diabetes research and clinical practice 204 (2023) 110910

Contents lists available at ScienceDirect

Diabetes Research and Clinical Practice

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

Predicting the onset of diabetes-related complications after a diabetes diagnosis with machine learning algorithms

Toni Mora [a], David Roche [a], Beatriz Rodríguez-Sánchez [b,*]

[a] Research Institute for Evaluation and Public Policies (IRAPP), Universitat Internacional de Catalunya (UIC), Carrer de la Immaculada, 22, 08017 Barcelona, Spain
[b] Applied Economics, Public Economics and Political Economy, Faculty of Law, Universidad Complutense de Madrid, Plaza Menéndez Pelayo, 4, 28040 Madrid, Spain

ARTICLE INFO

Keywords:
Diabetes mellitus
Diabetes-related complications
Machine learning
Deep learning
Administrative data

ABSTRACT

Aims: Using machine learning algorithms and administrative data, we aimed to predict the risk of being diagnosed with several diabetes-related complications after one-, two- and three-year post-diabetes diagnosis.

Methods: We used longitudinal data from administrative registers of 610,019 individuals in Catalonia with a diagnosis of diabetes and checked the presence of several complications after diabetes onset from 2013 to 2017: hypertension, renal failure, myocardial infarction, cardiovascular disease, retinopathy, congestive heart failure, cerebrovascular disease, peripheral vascular disease and stroke. Four different machine learning (ML) algorithms (logistic regression (LR), Decision tree (DT), Random Forest (RF), and Extreme Gradient Boosting (XGB)) will be used to assess their prediction performance and to evaluate the prediction accuracy of complications changes over the period considered.

Results: 610,019 people with diabetes were included. After three years since diabetes diagnosis, the area under the curve values ranged from 60% (retinopathy) to 69% (congestive heart failure), whereas accuracy rates varied between 60% (retinopathy) to 75% (hypertension). RF was the most relevant technique for hypertension, myocardial and retinopathy, and LR for the rest of the comorbidities. The Shapley additive explanations values showed that age was associated with an elevated risk for all diabetes-related complications except retinopathy. Gender, other comorbidities, co-payment levels and age were the most relevant factors for comorbidity diagnosis prediction.

Conclusions: Our ML models allow for the identification of individuals newly diagnosed with diabetes who are at increased risk of developing diabetes-related complications. The prediction performance varied across complications but within acceptable ranges as prediction tools.

1. Introduction

The burden of Diabetes has been ranked as the eighth cause in the ranking of causes of disability-adjusted life years (DALYs) all over the world [1,2], accounting for 2.8 % of total DALYs and with an increase of more than 148 % in 2019 as compared with the data obtained in 1990 [2]. Diabetes is one of the most significant factors increasing the risk of mortality, morbidity, and disability worldwide, and its economic burden demands new ways to curb diabetes healthcare expenditure [3]. Furthermore, people with diabetes are at greater risk of additionally suffering from cardiovascular diseases, such as heart attack or stroke [4–6]; kidney failure [4]; foot ulcers that might lead to amputation [7,8]; and functional impairment [9].

The literature has widely analysed the impact of micro and macrovascular diseases on people with diabetes quality of life [10,11], concluding that the quality of life in people with diabetes is affected by complications and not by diabetes itself [12], with differences in terms of quantitative impact between diabetes-related complications [10,12–14]. Given the negative burden that diabetes-related complications might bear on an individual's quality of life, mortality [2,15] and economic burden [16–18], identifying those individuals at higher risk of diabetes mellitus progression might ease targeted prevention programs [19,20].

Artificial intelligence (AI) can improve diabetes mellitus care and diabetes-related complications diagnosis, including improved glucose monitoring and automated insulin delivery [21–23]. Three key areas in which AI finds widespread use. First, its predictive analytics to identify patients who are at high risk for developing diabetes or its

* Corresponding author.
E-mail address: brodriguezsanchez@ucm.es (B. Rodríguez-Sánchez).

https://doi.org/10.1016/j.diabres.2023.110910
Received 12 April 2023; Received in revised form 1 August 2023; Accepted 15 September 2023
Available online 16 September 2023

complications. By analysing patient data such as age, gender, family history, and other health factors, AI can identify patients who are at risk and provide early intervention to prevent or delay the onset of diabetes or its complications. Moreover, AI can help detect the early signs of diabetes-related complications such as diabetic retinopathy, neuropathy, and myocardial infarction through the analysis of large amounts of medical images and identify any abnormalities that could indicate the presence of these complications. Lastly, AI can help create personalised treatment plans for diabetes patients based on their health data. AI algorithms can analyse data from electronic health records, glucose monitoring devices, and other sources to identify patterns and recommend personalised treatment plans. Thus, timely and accurate prediction of complications could help implement more specific and targeted measures, potentially preventing or slowing down their development. Consequently, slowing down the growth of complications would save significant economic resources needed for their treatment.

