diabetesresearchandclinicalpractice190(2022)110013

Contents lists available at ScienceDirect

Diabetes Research and Clinical Practice

journal homepage: www.journals.elsevier.com/diabetes-research-and-clinical-practice

Prediction of complications of type 2 Diabetes: A Machine learning approach

Antonio Nicolucci [a,*], Luca Romeo [b], Michele Bernardini [b], Marco Vespasiani [c], Maria Chiara Rossi [a], Massimiliano Petrelli [d], Antonio Ceriello [e], Paolo Di Bartolo [f], Emanuele Frontoni [b], Giacomo Vespasiani [c]

[a] Center for Outcomes Research and Clinical Epidemiology – CORESEARCH, Pescara, Italy
[b] Department of Information Engineering, Università Politecnica delle Marche, Ancona, Italy
[c] METEDA s.r.l., Rome, Italy
[d] Clinic of Endocrinology and Metabolic Diseases, Department of Clinical and Molecular Sciences, Marche Polytechnic University, Ancona, Italy
[e] IRCCS MultiMedica, Milan, Italy
[f] Diabetes Unit, AUSL Romagna, Ravenna, Italy

ARTICLE INFO

Keywords:
Diabetes complications
Machine learning
Artificial intelligence
Predictive models
Clinical inertia
Quality of care

ABSTRACT

Aim: To construct predictive models of diabetes complications (DCs) by big data machine learning, based on electronic medical records.

Methods: Six groups of DCs were considered: eye complications, cardiovascular, cerebrovascular, and peripheral vascular disease, nephropathy, diabetic neuropathy. A supervised, tree-based learning approach (XGBoost) was used to predict the onset of each complication within 5 years (task 1). Furthermore, a separate prediction for early (within 2 years) and late (3–5 years) onset of complication (task 2) was performed. A dataset of 147.664 patients seen during 15 years by 23 centers was used. External validation was performed in five additional centers. Models were evaluated by considering accuracy, sensitivity, specificity, and area under the ROC curve (AUC).

Results: For all DCs considered, the predictive models in task 1 showed an accuracy > 70 %, and AUC largely exceeded 0.80, reaching 0.97 for nephropathy. For task 2, all predictive models showed an accuracy > 70 % and an AUC > 0.85. Sensitivity in predicting the early occurrence of the complication ranged between 83.2 % (peripheral vascular disease) and 88.5 % (nephropathy).

Conclusions: Machine learning approach offers the opportunity to identify patients at greater risk of complications. This can help overcoming clinical inertia and improving the quality of diabetes care.

1. Introduction

The global diabetes burden is rising at an alarming rate, and the number of people affected is projected to increase from 537 million individuals in 2021 to 783 million by 2045 [1].

The clinical, social and economic impact of diabetes is mainly related to long-term complications, including cardio-cerebrovascular events, kidney disease, eye damage, and nervous system damage [2–4].

Without urgent action to prevent complications, the already huge sums of money being spent on treating diabetes will rise to unsustainable levels for any healthcare system. Therefore, predicting adverse outcomes of diabetes is crucial to improve patient quality and length of life, as well as to reduce economic costs.

A large body of scientific evidence supports the importance of

Abbreviations: ACCORD, Action to Control Cardiovascular Risk in Diabetes; ALLHAT, Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial; AMD, Associazione Medici Diabetologi - Italian Association of Clinical Diabetologists; AUC, area under the ROC curve; CKD, chronic kidney disease; CV-10, Tenfold Cross-Validation; CVOP-10, Tenfold Cross-Validation Over Patients; DCs, end-organ diabetes complications; EMRs, electronic medical records; FN, false negative; FP, false positive; HF, heart failure; ID, patient's identification number; ML, maximum likelihood; ROC, receiver operating characteristics; RSF, random survival forest; SMOTE, synthetic minority oversampling technique; TN, true negative; TP, true positive; TWOI, time-widow of interest; XGB, XGBoost model.

* Corresponding author at: Center for Outcomes Research and Clinical Epidemiology – CORESEARCH, Corso Umberto I, 103 – 65122 Pescara, Italy.
E-mail address: nicolucci@coresearch.it (A. Nicolucci).

https://doi.org/10.1016/j.diabres.2022.110013
Received 26 May 2022; Received in revised form 11 July 2022; Accepted 16 July 2022
Available online 21 July 2022

controlling hyperglycemia, hypertension, hypercholesterolemia, body weight, and promoting smoking cessation in order to reduce end organ damage associated with diabetes [5–11]. Furthermore, large randomized trials have shown that new classes of glucose lowering drugs, such as SGLT2 inhibitors and GLP-1 receptor agonists, significantly reduce cardiovascular morbidity and mortality and protect against the progression of chronic kidney disease [12–14]; furthermore, SGLT2 inhibitors markedly reduce the risk of hospitalization for heart failure [15]. However, the desired therapeutic targets are not met in a large proportion of patients, and novel drugs are still largely underused, due to the persistence of clinical inertia [16–17].

