in individuals with type 2 diabetes: A population cohort study. *Diabetes Care* 2022; 45: 710–716.

75. Bosnyak Z, Zhou FL, Jimenez J, et al. Predictive modeling of hypoglycemia risk with basal insulin use in type 2 diabetes: use of machine learning in the LIGHTNING study. *Diabetes Ther* 2019; 10: 605–615.

76. Yuan C and Yang H. Research on K-value selection method of K-means clustering algorithm. *J [Internet]* 2019; 2: 226–235.

77. van der Nest G, Lima Passos V, Candel MJJM, et al. An overview of mixture modelling for latent evolutions in longitudinal data: modelling approaches, fit statistics and software. *Adv Life Course Res* 2020; 43: 100323.

78. Anjana RM, Pradeepa R, Unnikrishnan R, et al. New and unique clusters of type 2 diabetes identified in Indians. *J Assoc Physicians India* 2021; 69: 58–61.

79. Raverdy V, Cohen RV, Caiazzo R, et al. Data-driven subgroups of type 2 diabetes, metabolic response, and renal risk profile after bariatric surgery: A retrospective cohort study. *Lancet Diabetes Endocrinol* 2022; 10: 167–176.

80. Kahkoska AR, Geybels MS, Klein KR, et al. Validation of distinct type 2 diabetes clusters and their association with diabetes complications in the DEVOTE, LEADER and SUSTAIN-6 cardiovascular outcomes trials. *Diabetes Obes Metab* 2020; 22: 1537–1547.

81. Zaharia OP, Kuss O, Strassburger K, et al. Diabetes clusters and risk of diabetes-associated diseases – Authors' reply. *Lancet Diabetes Endocrinol* 2019; 7: 828–829.

82. Tanabe H, Masuzaki H and Shimabukuro M. Novel strategies for glycaemic control and preventing diabetic complications applying the clustering-based classification of adult-onset diabetes mellitus: A perspective. *Diabetes Res Clin Pract* 2021; 180: 109067.