Hence, this paper aims to predict whether a list of nine diabetes-related complications (hypertension, renal failure, myocardial infarction, cardiovascular disease, retinopathy, congestive heart failure, cerebrovascular disease, peripheral vascular disease and stroke) will develop after one, two and three years since diabetes onset, using four different machine learning algorithms (logistic regression (LR), Decision tree (DT), Random Forest (RF), and Extreme Gradient Boosting (XGB)). Moreover, we can assess the prediction performance of these four AI approaches and evaluate the prediction accuracy of complications that might change over the period considered in case of varying time spans.

2. Research design and methods

2.1. Data sources and study population

We used a large administrative dataset from the Agency for Health Quality and Assessment of Catalonia (AQuAS), which combines information from several providers, although considering different periods, for the Catalan adult population from 2013 to 2017, resulting in 610,019 observations. The Ethical Review Board approved the study in Hospital Clínic, Barcelona (Spain). This database contains primary care, hospitalisations, emergency care, mental health hospitalisation, and community mental health care from 2013 to 2017. Note that the Spanish Health System provides universal care. In these files it is collected the individual identifier, the health care provider unit (N = 484), the date of the visit (length in case of hospitalisations), and all the diagnoses (N = 2,810) and procedures (N = 463) that were administered. In Catalonia, the International Classification of Diseases (ICD-9) diagnostic manual was used for diagnostic purposes. Diagnoses are shown in an ordinal sense, indicating the primary diagnosis for that visit and a list of secondary diagnoses. Via unique personal identifiers, the information is linked between all provider's datasets but also to some demographic information: gender, age, drug co-payment level, which is related to the socioeconomic status of their parents, individual nationality, date of decease and the sanitary health region they belong to.

2.2. Problem setting and data pre-processing

Each patient corresponds to one observation or row in our dataset. For each patient, we have demographic information about gender, socioeconomic level, co-payment level, nationality, and sanitary health region. Moreover, information on diagnoses and procedures is retained at the appearance date.

Because original data collection is in the event-row format, a pre-processing phase is mandatory. Initially, each row corresponds to patient identification, date, and event, with repeated individual demographic variables. Event information was codified to unique details to obtain predictors. For example, the event-row date (event information) was transformed into personal information creating two new individual variables for each patient: the overall number of visits during the period and the standard deviation of visit dates. Then, nominal variables (diagnoses, procedures, health care, and provider unit) were one-hot encoded. If one variable has n categories, n new variables were created, assigning one if the patient has that category and 0 if not. For instance, the economic level, initially, was a categorical variable with four classes and was transformed into four dichotomous variables. Continuous variables were kept continuous (age, number of visits and the standard deviation of visit dates).

Next, we build a dummy variable representing the presence of the accounted comorbidity (hypertension, renal failure, myocardial infarction, cardiovascular disease, retinopathy, congestive heart failure, cerebrovascular disease, peripheral vascular disease and stroke[1]) after one-, two- and three-years post-diabetes diagnosis. For each comorbidity analysis, we deleted individuals diagnosed with these comorbidities before the diabetes diagnosis date. After this processing for the three years, the final number of considered individuals was, concerning each comorbidity aforementioned: 432,924; 485,747; 487,567; 489,419; 492,648; 490,762; 490,788; 491,911; and 491,222; respectively. We carried out the analysis ~~to~~ for those with available data to cover one to three years in the post-diagnosed period (2013–2017).

2.3. Machine learning algorithms

Once data is pre-processed, a battery of models is performed. Following the standard procedure of previous empirical analyses implementing AI algorithms, logistic regression (LR), Decision tree (DT), Random Forest (RF), and Extreme Gradient Boosting (XGB) were performed. For interpretability reasons, given that it is considered a baseline algorithm, we considered the logistic approach in the stacked model. The same reason relies on using the DT based on tree-based algorithms. Then, we considered more complex tree-based algorithms that avoid overfitting issues (RF and XGB). An 80/20 split of train/test samples was carried out for each model. Next, we standardised all variables, and each model was tuned concerning its hyperparameters. These hyperparameters allow control of the learning process and should be set before the algorithm's implementation. Cross-validation was used in the train set for this hyperparameter setting purpose with a grid search paradigm. For DT, RF and XGB, a regular grid was computed with five folds. Cross-validation is a known technique to avoid over-fitting and over-performed evaluation in ML techniques dividing the dataset into k folds and computing k evaluations using one fold for testing and the other to train in each iteration. Finally, a stacked model based on logistic regression was performed to combine our four models for better performance.

LR is the most common medical research technique for solving classification problems using binary outcomes. DT is, among other applications, a classification ML task based on the divide and conquer strategy. The process iterates by splitting the independent variables into subsets repeatedly divided into smaller subsets until a stop condition is met. To set which variable is split, Information Gain (IG) measuring the quality of each possibility is computed, and a variable with maximum IG is selected in each iteration. RF combines several decision trees, a restriction over what variables and cases are made based on hyperparameters setting. The final model classifies instances averaging over all tree classifications. This technique overperforms a simple decision tree to avoid over-fitting, thus better generalising our training data to unseen data. XGB is an ensemble classification algorithm based on a combination of several weak trees in a sequential form. The critical point is that each tree deals with cases poorly classified in the previous tree during the process. Often, this scheme outperforms other techniques

[1] Diabetes-related comorbidities were included depending on their prevalence, as Figure A1, Appendix, shows. Dementia, neuropathy, heart failure and other vascular diseases were not included given their low prevalence at the long timespan (3 years after diabetes diagnosis).

with better generalisations. A stacked model is an ensemble machine learning algorithm that learns how to best combine the predictions from multiple well-performing individual machine learning models. Strong calibration [24] was implemented for the stacked models.

