Received: 27 March 2025 | Revised: 18 June 2025 | Accepted: 27 June 2025

DOI: 10.1111/dom.16598

ORIGINAL ARTICLE

WILEY

Prediction of glycaemic control and quality of life in people with type 2 diabetes using glucose-lowering drugs with machine learning—The Maastricht study

Nikki C. C. Werkman PhD[1,2] | Johannes T. H. Nielen PhD[1,2] | José Tapia-Galisteo PhD[3,4] | Francisco J. Somolinos-Simón PhD[3] | Maria Elena Hernando PhD[3,4] | Junfeng Wang PhD[5] | Li Jiu PhD[5] | Wim G. Goettsch PhD[5,6] | Hans Bosma PhD[7] | Miranda T. Schram PhD[1,8,9,10] | Marleen M. J. van Greevenbroek PhD[1,8] | Anke Wesselius PhD[11] | Coen D. A. Stehouwer PhD[1,8] | Johanna H. M. Driessen PhD[1,2] | Gema Garcia-Sáez PhD[3,4]

[1]Department of Clinical Pharmacy, Cardiovascular Research Institute Maastricht (CARIM), Maastricht University, Maastricht, The Netherlands

[2]Department of Clinical Pharmacy and Toxicology, Maastricht University Medical Center+, Maastricht, The Netherlands

[3]Bioengineering and Telemedicine Group, Centro de Tecnología Biomédica, ETSI de Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain

[4]CIBER-BBN: Networking Research Center for Bioengineering, Biomaterials and Nanomedicine, Madrid, Spain

[5]Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands

[6]National Health Care Institute, Diemen, The Netherlands

[7]School for Public Health and Primary Care (CAPHRI), Maastricht University, Maastricht, The Netherlands

[8]Department of Internal Medicine, Maastricht University Medical Center+, Maastricht, The Netherlands

[9]School for Mental Health & Neuroscience, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands

[10]Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands

[11]School for Nutrition and Translational Research in Metabolism (NUTRIM), Maastricht University, Maastricht, The Netherlands

Correspondence
Johannes T. H. Nielen, Department of Clinical Pharmacy, Cardiovascular Research Institute Maastricht (CARIM), Maastricht University, Maastricht, The Netherlands.
Email: yannick.nielen@mumc.nl

Funding information
European Regional Development Fund via OP-Zuid; Province of Limburg, the Dutch Ministry of Economic Affairs, Grant/Award Number: 31O.041; Stichting De Weijerhorst; Pearl String Initiative Diabetes; Cardiovascular Center; CARIM School for Cardiovascular Diseases; CAPHRI Care and Public Health Research Institute; NUTRIM School for Nutrition and Translational Research in Metabolism; Stichting Annadal; Health Foundation Limburg; unrestricted grants from

Abstract

Background: Despite the heterogeneity of type 2 diabetes (T2D), all patients are treated according to the same guideline. Some people have more difficulty reaching treatment goals (adequate glycaemic control) and maintaining quality of life (QoL). Therefore, a prediction model identifying who is unlikely to reach these goals within the next year would be useful to allow specific attention to these people.

Aim: To investigate if machine learning algorithms can predict which individuals are unlikely to reach glycaemic control and likely to deteriorate in QoL in 1 year.

Methods: We used data from The Maastricht Study, including 842 people with T2D and information on HbA1c values, and 964 people with T2D and information on QoL. We evaluated several machine learning algorithms with feature selection methods

Janssen-Cilag B.V.; Novo Nordisk Farma B.V.; Sanofi-Aventis Netherlands B.V.

and hyperparameter tuning in fivefold cross-validation for the corresponding outcomes.

Results: The prediction of inadequate glycaemic control showed good performance. The support vector machine classifier performed best in terms of accuracy (0.76 (95% CI 0.71–0.79)), precision (0.79 (95% CI 0.71–0.83)) and area under the receiver operating characteristic curve (AUC-ROC) (0.85 (95% CI 0.80–0.89)). The multi-layer perceptron classifier performed best in terms of recall (0.72 (95% CI 0.64–0.79)) and F1-score (0.73 (95% CI 0.64–0.79)). The prediction of deterioration in QoL showed inadequate performance and did not seem feasible.