Helping the clinicians to identify patients at high risk of developing diabetes complications and inform treatment decisions can represent an important aid to overcome clinical inertia and improve the quality of diabetes care. To this purpose, several prognostic models have been developed for diabetes complications [18–20]. However, these models are usually based on a restricted number of patient characteristics (features) and have been developed in populations of limited size, making their use for individual prediction problematic. More recently, machine learning approaches have been increasingly used [21–28], allowing the analysis of the complex interplay between a large array of different features, generally extracted from electronic medical records (EMRs) [21–25], administrative health data [26], or clinical trials [27–28]. However, the models developed usually predict one single complication [22,23,27,28], or the hospitalization due to complications [24], or address specific populations, such as obese individuals undergoing metabolic surgery [21]. The temporal window for the prediction of the complication varied in the different studies from 6 months [23] to 10 years [21], while the number of features used in the predictive model ranged from less than ten [25] to more than one thousand [22–24].

In this study, we constructed new predictive models of different diabetes complications by big data machine learning, based on electronic medical records (EMRs). We also made the prediction of complications separately in the short term (i.e. within two years) and the medium-term (between 3 and 5 years).

2. Methods

2.1. Dataset

The dataset was constructed by aggregating patients from 23 Italian diabetes centers, all sharing the same EMR system (Smart Digital Clinic, METEDA s.r.l). The dataset consists of 147.664 patients seen during 15 years, and was organized in the following 3 different fields:

a) the demographics field, storing the patient's identification number (patient ID), gender, year of birth, and date of diagnosis of diabetes. In particular, the first name and the surname of the patients were anonymized and associated with a random numeric patient ID;
b) the diseases field, which contains patient ID, disease codes, and disease diagnosis date;
c) the lab tests field, which stores patient ID, laboratory tests codes, laboratory tests values, and tests prescription date.

2.2. Pre-processing

Six groups of diabetes complications (DCs) were considered: eye complications, cardiovascular disease (including coronary artery disease and heart failure), cerebrovascular disease, peripheral vascular disease, nephropathy, diabetic neuropathy. Disease ICD-9-CM codes associated with DCs are summarized in supplementary Table 1. For each DC, a specific code corresponding to a specific field in the EMR indicated the absence of the pathology (for example, the code −3001 indicated the absence of eye complications). This code was utilized to correctly identify patients free from the specific complication. The first mention of the specific complication, irrespective of its severity, was considered as the onset of the complication. The presence of albuminuria was not considered as onset of diabetes nephropathy, due to the variability of this measure over time.

All the pathology codes that were not included in supplementary Table 1 were removed from the dataset. Then, for each patient, both pathology codes and laboratory tests codes were removed if pathology diagnosis date and laboratory tests prescription date preceded the date of diagnosis of diabetes. Finally, the inclusion criteria to select the time-widow of interest (TWOI) were applied for both control patients and DC patients as depicted in Fig. 1. TWOI of DC and control patients was calculated separately for each of the six groups of complications.

In particular, a DC patient had to have at least one code indicative of the absence of the specific complication, followed by at least one of the remaining pathology codes reported in supplementary Table 1. A TWOI of a DC patient (Fig. 1 - bottom side) was therefore delimited by the earliest pathology code indicating the absence of the complication and the earliest pathology code indicative of DC. A patient was included in the study only if the date of the earliest non-DC code was antecedent to the earliest date of a DC code.

A control patient had to have during the period of observation at least 2 codes indicating the absence of the complication and none of the remaining pathology codes reported in supplementary Table 1. A TWOI of a control patient (Fig. 1 - upper side) was therefore delimited by the earliest pathology code of non-DC and the latest code of non-DC.

Table 1
Predictive performance experimental results for task 1.

	Accuracy	Sensitivity (Recall)	Specificity	AUC
Eye complications	74.1	82.0	71.9	0.857
Cardiovascular disease	74.8	70.5	75.8	0.817
Cerebrovascular disease	70.5	89.1	59.2	0.846
Peripheral vascular disease	80.5	72.2	82.1	0.857
Nephropathy	89.7	92.8	88.0	0.970
Diabetic neuropathy	76.0	74.6	76.4	0.840

2.3. Tasks definition

Task 1, defined as the prediction of the onset of a complication among patients free from that specific complication at baseline, was evaluated by taking the average of all the lab tests values enclosed in the range of the TWOI. In addition to these predictors, also the information of gender, age, weight, height, waist circumference, blood pressure, ankle/brachial index and duration of diabetes was added. Overall, 46 features were considered in the predictive models (supplementary table 2).

All the missing values were filled with an extra-values imputation (i.e., −999).

Task 2, defined as the temporal stratification of the DC risk, was evaluated only among DC patients. For each patient, the unique lab tests prescription dates enclosed in the TWOI were considered. Each of those represented an observation of the patient. Thus, for each patient, starting from the earliest observation close to the lower boundary of the TWOI, the mobile averages of all the lab tests values inside the range of the dynamic time-windows were taken, observation by observation, until the latest observation close to the upper boundary of the TWOI. In addition to the already existing predictors (i.e., unique lab tests codes), also the information of gender, age, weight, height, waist circumference, blood pressure, ankle/brachial index and duration of diabetes was added and incremental number of observations per patient was added. All the missing values were filled with an extra-values imputation (i.e., −999).