3. Results

Table A1, Appendix, shows the main descriptive sociodemographic variables and unhealthy behaviours (smoking, drinking and BMI categories). The average age in our sample is 69.56 years old (SD 13.30), although diabetes-related complications observe higher average ages, as Figure A2, Appendix, shows. 45.67 % are women, 93.68 % are Spaniards, and 74 % have the lowest co-payment level (10 %). 33.94 and 27.51 % are overweight or obese, respectively, whereas 84.78 % are non-drinkers, and 27.32 % of the sample are current smokers.

Considering pre-processed steps and the best hyperparameters obtained from cross-validation techniques, the results from applying each algorithm to the test set are shown in Fig. 1. In the final stacked model, all the measures (the area under the curve (AUC), accuracy and precision) were above 63 % general performance, regardless of the diabetes-related complication. The AUC measure uses a receiver operating curve's characteristic to capture the trade-offs between the actual and false-positive rates. Values close to one are preferable. Accuracy constitutes the percentage of correctly predicted data from all the test sets. In contrast, precision is related to the rate of correctly predicted data within the positive values (presence of the comorbidity). We tested for sensitivity to the chosen period before the first diagnosis of the comorbidity. Indeed, we considered three alternative spans: 1, 2 and 3 years. Fig. 1 shows these performances. The performance hardly changed according to the information length. It is better to account for further information (and diagnoses) in a clinical sense than to rely on shorter clinical spans.

Concretely, for the hypertension analysis, which was the more prevalent diabetes-related complication (Figure A1, Appendix), accuracy was 75 %, whereas AUC and precision were 66 % and 77 %, respectively. In our case, a value of 75 % indicated a moderately excellent performance model, which corroborates the rest of the metrics information. The AUC was similar for the other comorbidities, ranging from 60 % for retinopathy to 69 % for congestive heart failure.