Conclusion: Prediction of glycaemic control after 1 year in T2D is feasible with good model performance. However, the prediction of deterioration in QoL remains a challenge and needs further work.

KEYWORDS
antidiabetic drug, glycaemic control, observational study, type 2 diabetes

1 | INTRODUCTION

There are roughly half a billion people living with type 2 diabetes (T2D) globally, presenting with a wide range of profiles and disease characteristics of this highly heterogeneous disease.[1] Although there are many differences in metabolic profiles and disease severity, most individuals are treated according to standardised guidelines. Providing each patient with standard care despite their differences may stand in the way of reaching treatment goals. It would be valuable if we could predict who is going to have trouble reaching treatment goals.

From a clinical point of view, the treatment goal is to reach adequate glycaemic control, defined as a glycated haemoglobin A1c (HbA1c) level of less than 53 mmol/mol.[2] The tools used to reach this goal are initially lifestyle advice, followed by the prescription of glucose-lowering drugs. Different drug classes are used in different disease stages as defined by national treatment guidelines, and the effect on HbA1c values differs per drug class.[3,4] Reaching adequate glycaemic control is important as high HbA1c levels have been associated with various comorbidities and mortality.[5] However, a patient's own perspective on life is important too and can be evaluated looking at quality of life (QoL) measures through different validated questionnaires. QoL is important in diabetes as people with diabetes report a lower QoL compared to those without.[6,7] Additionally, there is an intricate relationship between diabetes, mental health[8] and QoL.[9] This network of bidirectional interactions creates a risk for a patient to enter a cycle in which diabetes worsens QoL and mental health, which in turn can cause an increase in diabetes severity.[10] Therefore, preventing deterioration in QoL might not only improve the patient's experience, but also clinical parameters.

Machine learning has been used before in predicting HbA1c, but mostly based on blood glucose levels in type 1 diabetes.[11,12] Additionally, a data-driven approach has been adopted to predict glycaemic control trajectories over 6 years in Finland[13] and other authors have created a prediction model for glycaemic control in 6 months using wearable devices.[14] However, to our knowledge, no prediction model has been adopted for predicting how likely a T2D patient is to have adequate glycaemic control within a year from now, without having to use wearable or follow-up data. Prediction algorithms for QoL have been used in some other diseases,[15–17] but not yet for diabetes. Being able to predict which individual with diabetes will be likely to have inadequate glycaemic control, or to experience deterioration in QoL in the next year, would allow clinicians to prioritise patient monitoring and implement strategies to avert these negative outcomes.

Therefore, in the current study, we aimed to investigate if machine learning algorithms can be used to predict which T2D individuals are likely not to reach treatment goals in terms of glycaemic control and QoL.

2 | METHODS

2.1 | Data source

We used data from The Maastricht Study, an observational, prospective, population-based cohort study. The rationale and methodology have been described previously.[18] In brief, the study focuses on the aetiology, pathophysiology, complications and comorbidities of T2D and is characterised by an extensive phenotyping approach. Eligible for participation were all aged between 40 and 75 years and living in the southern part of the Netherlands. Participants were recruited through mass media campaigns and from the municipal registries and the regional Diabetes Patient Registry via mailings. Recruitment was stratified according to known T2DM status, with an oversampling of individuals with T2DM, for reasons of efficiency. The present study includes cross-sectional data from the first 9187 participants, who were included in the baseline survey between November 2010 and October 2020. The examinations of each participant were performed within a time window of 3 months after the baseline visit. The study has the approval of the institutional medical ethics committee (NL31329.068.10) and the Dutch Ministry of Health, Welfare and

Sport (Permit 131 088-105 234-PG). All participants gave their written informed consent.