The task 2 consisted in the prediction of the temporal distance between the date of each patient's observation and the date of DC diagnosis. The risk was defined "early" if the temporal distance was within the range of 0–2 years, otherwise was defined as "late" if within the

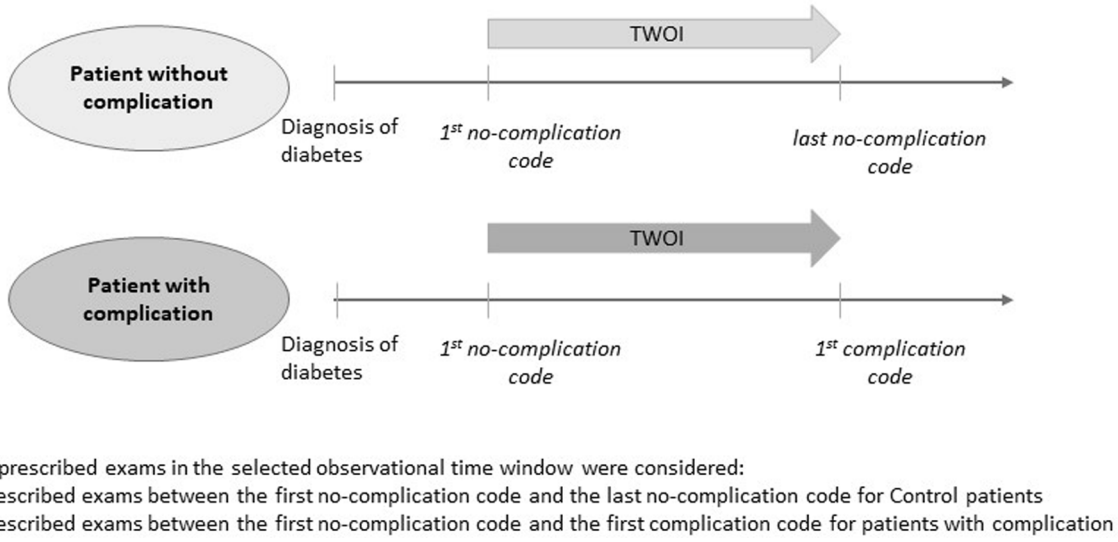

Fig. 1. Definition of the time window of interest (TWOI).

range of 3–5 years. DC patients whose temporal distance was>5 years were excluded from the study.

2.4. Selection of features

Among all laboratory tests available in EMRs, only a subset of exams routinely prescribed to the patients in clinical practice was selected for this study. This subset of selected laboratory tests features, along with other types of features (supplementary table 2) was utilized to feed the predictive model.

2.5. Train-test split

A common problem with machine learning models is represented by overfitting. Overfitting occurs when the model fits too well with the data, and fails to make accurate predictions when it is presented with new data [29]. The remedy to this problem is cross-validation, which implies the split of the data into two sets; the training set, used to train the data, and the test set, used to test the data. The test set, which is new data, is used to determine if the model is overfitting.

To perform task 1, a Tenfold Cross-Validation (CV-10) experimental procedure was chosen. In particular, CV-10 was implemented by dividing all patients in ten folds, by selecting nine folds for training and onefold for testing. CV-10 procedure was implemented without considering the temporal evolution of predictors, providing an overall average of the patient's clinical history.

On the contrary, to perform task 2, a Tenfold Cross-Validation Over Patients (CVOP-10) was chosen. CVOP-10 was implemented dividing all observations grouped by patients in ten folds, by selecting nine folds for training and onefold for testing. CVOP-10 procedure was implemented considering the temporal evolution of the patient's predictors.

2.6. Class imbalance and oversampling

The dataset included fewer instances where the complication was present (minority class) and more instances where it was not present (majority class). If the dataset is left as it is, the minority class tends to achieve poor predictive accuracy since the algorithm predicts the majority class more often [30]. To avoid this problem, the classes were balanced by increasing the minority class (oversampling). The first step consisted in dividing the dataset into training and test set. After that, the training set was oversampled using the SMOTE algorithm (Synthetic Minority Oversampling Technique) [31].

2.7. Performance metrics

The proposed task 1 and task 2 were evaluated by considering accuracy, sensitivity, specificity, and area under the ROC curve (AUC). Accuracy determines the number of correct predictions over the total number of predictions made by the model, and is calculated as the sum of true positive (TP) and true negative (TN), divided by the total number of predictions: (TP + TN)/(TP + TN + FP (false positive) + FN (false negative)).

Sensitivity (recall) is a measure of the proportion of patients that were predicted to have the complication of interest among those patients that actually had the complication. The formula for recall is TP/(TP + FN).

Specificity is a measure of the proportion of patients that were predicted not to have complication of interest among those patients that actually did not have the complication. The formula for specificity is TN/(TN + FP).

ROC or Receiver Operating Characteristics is a graphical plot of Sensitivity (y axis) against (1-Specificity) (x axis) or, in other words, a comparison of true positive rate and false positive rate. It is used to visualize a classifier's performance at different thresholds to determine the best threshold point for the classifier. AUC is the entire area under the ROC Curve that is used to determine the performance of a classifier across all classes. It ranges from 0 to 1; the higher the value, the better the performance.