Next, we computed the relevance of each ML approach for the

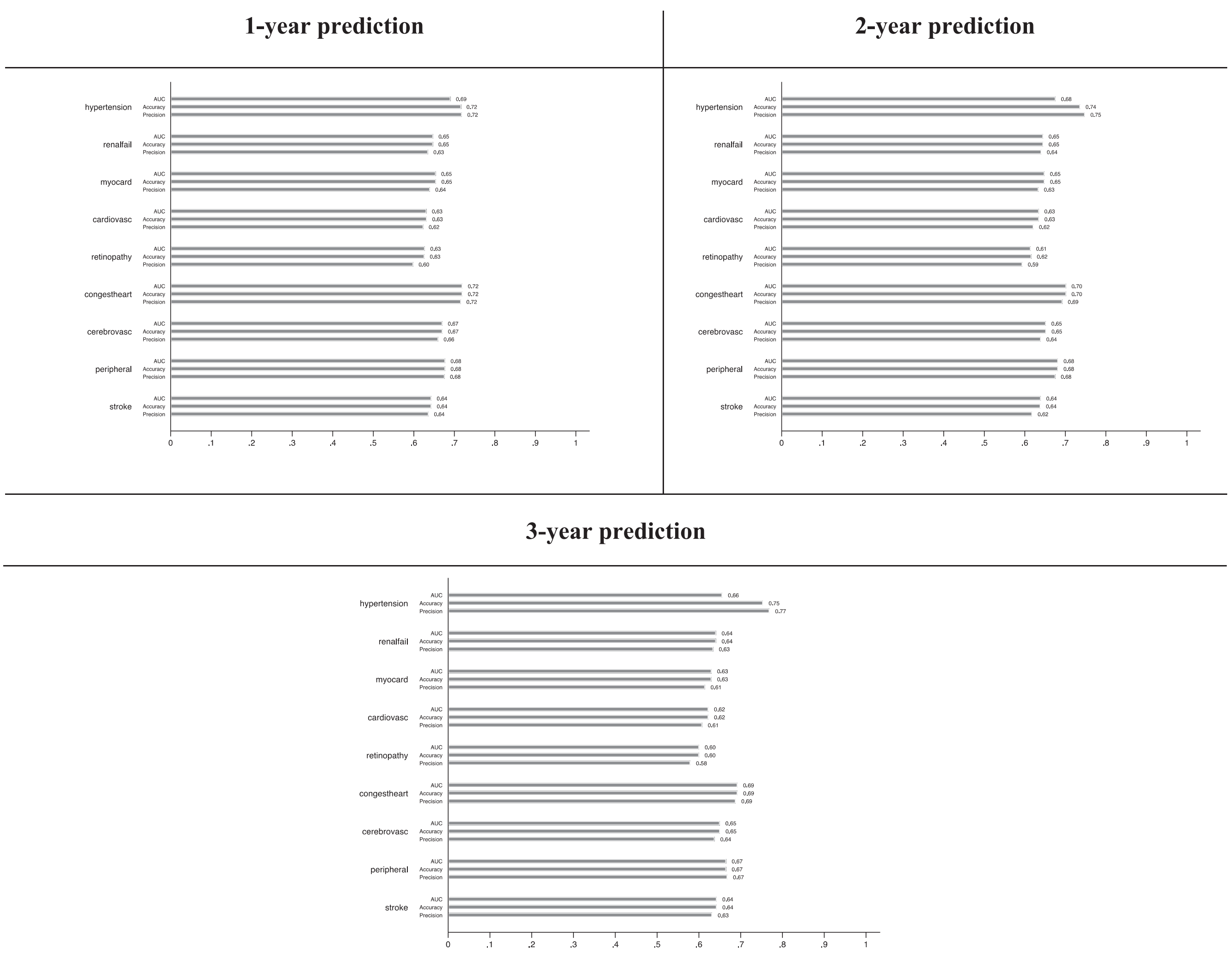

Fig. 1. Performance measures for stacked models and every comorbidity: 1-, 2- and 3-year predictions.
Note: "hypertension" stands for hypertension; "renalfail" for renal failure; "myocard" for myocardial infarction; "cardiovasc" for cardiovascular disease; "retinopathy" for retinopathy; "congestheart" for congestive heart failure; "cerebrovasc" for cerebrovascular disease; "peripheral" for peripheral vascular disease; and "stroke" for stroke.

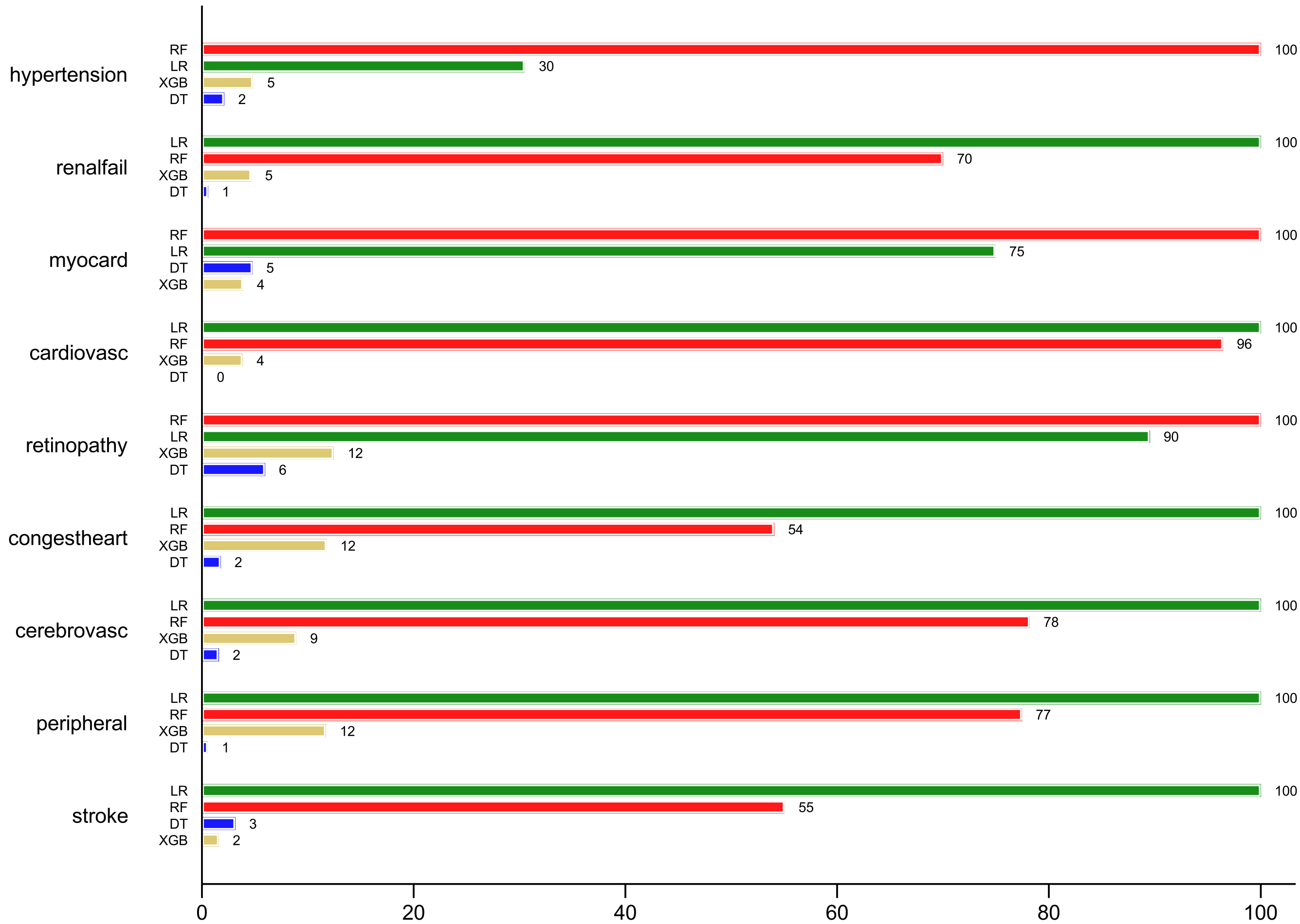

Fig. 2. Algorithm relevance within the stacked model for each comorbidity for a 3-year span.
Note: "hypertension" stands for hypertension; "renalfail" for renal failure; "myocard" for myocardial infarction; "cardiovasc" for cardiovascular disease; "retinopathy" for retinopathy; "congestheart" for congestive heart failure; "cerebrovasc" for cerebrovascular disease; "peripheral" for peripheral vascular disease; and "stroke" for stroke.