The current study is part of the HTx Project, which is a Horizon 2020 project supported by the European Union that lasted for 5 years from January 2019. The main aim of HTx is to create a framework for the Next Generation Health Technology Assessment (HTA) to support patient-centred, societally oriented, real-time decision-making on access to and reimbursement for health technologies throughout Europe.

2.2 | Study population

From The Maastricht Study dataset, we selected all people with T2D based on the oral glucose tolerance test performed during their first (baseline) visit to the study centre or use of glucose-lowering drug based on World Health Organization (WHO) definition.[19] T2D was defined by a fasting glucose ≥7.0 mmol/L and 2 h post-load glucose ≥11.1 mmol/L, or the use of glucose-lowering drugs and the absence of a type 1 diabetes diagnosis. From the people with T2D, we selected those who used glucose-lowering drugs at baseline. For the prediction of glycaemic control, we selected individuals with no missing baseline glycated haemoglobin A1c (HbA1c) measurement (baseline visit) and a follow-up measurement available at 365 ± 120 days after baseline (hospital records). This population will be referred to as 'population GLUC'. For the prediction of QoL, we selected individuals with T2D and glucose-lowering drug use, and no missing values in the short form 36 (SF-36) data at baseline and follow-up questionnaire 1, completed 1 year after baseline. This population will be referred to as 'population QoL'. In both populations, there were no users of sodium-glucose cotransporter 2 inhibitors or glucagon-like peptide 1 receptor agonists.

2.3 | Features

A wide range of features from the Maastricht Study were used in this study. These features are listed in short below and additional information can be found in Table S1. All these features were measured at baseline.

1. General: Sex, age
2. Socio-economic: education level, income
3. Lifestyle: Dutch healthy diet (DHD) score, alcohol use, smoking
4. Fitness: sedentary wake minutes per day, sedentary bouts, percentage of moderate to vigorous activity of wake time, maximum power output at bicycle test (W/kg)
5. Diabetes-related: diabetes duration, HbA1c, body mass index (BMI)
6. Questionnaires: EQ-5D, SF-36, Big5
7. Comorbidities: depression, anxiety, albuminuria, impaired renal function, cardiovascular diseases
8. Laboratory values: high-density lipoprotein (HDL), low-density lipoprotein (LDL), triglycerides, systolic blood pressure (SBP) and diastolic blood pressure (DBP)
9. Drug use: number of different glucose-lowering drug classes used, biguanides, sulphonylureas, DPP4-Is, insulin, other glucose-lowering drugs and sleep medication/hypnotic drugs.

2.4 | Outcomes

The outcomes to be predicted in this study were glycaemic control and deterioration in QoL. Inadequate glycaemic control was defined as an HbA1c level of 53 mmol/mol or higher at 365 ± 120 days after baseline.[2] Follow-up HbA1c measurements were available from routine care through linkage with hospital data. In the case of multiple measurements, the HbA1c value closest to 365 days after baseline was selected. HbA1c measurements were, into adequate and inadequate glycaemic control. Deterioration in QoL was defined as a reduction of 3 points[6,20] in the SF-36 score in the online questionnaire at follow-up 1 compared to baseline, according to the SF-36 manual's definition of a relevant difference in score. QoL follow-up 1 took place 1 year after the baseline visit. The SF-36 produces a mental component summary (MCS) score and a physical component summary (PCS) score, which were used as separate outcomes and referred to as mental QoL and physical QoL, respectively.

2.5 | Pre-processing and machine learning

After selection of the study population and definition of the features, all features with more than 30% missing values were removed. Subsequently, we evaluated the Pearson correlation matrix and removed features with a Pearson correlation coefficient of more than 0.6.