2.8. Extreme gradient boosting model

The XGBoost method [32] is a supervised, tree-based learning approach. It was applied as a prediction model in consideration of its characteristics of high generalization performance and the low risk of overfitting that outperforms other data mining methods widely used for solving predictive medicine tasks [33]. The gradient tree boosting algorithms extend the concept of adaptive boosting by sequentially adding predictors and correcting previous models using the gradient descent algorithm.

2.9. Validation procedure

For what concerns the CV-10 and CVOP-10 experimental procedures, the optimization of the hyperparameters of the maximum likelihood (ML) models was performed by implementing a grid-search and optimizing the sensitivity (recall) in a nested Fivefold Cross-Validation.

Sensitivity was preferred over other optimization objectives, because the minimization of false negatives had the most clinical relevance for the task 1 experiment. Hence, each split of the outer loop was trained with the optimal hyper-parameters tuned in the inner loop. Although this procedure was computationally expensive, it allowed to obtain an unbiased and robust performance evaluation. The features importance of the XGB model was extracted according to the logic of showing the number of times the feature is used to split data across all trees during the XGB model training. The importance of every predictor in terms of percentage expresses its relative weight with respect to the total of the predictors.

2.10. External validation

The models for predicting complications of diabetes were externally validated using the databases of five different diabetes clinics not involved in the previous steps of the study. The sample size of the validation cohorts ranged between a minimum of 3,912 and a maximum of 20,007 patients.

2.11. Compliance with ethics guidelines

Anonymous data retrospectively collected in EMRs relative to 23 centers were provided by the Marche Regional Health Authority to the Department of Information Engineering of Università Politecnica delle Marche - Ancona for research purposes (i.e. development of predictive models).

Anonymous data retrospectively collected in EMRs relative to 5 centers were provided by the Associazione Medici Diabetologi (AMD) scientific society to the Department of Information Engineering of Università Politecnica delle Marche - Ancona for external validation.

Data were anonymous by design and neither Ethics Committees approval nor signed patient informed consent were required.

3. Results

Supplementary table 3 reports the number of patients evaluated in each diabetes complication prediction model. The sample ranged from a minimum of 7,852 patients for cerebrovascular disease to 40,555 patients for eye complications.

Table 1 shows the predictive performance of the models for task 1. For all the complications considered, the predictive models showed an accuracy over 70 % (from 70.5 % for cerebrovascular disease to 89.7 % for nephropathy), and AUC exceeded 80 %, reaching 97 % for the prediction of diabetic nephropathy. Similarly, sensitivity ranged from 70.5 % for the prediction of cardiovascular disease to 92.8 % for nephropathy, while specificity ranged between 59.2 % (cerebrovascular disease) and 88.0 % (nephropathy).

Supplementary table 4 reports the ranking of the ten variables providing the greatest contribution to the prediction of each end organ complication. For all the complications, each of the variables considered contributed for less than 5 % in predicting the outcome.

Table 2 shows the predictive performance of the models for task 2. All predictive models showed an accuracy largely exceeding 70 % (from 76.4 % for peripheral vascular disease to 78.8 % for cardiovascular disease) and an AUC over 0.85. Among all patients who developed the specific complication, sensitivity in predicting the occurrence of the complication within 2 years ranged between 83.2 % (peripheral vascular disease) and 88.5 % (nephropathy), while sensitivity in predicting the occurrence of the complication in the following 3–5 years ranged between 62.2 % (nephropathy) and 69.5 % (cardiovascular disease).

Results of external validation, performed using the databases of five different diabetes clinics not involved in the previous steps of the study, are summarized in Table 3.. We observed a variability in the performance of the predictive models across centers; however, for all end organ complications the performance was satisfactory (AUC > 0.6 for all complications in all centers, with the only exception of cerebrovascular disease, with an AUC of 0.59 in one center), being excellent (AUC > 0.80 in all centers) for the prediction of nephropathy. In task 2, the models showed a better performance in predicting complications occurring within 2 years than in predicting complications occurring after 3–5 years.

Table 2
Predictive performance experimental results for task 2.

	Accuracy	Recall (sensitivity) 0–2 years	Recall (sensitivity) 3–5 years	AUC
Eye complications	76.7	83.6	67.7	0.861
Cardiovascular disease	78.8	85.8	69.5	0.871
Cerebrovascular disease	77.3	86.4	64.2	0.855
Peripheral vascular disease	76.4	83.2	67.5	0.859
Nephropathy	78.5	88.5	62.2	0.866
Diabetic neuropathy	77.2	85.7	65.4	0.859

Table 3
Results of external validation in five centers: task 1 and task 2.