stacked model. RF was the most relevant technique for hypertension, myocardial and retinopathy, whereas the best model for the rest of the comorbidities was the LR. Fig. 2 shows the relevance of each algorithm as part of the final stacked model considered for each comorbidity for the three years case. We do not report findings for 1-year and 2-year spans for redundancy reasons.

To detect the variables with the highest predicting power, an importance variable rank was computed for every algorithm. Fig. 3 depicts the contribution of the essential variables obtained from the weighted relevance of each model in the stacked one to the model's performance. A closer value to 100 indicates the highest variable's importance on performance. The different techniques found that the following variables were among the four more relevant factors to predict the presence of a comorbidity diagnosis: (i) gender; (ii) some diagnoses; (iii) copayment levels; and (iv) age.

Next, Fig. 4 displays the computed Shapley additive explanations values (SHAP) for the three main relevant covariates for the 3-year timespan and each comorbidity. These values indicate how a change in a covariate would alter the prediction. SHAP values in the graphs represent each individual with a point in two different colours. Blue points indicate lower values in the covariate, whereas red ones represent the higher values for covariates. As for the position in abscise, it constitutes the probability of having the specific comorbidity. The higher the SHAP value, the greater the probability of that comorbidity. Likewise, the order in the y-axis is in descending order. That is, the higher the position, the more relevant for the computation of the SHAP values. Feature effects come mainly from individuals' age and copayment levels. Still, other variables related to unhealthier habits also showed some relevance (BMI and alcohol consumption).

4. Conclusions

Our study used administrative data to detect the most often diabetes-related complications along different time spans after a diabetes diagnosis. The results showed that RF outperformed the other machine learning algorithms for hypertension, myocardial and retinopathy, whereas, for the rest of the comorbidities, the best model was the LR. The four machine learning methods showed high predicting power, with AUC values ranging from 60 % to a maximum of 69 %, depending on the diabetes-related complication considered.

To our knowledge, no prediction models for micro- and macro-vascular complications exist for individuals considering different time spans after a diabetes diagnosis; hence, a comparison with prior work is impossible. Still, some authors have already used prediction models for individuals with diabetes. The most similar work was that performed by Schallmoser et al. (2023) [25]. They used machine learning models to predict the risk of developing three micro- (retinopathy, nephropathy and neuropathy) and three macro-vascular (peripheral vascular disease, cerebrovascular and cardiovascular diseases) complications after five years since a diagnosis of diabetes or pre-diabetes. Although the