Data were analysed using Python v3.10.9 and scikit-learn v1.2.1. We applied several supervised machine learning prediction algorithms to predict the outcomes, that is, to classify patients according to the two values of the binary outcome. We selected a diverse set of widely used and well-validated machine learning algorithms to ensure a comprehensive evaluation across different methodological families, including tree-based, instance-based, probabilistic, kernel-based and neural network approaches. These algorithms have been extensively applied in biomedical and clinical research, demonstrating solid performance across various prediction tasks. Their inclusion enables comparison under different modelling assumptions and supports the identification of robust solutions for clinical decision-making. The prediction algorithms used in this study are: Decision Tree classifier (DT), Random Forest classifier (RF), K-Nearest Neighbour classifier (KNN), Gaussian Naïve Bayes (GNB), Support Vector classifier (SVC) and Multi-Layer Perceptron classifier (MLP). All methods applied are supervised algorithms, meaning that we provide the model with the true outcome values to predict.

Additionally, we applied a logistic regression (LR) to assess how a regression model would perform compared to machine learning algorithms. LR has historically been one of the most widely used models for binary classification in medicine, making it a solid benchmark for comparing the performance of more complex models. It allows

assessing whether the use of more sophisticated models provides a real improvement over a well-established approach.[21] Detailed information on the algorithms used is provided in Data S2.

In order to find the optimal set of features for prediction, we applied three widely used different feature selection methods: Recursive Feature Elimination (RFE), meta-transformer selecting features based on importance (SelectFromModel (SFM)) and forward sequential feature selection (SFS). RFE attempts to select the optimal feature set based on the learned model and classification accuracy. It removes the worst feature that causes a drop in accuracy after building the model and does so recursively until the prespecified number of features is reached.[22] SFM is a meta-transformer selecting features with higher importance than the threshold value as indicated by the model on the training set.[23] Forward SF is a greedy procedure finding the best feature to add to the model. It iteratively finds the next best feature to add until the prespecified number of features is reached.[24] Detailed information on the feature selection methods is provided in Data S2.

We evaluated each of the prediction algorithms (default setting) with each of the feature selection methods in fivefold cross-validation, setting the number of characteristics to select to 20. In each of the folds of the cross-validations, missing values in features were imputed (using the scikit-learn function *IterativeImputer*), numeric features were scaled (using the scikit-learn function *StandardScaler*) and categorical features were coded (using the scikit-learn function OneHotEncoder).

Since each feature selection method is based on different assumptions and selection strategies, it is well known that they often produce partially overlapping but not identical sets of features. These differences are intrinsic to the different methodologies and are not necessarily a weakness, but rather a reflection of the complexity and multidimensionality of biomedical data.

Therefore, this work employed a combination of the results from multiple feature selection techniques in an ensemble strategy, the aim being to identify features that demonstrate consistent importance across methods, increasing the robustness and generalisability of the selected subset. Specifically, features retained by at least 50% of the folds in each method were included in the final feature set. This voting-based strategy reduces the reliance on a single selection method and mitigates the biases inherent in each approach, thus favouring a more stable and reliable feature space for subsequent predictive modelling.

The 50% threshold in our ensemble feature selection strategy was chosen as a balance between sensitivity and specificity in retaining relevant features.[25–28] This threshold ensures that only consistently selected features are retained in at least half of the cross-validation folds within each selection method, thus filtering out spurious or unstable variables that might arise due to sampling variability or model-specific bias. This final ensemble feature set was applied to all prediction algorithms.

Finally, we performed a second analysis with hyperparameter tuning on the three algorithms with the best results from the previous analysis to find the best combination and evaluate its impact on the models' performance. This analysis was done in conjunction with RFE feature selection.

2.6 | Evaluation

The models were evaluated by fivefold cross-validation, using a pipeline including variable transformation, imputation, feature selection and hyperparameter selection when applicable.

The final models were scored using accuracy, precision, recall, F1-score and the area under the receiver operating characteristic curve (AUC-ROC). All scoring parameters range from 0 to 1, with 1 being a perfect score. Both the mean and the 95% confidence interval (CI) were calculated. The scoring parameters are based on the number of true positive (TP), false positive (FP), true negative (TN) and false negative (FN) individuals, in which positive refers to the label "1" and negative refers to the label "0". True refers to correctly classified individuals and false to incorrectly classified individuals.

Accuracy represents the number of correctly classified people over the total number of people: $\text{Accuracy} = \frac{TN+TP}{TN+FP+TP+FN}$.