	Task 1			
	Accuracy Range	Sensitivity Range	Sensitivity Range	AUC Range
Eye complications	57.9–83.5	58.0–86.3	52.6–81.7	0.651–0.932
Cardiovascular disease	48.5–82.0	66.4–91.4	30.4–70.7	0.629–0.894
Cerebrovascular disease	58.5–83.9	63.6–96.9	12.1–58.3	0.590–0.860
Peripheral vascular disease	60.2–75.4	59.3–77.0	60.9–76.6	0.635–0.820
Nephropathy	73.6–96.8	76.4–91.6	71.5–98.3	0.817–0.979
Diabetic neuropathy	55.5–80.6	46.9–75.0	44.1–83.7	0.628–0.858
	Task 2			
	Accuracy Range	Sensitivity 0-2 years Range	Sensitivity 3-5 years Range	AUC Range
Eye complications	66.0–88.6	58.4–95.7	52.7–83.9	0.730–0.953
Cardiovascular disease	66.1–86.3	85.6–93.0	39.8–81.1	0.747–0.947
Cerebrovascular disease	65.4–85.9	73.0–93.9	36.2–82.4	0.730–0.939
Peripheral vascular disease	64.7–79.2	75.4–93.2	34.5–69.8	0.635–0.820
Nephropathy	65.4–86.0	86.4–98.3	30.5–73.5	0.715–0.939
Diabetic neuropathy	62.4–80.0	67.3–95.8	22.7–63.5	0.690–0.928

4. Conclusion

4.1. Major findings

Helping the clinicians to identify patients at high risk of developing diabetes complications in the short/medium term and inform treatment decisions can represent an important aid to overcome clinical inertia and improve the quality of diabetes care. In this study, we applied big data machine learning to construct new predictive models of different diabetes complications. Using data derived from electronic medical records, relative to 147.664 patients seen in 23 diabetes centers during 15 years, we developed models for the prediction of eye complications, cardiovascular disease, cerebrovascular disease, peripheral vascular disease, nephropathy, and diabetic neuropathy. For all the complications considered, the predictive models showed very good accuracy and high sensitivity, reflected by AUC values over 0.85. Predictive models

showed a particularly good performance in identifying the cases of early occurrence of the complications, making the models particularly attractive for their use in clinical practice, allowing the identification of individuals with more urgent need of attention. Furthermore, the models were built to predict the onset of complications among individuals for whom it was explicitly reported in EMRs they were free of that specific complication, thus avoiding misclassification.

Of note, each of the variables investigated gave a little contribution in predicting the complications, supporting the concept that the simultaneous effect of multiple factors and their interaction play a role in determining the individual risk of developing end organ damage. Such a complex interaction cannot be captured by the clinician, and emphasizes the role of machine learning techniques in improving prediction and helping decision making.

When the models were applied to patients from five additional diabetes centers, we found a variable level of predictive performance, ranging from acceptable to excellent. There was no obvious explanation for such a variability, which was not associated with the size of the population or the degree of missingness in the different features considered. It can be speculated that better performances of the predictive models could be obtained in those centers with more accurate registration of the data relative to presence or absence of the different complications. However, the models allowed in any case to identify a large number of patients at risk of complications, for whom it would be possible to improve diabetes care.

4.2. Comparison with existing evidence

Recently, machine learning approaches have been increasingly used to predict diabetes complications [21–28]. However, the models developed frequently predict one single complication [22,23,27,28], usually chronic kidney disease (CKD) [22,23,28]. In one study based on EMRs of 64,059 diabetes patients, the aggravation of CKD within six months was predicted using a model with 3073 features [23]. The model predicted CKD aggravation with 71 % accuracy.

In another study, a temporal-enhanced gradient boosting machine model was applied for the prediction of CKD, using a broad spectrum of EMR data and thousands of features on a retrospective cohort of 14,039 adult patients with diabetes [22]. The model achieved an AUC of 0.83 (95 % CI 0.76–0.85), 0.78 (95 % CI 0.75–0.82), and 0.82 (95 % CI 0.78–0.86) in predicting CKD risk in years 2, 3, and 4 since diabetes mellitus onset, respectively. CKD prediction models have also been developed using the data of the ACCORD trial [28]. Data on 10,251 patients and 22 features have been used applying different machine learning approaches to predict early (within 2 years) and late (within 7 years) occurrence of CKD. AUC values $>$ 0.70 were obtained in the different time windows. The prediction of development of nephropathy was among the complications considered in another study, along with retinopathy and neuropathy [25]. The study tested different machine learning approaches, with the best performance provided by logistic regression with rebalanced classes (AUC of 0.701, 0.734, and 0.721 for 3, 5 and 7 years prediction of nephropathy, respectively). However, the study population included only 943 patients and 7 features, and no external validation was performed. In our study, based on a very large population and a manageable number of features generally present in EMRs, the application of a XGBoost approach led to excellent predictive properties for the development of nephropathy in individuals with any diabetes duration, with an accuracy of 89.7 %, a sensitivity of 92.8 % and an AUC of 0.97. The external validation in five centers yielded AUC values ranging between 0.72 and 0.94.