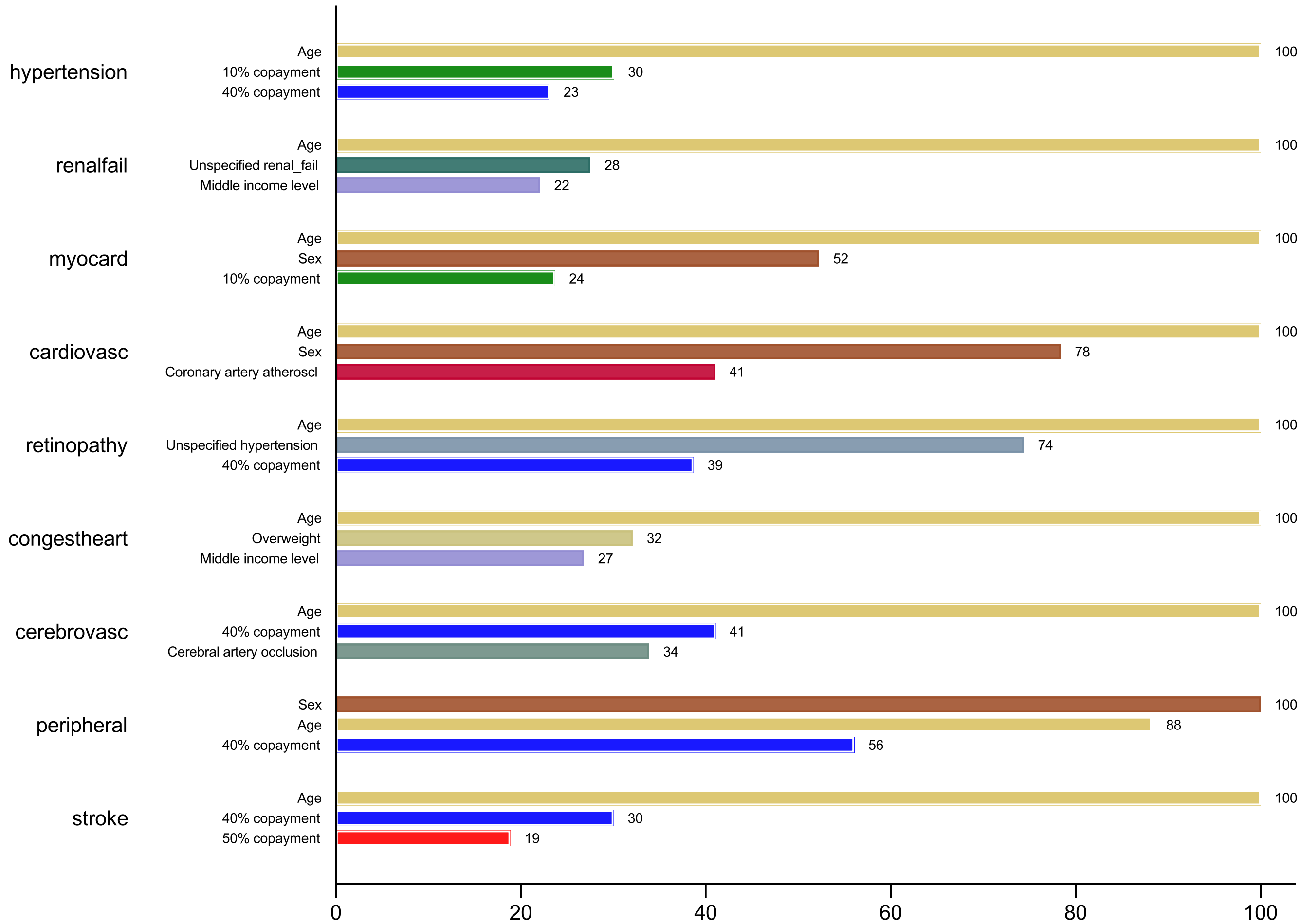

Fig. 3. Main three relevant covariates for each comorbidity.
Note: "hypertension" stands for hypertension; "renalfail" for renal failure; "myocard" for myocardial infarction; "cardiovasc" for cardiovascular disease; "retinopathy" for retinopathy; "congestheart" for congestive heart failure; "cerebrovasc" for cerebrovascular disease; "peripheral" for peripheral vascular disease; and "stroke" for stroke.

forecasted time differed (they only considered the period of five years, and we considered one, two and three years since diabetes diagnosis), our prediction models for cardiovascular diseases (myocardial infarction and congestive heart failure) would potentially outperform their models at 5-years since figures are already similar at our three-year timespan to theirs (0.69). Their models did better than ours for retinopathy. Still, the results are not fully comparable given the different time frames and the fact that they grouped diabetes-related macro-vascular complications, limiting the comparison at disaggregated levels.

Similarly, Ljubic et al. (2020) [26] modelled the prediction risk for ten diabetes-related complications over nine years. In addition to considering a longer time span, comparability is restricted since performance measures are based on a potentially sicker group of people with diabetes (four visits to hospitals between diabetes diagnosis and diabetes-related complications) who are at higher risk of developing any health problem. Moreover, machine learning algorithms differ (Ljubic et al. report only the performance metrics for recurrent neural networks). Dagliati et al. (2018) [27] focused only on microvascular complications after the first visit to the hospital, not since diabetes diagnosis as we do, which outperformed our results when balancing the algorithms, but not in the raw ones. Moreover, our results confirm the role that some individual characteristics might pose on the onset of some comorbidities, such as the role of age [28,29] with respect to the development of macrovascular complications.

Overall, the prediction performance of our models for individuals with diabetes is somehow comparable to the performances reported in prior work, with differences between studies explaining differences and providing new evidence in the existing literature. As recent systematic review literature showed (Gosak et al., 2022) [30], work on applying ML algorithms to predict the risk of diabetes-related complications is very scarce. Hence, one of the advantages of our study is that we are the first to account for different time spans after a diabetes diagnosis, which allows us to set short- and medium-term prediction risk models. Furthermore, the availability of administrative data, which refers to a representative population, allows for a large set of individual information (sociodemographics, biomarkers, comorbidities and other items included in the clinical history). The information provided here could be especially relevant for the clinical setting since it allows for early identification of individuals at risk who could benefit from prompt treatment, increasing the odds of preventing or delaying diabetes-related complications onset.

This is the first paper that applies different machine learning methods to study their prediction performance of diabetes-related complications after a diabetes diagnosis and whether it persists over different short-term periods. Moreover, our analysis is enriched using a large administrative dataset, allowing us to split the sample into nine diabetes comorbidities after a diabetes diagnosis. However, some limitations should be mentioned. The main limitations of our study are