Machine learning approaches to predict other diabetes complications have been less frequently utilized. In a large study using administrative health data from the single-payer health system in Ontario, Canada, a Gradient Boosting Decision Tree model was applied to predict three-year risk of adverse outcomes due to diabetes complications (hyper/hypoglycemia, tissue infection, retinopathy, cardiovascular events, amputation) [26]. The model was trained on data from 1,029,366 patients, validated on 272,864 patients, and tested on 265,406 patients. It included 700 features from multiple diverse data sources and showed an AUC of 0.795 for cardiovascular events. Another study used data from 8,756 patients free at baseline of heart failure (HF), with less than 10 % missing data, and enrolled in the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial, to predict incident HF [27]. Random survival forest (RSF) methods were applied, and the model was externally validated in a cohort of individuals with T2DM using the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). The RSF-based model for predicting incident HF at year 5, based on 22 features, showed an AUC of 0.77 and 0.74 in the ACCORD and ALLHAT cohorts, respectively. In our study, the prediction model for cardiovascular disease (including coronary heart disease and HF) showed an AUC of 0.817 in task 1 and 0.871 in task 2. In the five centers of the external validation, AUC ranged between 0.629 and 0.894 for task 1 and between 0.747 and 0.947 for task 2. A similar, strong discrimination was documented for the prediction of cerebrovascular disease and peripheral vascular disease. As for retinopathy, while several studies have been conducted on the application of machine learning to retinal fundus images for automating screening and diagnosis [34], the prediction of development of eye complications has been seldom investigated. In the study previously described conducted in Canada by Ravaut et al [26], an AUC of 0.807 for the prediction of retinopathy was obtained. The study by Dagliati et al. previously mentioned [25, including 943 patients and 7 features, showed AUC of 0.808, 0.769, and 0.726 for 3, 5 and 7 years prediction of retinopathy, respectively. In our study, AUC of 0.857 and 0.861 was documented for task 1 and task 2, respectively. Finally, the same study [25] documented AUC of 0.799, 0.714, and 0.769 for 3, 5 and 7 years prediction of neuropathy, respectively. In our study, we obtained an AUC for the prediction of diabetic neuropathy of 0.840 for task 1 and 0.859 for task 2.

4.3. Implications for clinical practice

The availability of prediction models that can be easily incorporated into EMRs can represent an important aid to the clinician for the identification of patients at higher risk of developing end-organ damage. Our models were based on information easily available and provided the prediction of all the major complications, thus allowing an overall assessment of the risk profile of the individual patient. The models also made the prediction in the short (2 years) and medium term (3–5 years), facilitating the identification of those patients who need immediate attention and more frequent follow-up visits, thus allowing a more rational use of healthcare resources. This process can have a great impact on clinical inertia and improve the quality of diabetes care. In particular, a mounting body of evidence supports the cardiovascular and renal benefits of new glucose-lowering drugs (SGLT2 inhibitors, GLP1 receptor agonists) [11–14]. Similarly, the importance of controlling hyperglycemia, hypertension, hypercholesterolemia, body weight, and promoting smoking cessation in order to reduce end organ damage associated with diabetes is widely recognized [5–10]. However, the desired therapeutic targets are not met in a large proportion of patients, and novel drugs are still largely underused, due to the persistence of clinical inertia [15–17].

The impact of the incorporation of the prediction models into EMRs will be evaluated in a large network of over 250 diabetes clinics participating in a continuous quality improvement initiative promoted by the Italian Association of Clinical Diabetologists (Associazione Medici Diabetologi - AMD) [17].

4.4. Strengths and limitations

Our study has strengths and limitations. Among the strengths, it should be underlined the large size of the population, the manageable number of features making the models useful in different healthcare

settings, the possibility to accurately predict all major complications, and the good performance of the models when externally validated in five different diabetes centers.

Among the limitations, it should be mentioned that the prediction models have been developed to maximize sensitivity (true positive rate), with a possible penalty in terms of specificity (true negative rate). However, for all complications, with the only exception of cerebrovascular disease, specificity exceeded 70 %, indicating that the expected proportion of false-positive predictions is reasonable and, more important, unlikely to have negative consequences for the patients. As an additional limitation, the performance of the predictive models was not uniform in the five centers involved in the external validation, and we did not find an obvious explanation for such variability. However, the models allowed to identify a large number of patients at risk of complications, for whom it would be possible to improve diabetes care. In any case, clinicians using the predictive models should be warned against the false reassurance provided by a negative prediction (i.e. low risk of developing complications); from this point of view, artificial intelligence must be considered as an aid, and not a substitute, for clinical judgment.

Preventing or delaying the onset of complications represents the primary goal of diabetes care. The growing availability of patient electronic medical records and the development of sophisticated algorithms that can learn from the data offer the unique opportunity to improve our ability to identify those patients at greater risk of complications and needing a closer monitoring or more appropriate interventions.

The integration of machine learning models with clinical experience could represent the new frontier for the provision of care ever closer to the needs of the individual patient.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Source of funding

The study was supported by a grant provided by METEDA s.r.l. (Rome, Italy) to Università Politecnica delle Marche (Ancona, Italy) Grant No. 153, 30/9/2019.

MV is the General Manager of METEDA.

AN and MCR received funding for research from Alfasigma, AstraZeneca, NovoNordisk, METEDA, Pikdare, Sanofi, Shionogi, SOBI, and Theras.