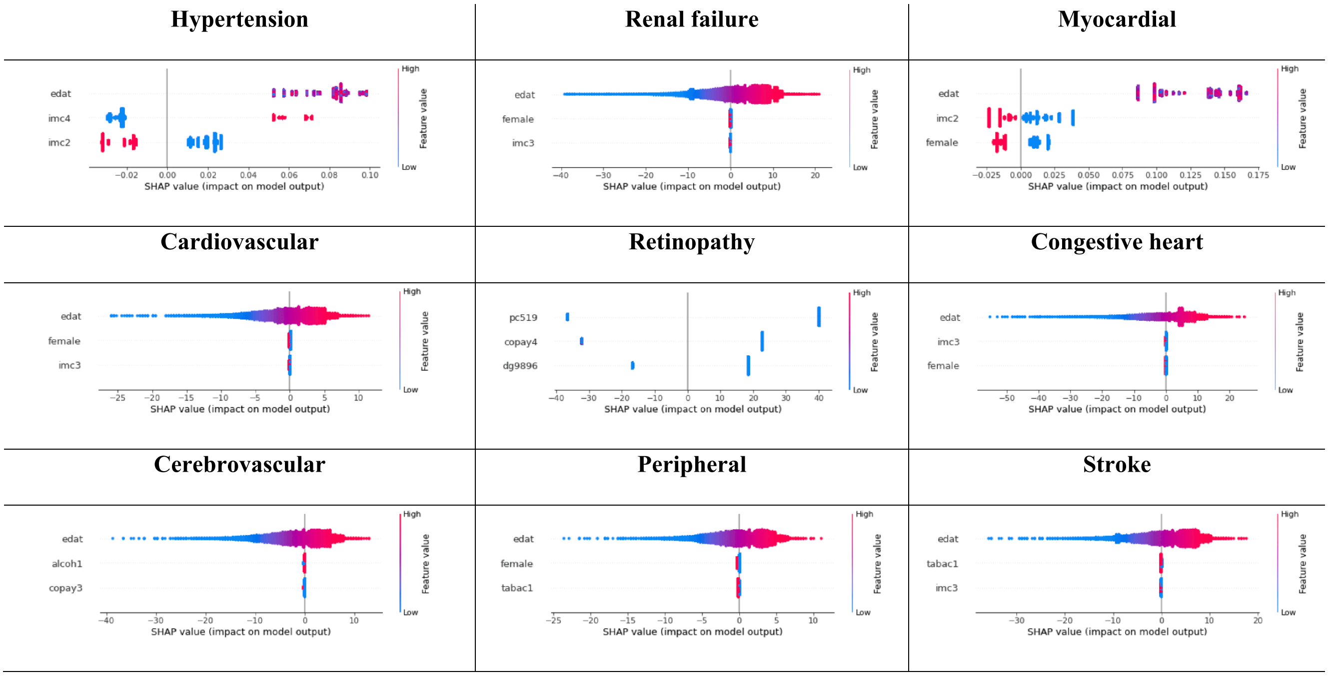

Fig. 4. Shapley values for 3-year predictions and the best model for each comorbidity.
Note: notation of the predictors corresponds to edat (age), imc2 (normal weight), imc3 (overweight), imc4 (obese), tabac1 (no smoker), alcoh1 (not drinker), copay3 (40% drug co-payment level), copay4 (50% drug co-payment level), and female (being female). pc519 refers to a particular procedure related to cataracts, and dg9896 represents a visit for a non-specific generic medical exam.

related to the availability of data, mainly due to the lack of data on more extended follow-up periods to check whether the pattern observed within our timeframe remains in the long term. Likewise, the lack of a longer follow-up period in case of a possible change in the characteristics of the diabetic population may lead to a dataset shift bias for the models' applicability. Second, we did not include information on blood glucose measurements since this information was not available for all individuals and could contain measurement errors. For this purpose, we ran sensitivity analyses for the three-year period and the most important complications, and there was no impact on previous results. Third, our analysis is restricted to public healthcare use. However, it is common practice among private hospital patients to pick up prescriptions from primary public centres. This is probably true for moderate to severe patients; we may still lose the milder ones that might be using only private resources.

Overall, AI has the potential to improve diabetes care and help patients manage their condition more effectively to prevent or delay diabetes progression into the development of diabetes-related complications. However, it's important to note that AI is not a substitute for medical advice or treatment, and patients should always consult with their healthcare providers for any medical concerns.

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.

Data availability

The data that support the findings of this study are available from the Agency for Health Quality and Assessment of Catalonia but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of the Agency for Health Quality and Assessment of Catalonia.

Acknowledgements

We thank Antonio Valero from Hospital Clínic of Barcelona and the Agency for Health Quality and Assessment of Catalonia for the access to the dataset. Toni Mora and David Roche gratefully acknowledge the financial support from the PID2021-124067OB-C21.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.diabres.2023.110910.