AC received consulting fees from Astra Zeneca, Bayer, Elsevier, Eli Lilly, Theras, and Roche Diagnostic; payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events from Novo Nordisk, Astra Zeneca, Berlin Chemie, Roche Diagnostics; participation on a Data Safety Monitoring Board or Advisory Board of Eli Lilly.

PDB received funding for research from Astra Zeneca, Eli Lilly, Novo Nordisk, Mundipharma, Abbvie Allergan, Boheringer, Sanofi, MSD, Bayer, Theras Italy, Ypsomed; consulting fees from Astra Zeneca, Eli Lilly, Novo Nordisk, Abbvie Allergan, Boehringer; support for attending meetings and/or travel from Astra Zeneca, Eli Lilly, Novo Nordisk, Mundipharma, Abbvie Allergan, Boheringer, Sanofi, MSD, Bayer, Theras Italy, Ypsomed; participation on a Data Safety Monitoring Board or Advisory Board of Astra Zeneca, Eli Lilly, Novo Nordisk, Abbvie Allergan, Boehringer; received equipment, materials, drugs, medical writing from Astra Zeneca; role of Past President of the Associazione Medici Diabetologi (AMD).

GV received consulting fees from METEDA.

LR, MB, MP, EF had nothing to disclose.

Author contributions

AN, GV, LR, MB, EF contributed to study concept. LR, MB, EF, AN analyzed data. GV, LR, MB, EF, AN developed models. All authors contributed to interpretation of data and critical revision of the manuscript for important intellectual content. GV is the guarantor of this work and, as such, has full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.diabres.2022.110013.

References

[1] Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 2022;183:109119. https://doi.org/10.1016/j.diabres.2021.109119.

[2] Deshpande AD, Harris-Hayes M, Schootman M. Epidemiology of diabetes and diabetes-related complications. Phys Ther 2008;88:1254–64. https://doi.org/10.2522/ptj.20080020.

[3] Harding JL, Pavkov ME, Magliano DJ, Shaw JE, Gregg EW. Global trends in diabetes complications: a review of current evidence. Diabetologia 2019;62:3–16. https://doi.org/10.1007/s00125-018-4711-2.

[4] Hazel-Fernandez L, Li Y, Nero D, Moretz C, Slabaugh L, Meah Y, et al. Relationship of diabetes complications severity to healthcare utilization and costs among Medicare Advantage beneficiaries. Am J Manag Care 2015;21:e62–70. PMID: 25880269.

[5] UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). Lancet 1998;352:837-853. PMID: 9742976.

[6] UK Prospective Diabetes Study Group. Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 38. UK Prospective Diabetes Study Group. BMJ 1998;317:703–713. PMID: 9732337.

[7] Colhoun HM, Betteridge DJ, Durrington PN, Hitman GA, W Neil HA, Livingstone SJ, et al. Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the Collaborative Atorvastatin Diabetes Study (CARDS): multicentre randomised placebo-controlled trial. Lancet 2004;364 (9435):685–96.

[8] Gæde P, Vedel P, Larsen N, Jensen GVH, Parving H-H, Pedersen O. Multifactorial intervention and cardiovascular disease in patients with type 2 diabetes. N Engl J Med 2003;348(5):383–93.

[9] Look AHEAD Research Group; Gregg E, Jakicic J, Blackburn G, et al. Association of the magnitude of weight loss and changes in physical fitness with long-term cardiovascular disease outcomes in overweight or obese people with type 2 diabetes: a post-hoc analysis of the Look AHEAD randomised clinical trial. Lancet Diabetes Endocrinol 2016;4:913–921. DOI: 10.1016/S2213-8587(16)30162-0.

[10] Pan A, Wang Y, Talaei M, Hu FB. Relation of smoking with total mortality and cardiovascular events among patients with diabetes mellitus. Circulation 2015; 132:1795–804. https://doi.org/10.1161/CIRCULATIONAHA.115.017926.

[11] Zelniker TA, Wiviott SD, Raz I, Im K, Goodrich EL, Bonaca MP, et al. SGLT2 inhibitors for primary and secondary prevention of cardiovascular and renal outcomes in type 2 diabetes: a systematic review and meta-analysis of cardiovascular outcome trials. Lancet 2019;393(10166):31–9.

[12] Kristensen SL, Rørth R, Jhund PS, Docherty KF, Sattar N, Preiss D, et al. Cardiovascular, mortality, and kidney outcomes with GLP-1 receptor agonists in patients with type 2 diabetes: a systematic review and meta-analysis of cardiovascular outcome trials. Lancet Diabetes Endocrinol 2019;7(10):776–85.

[13] Neuen BL, Young T, Heerspink HJL, Neal B, Perkovic V, Billot L, et al. SGLT2 inhibitors for the prevention of kidney failure in patients with type 2 diabetes: a systematic review and meta-analysis. Lancet Diabetes Endocrinol 2019;7(11): 845–54.

[14] Butler J, Usman MS, Khan MS, Greene SJ, Friede T, Vaduganathan M, et al. Efficacy and safety of SGLT2 inhibitors in heart failure: systematic review and meta-analysis. ESC Heart Fail 2020;7(6):3298–309.