References

[1] Safiri S, Karamzad N, Kaufman JS, Bell AW, Nejadghaderi SA, Sullman MJ, et al. Prevalence, deaths and disability-adjusted-life-years (DALYs) due to type 2 diabetes and its attributable risk factors in 204 countries and territories, 1990–2019: results from the global burden of disease study 2019. Front Endocrinol 2022;13:98.
[2] Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global Burden of Disease study 2019. Lancet 2020;396 (10258):1204–22.
[3] de Lagasnerie G, Aguadé AS, Denis P, Fagot-Campagna A, Gastaldi-Menager C. The economic burden of diabetes to french national health insurance: a new cost-of-illness method based on a combined medicalized and incremental approach. Eur J Health Econ 2018;19(2):189–201.
[4] OECD/EU, 2022. Health at a Glance: Europe 2022 – State of Health in the EU Cycle. OECD Publishing, Paris.
[5] Engelmann J, Manuwald U, Rubach C, Kugler J, Birkenfeld AL, Hanefeld M, et al. Determinants of mortality in patients with type 2 diabetes: a review. Rev Endocr Metab Disord 2016;17:129–37.
[6] Constantino MI, Molyneaux L, Limacher-Gisler F, Al-Saeed A, Luo C, Wu T, et al. Long-term complications and mortality in young-onset diabetes: type 2 diabetes is more hazardous and lethal than type 1 diabetes. Diabetes Care 2013;36(12): 3863–9.
[7] Lombardo FL, Maggini M, De Bellis A, Seghieri G, Anichini R. Lower extremity amputations in persons with and without diabetes in italy: 2001–2010. PLoS One 2014;9(1):e86405.
[8] Chand G, Mishra AK, Kumar S, Agarwal A. Diabetic foot. Clinical queries. Nephrology 2012;1(2):144–50.
[9] Sinclair A, Dunning T, Rodriguez-Mañas L. Diabetes in older people: new insights and remaining challenges. Lancet Diabetes Endocrinol 2015;3(4):275–85.

[10] Shao H, Yang S, Fonseca V, Stoecker C, Shi L. Estimating quality of life decrements due to diabetes complications in the united states: the health utility index (HUI) diabetes complication equation. Pharmacoeconomics 2019;37:921–9.
[11] Vaidya V, Gangan N, Sheehan J. Impact of cardiovascular complications among patients with type 2 diabetes mellitus: a systematic review. Expert Rev Pharmacoecon Outcomes Res 2015;15(3):487–97.
[12] Venkataraman K, Wee HL, Leow MKS, Tai ES, Lee J, Lim SC, et al. Associations between complications and health-related quality of life in individuals with diabetes. Clin Endocrinol 2013;78(6):865–73.
[13] Jacobson AM, Braffett BH, Cleary PA, Gubitosi-Klug RA, Larkin ME, DCCT/EDIC research group. The long-term effects of type 1 diabetes treatment and complications on health-related quality of life: a 23-year follow-up of the Diabetes Control and Complications/Epidemiology of Diabetes Interventions and Complications cohort. Diabetes Care 2013;36(10):3131–8.
[14] Solli O, Stavem K, Kristiansen IS. Health-related quality of life in diabetes: the associations of complications with EQ-5D scores. Health Qual Life Outcomes 2010; 8(1):1–8.
[15] Susan van D, Beulens JW, Yvonne T. van der S, Grobbee DE, Nealb B. The global burden of diabetes and its complications: an emerging pandemic. Eur J Cardiovasc Prev Rehabil 2010;17(1_suppl):s3–8.
[16] Williams R, Karuranga S, Malanda B, Saeedi P, Basit A, Besançon S, et al. Global and regional estimates and projections of diabetes-related health expenditure: Results from the international diabetes federation diabetes atlas. Diabetes Res Clin Pract 2020;162:108072.
[17] American Diabetes Association. Economic costs of diabetes in the US in 2012. Diabetes Care 2013;36(4):1033–46.
[18] Herman WH. The economic costs of diabetes: is it time for a new treatment paradigm? Diabetes Care 2013;36(4):775–6.
[19] Bailey CJ, Day C. Treatment of type 2 diabetes: future approaches. Br Med Bull 2018;126(1):123–37.
[20] Cahn A, Shoshan A, Sagiv T, Yesharim R, Raz I, Goshen R. Use of a machine learning algorithm improves prediction of progression to diabetes. Diabetes 2018; 67(Supplement_1).
[21] Makroum MA, Adda M, Bouzouane A, Ibrahim H. Machine learning and smart devices for diabetes management: Systematic review. Sensors 2022;22(5):1843.
[22] Ellahham S. Artificial intelligence: the future for diabetes care. Am J Med 2020;133 (8):895–900.
[23] Contreras I, Vehi J. Artificial intelligence for diabetes management and decision support: literature review. J Med Internet Res 2018;20(5):e10775.
[24] Van Calster B, McLernon DJ, Van Smeden M, Wynants L, Steyerberg EW. Calibration: the achilles heel of predictive analytics. BMC Med 2019;17(1):1–7.
[25] Schallmoser S, Zueger T, Kraus M, Saar-Tsechansky M, Stettler C, Feuerriegel S. Machine learning for predicting micro-and macrovascular complications in individuals with prediabetes or diabetes: Retrospective cohort study. J Med Internet Res 2023;25:e42181.
[26] Ljubic B, Hai AA, Stanojevic M, Diaz W, Polimac D, Pavlovski M, et al. Predicting complications of diabetes mellitus using advanced machine learning algorithms. J Am Med Inform Assoc 2020;27(9):1343–51.
[27] 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.
[28] 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 Digital Med 2021;4(1):24.
[29] Vimont A, Béliard S, Valéro R, Leleu H, Durand-Zaleski I. Prognostic models for short-term annual risk of severe complications and mortality in patients living with type 2 diabetes using a national medical claim database. Diabetol Metab Syndr 2023;15(1):1–10.
[30] Gosak L, Martinović K, Lorber M, Stiglic G. artificial intelligence based prediction models for individuals at risk of multiple diabetic complications: A systematic review of the literature. J Nurs Manag 2022;30(8):3765–76.