[15] Reach G, Pechtner V, Gentilella R, Corcos A, Ceriello A. Clinical inertia and its impact on treatment intensification in people with type 2 diabetes mellitus. Diabetes Metab 2017;43:501–11. https://doi.org/10.1016/j.diabet.2017.06.003.

[16] Bailey CJ. Under-treatment of type 2 diabetes: Causes and outcomes of clinical inertia. Int J Clin Pract 2016;70:988–95. https://doi.org/10.1111/ijcp.12906.

[17] Cucinotta D, Nicolucci A, Giandalia A, Lucisano G, Manicardi V, Mannino D, et al. Temporal trends in intensification of glucose-lowering therapy for type 2 diabetes in Italy: Data from the AMD Annals initiative and their impact on clinical inertia. Diabetes Res Clin Pract 2021;181:109096. https://doi.org/10.1016/j.diabres.2021.109096.

[18] Selby JV, Karter AJ, Ackerson LM, Ferrara A, Liu J. Developing a prediction rule from automated clinical databases to identify high-risk patients in a large

population with diabetes. Diabetes Care 2001;24:1547–55. https://doi.org/10.2337/diacare.24.9.1547.
[19] Hayes AJ, Leal J, Gray AM, Holman RR, Clarke PM. UKPDS outcomes model 2: a new version of a model to simulate lifetime health outcomes of patients with type 2 diabetes mellitus using data from the 30 year United Kingdom Prospective Diabetes Study: UKPDS 82. Diabetologia 2013;56:1925–33. https://doi.org/10.1007/s00125-013-2940-y.
[20] Parrinello CM, Matsushita K, Woodward M, Wagenknecht LE, Coresh J, Selvin E. Risk prediction of major complications in individuals with diabetes: the Atherosclerosis Risk in Communities Study. Diabetes Obes Metab 2016;18:899–906. https://doi.org/10.1111/dom.12686.
[21] Aminian A, Zajichek A, Arterburn DE, Wolski KE, Brethauer SA, Schauer PR, et al. Predicting 10-year risk of end-organ complications of type 2 diabetes with and without metabolic surgery: a machine learning approach. Diabetes Care 2020;43(4):852–9.
[22] Song X, Waitman LR, Yu AS, Robbins DC, Hu Y, Liu M. Longitudinal risk prediction of chronic kidney disease in diabetic patients using a temporal-enhanced gradient boosting machine: retrospective cohort study. JMIR Med Inf 2020;8:e15510. https://doi.org/10.2196/15510.
[23] Makino M, Yoshimoto R, Ono M, Itoko T, Katsuki T, Koseki A, et al. Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning. Sci Rep 2019;9(1). https://doi.org/10.1038/s41598-019-48263-5.
[24] Brisimi TS, Xu T, Wang T, Dai W, Paschalidis IC. Predicting diabetes-related hospitalizations based on electronic health records. Stat Methods Med Res 2019;28:3667–82. https://doi.org/10.1177/0962280218810911.
[25] Dagliati A, Marini S, Sacchi L, Cogni G, Teliti M, Tibollo V, et al. Machine learning methods to predict diabetes complications. J Diabetes Sci Technol 2018;12(2):295–302.
[26] Ravaut M, Sadeghi H, Leung KK, Volkovs M, Kornas K, Harish V, et al. Predicting adverse outcomes due to diabetes complications with machine learning using administrative health data. NPJ Digit Med 2021;4(1).
[27] Segar MW, Vaduganathan M, Patel KV, McGuire DK, Butler J, Fonarow GC, et al. Machine learning to predict the risk of incident heart failure hospitalization among patients with diabetes: the WATCH-DM risk score. Diabetes Care 2019;42(12):2298–306.
[28] Rodriguez-Romero V, Bergstrom RF, Decker BS, Lahu G, Vakilynejad M, Bies RR. Prediction of nephropathy in type 2 diabetes: an analysis of the ACCORD trial applying machine learning techniques. Clin Transl Sci 2019;12:519–28. https://doi.org/10.1111/cts.12647.
[29] Hawkins DM. The problem of overfitting. Journal of chemical information and computer sciences. J Chem Inf Comput Sci 2004;44(1):1–12.
[30] Japkowicz N, Stephen S. The class imbalance problem: A systematic study. Intelligent data analysis Intelligent Data Analysis 2002;6:429–49. https://dl.acm.org/doi/10.5555/1293951.1293954.
[31] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research JAIR 2002;16:321–57. https://doi.org/10.1613/jair.953.
[32] Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, et al. Xgboost: extreme gradient boosting. R package version 04-2 2015;1(4):1–4.
[33] Bernardini M, Romeo L, Misericordia P, Frontoni E. Discovering the type 2 diabetes in electronic health records using the sparse balanced support vector machine. IEEE J Biomed Health Inform 2019;24:235–46. https://doi.org/10.1109/JBHI.2019.2899218.
[34] Jeong Y, Hong Y-J, Han J-H. Review of Machine Learning Applications Using Retinal Fundus Images. Diagnostics (Basel) 2022;12:134. oi:;12(1):134.