Precision represents the number of correctly classified positives over the total number of predicted positives: $\text{Precision} = \frac{TP}{TP+FP}$. Recall represents the number of correctly classified positives over the number of actual positives:

$\text{Recall} = \frac{TP}{TP+FN}$. The F1-score is a harmonic mean of precision and recall, only high when both precision and recall are high: $\text{F1} - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precsion} + \text{Recall}}$.

The receiver operating characteristic (ROC) curve is composed by plotting the model's true positive rate (TPR) versus its false positive rate (FPR) across all possible classification thresholds. TPR is the probability that a positive value is correctly predicted as positive, whereas the FPR is the probability that a negative value is correctly predicted as negative. The AUC-ROC is a summary statistic representing the probability that the model will rank a randomly chosen positive value more highly than a randomly chosen negative value, in which 'higher' means further towards positivity.

3 | RESULTS

3.1 | Study population

Figure 1 shows the selection of people with diabetes and glucose-lowering drug use into two groups of people: one with no missing values at baseline and follow-up HbA1c ($N = 842$, population GLUC) and one with no missing values at baseline and follow-up SF-36 ($N = 964$, population QoL). The baseline characteristics of both groups are shown in Table 1. The mean age in both populations was around 63 years old and almost half of the people had a low education level. The mean diabetes duration was over 9 years in both populations, although with a substantial standard deviation (SD). In population GLUC for the prediction of glycaemic control (Table 1), baseline HbA1c was slightly elevated (mean: 54 mmol/mol) and 43.2% of individuals had inadequate glycaemic control at baseline. After 1 year, the mean HbA1c had increased slightly to 54.4 mmol/mol and 46.4% of the individuals had inadequate glycaemic control. In

population QoL for the prediction of deterioration in QoL (Table 1), the mean mental- and physical QoL scores at baseline were 53.1 and 46.9, respectively, and 10.9% reported their health to be worse compared to 12 months ago. One year after baseline, there was a mean reduction in both mental QoL (−1.0) and physical QoL (−1.8) scores, and 29.3% and 34.3% of the individuals had reported a deterioration in these scores, respectively.

3.2 | Feature selection

Features with more than 30% missing values (homeostatic model assessment of insulin resistance and beta-cell function, glomerular filtration rate and percentage liver fat) were excluded. We excluded weight, height and waist circumference due to correlation with BMI and total cholesterol due to correlation with LDL and HDL. The other features did not show high correlation. Table 2 provides an overview of the features selected for each prediction model and each feature selection method in the first analysis. The following features were selected by at least 50% of the folds for the prediction of inadequate glycaemic control (population GLUC) and were therefore included in the final ensemble feature set: sex, age, income, diet score, baseline HbA1c, depression, anxiety, albumin excretion, renal function, use of sleep medication and the use of biguanides, sulphonylurea, insulin or other glucose-lowering drugs. Baseline HbA1c was selected in 100% of the folds and the use of sleep medication in 74.4% of the folds. The other features included were selected in 50.0%–66.7% of the folds, with most features relating to comorbidities or drug use.

The following features were selected for the prediction of deterioration in physical QoL (population QoL): diet score, EQ5D score, baseline mental QoL, baseline physical QoL, Big5 conscientiousness, depression, unhealthy LDL, triglycerides, DBP and use of sleep medication. The following features were selected for the prediction of deterioration in mental QoL (population QoL): age, income, baseline physical QoL, baseline mental QoL, Big5 emotional stability, depression, renal function, triglycerides, DBP and insulin use. In these two models, baseline mental QoL and physical QoL were chosen in most folds. The other features included were selected in 50.0%–65.6% and 54.4%–71.1% of the folds for the mental QoL and physical QoL model, respectively. Most features included in the models related to questionnaires or laboratory values.

The feature selection in the second analysis with hyperparameter tuning is detailed below and the overview of the features selected is included in Table S3.

In the prediction of glycaemic control (population GLUC), the following features were selected both for GNB (tuned hyperparameters: *var_smoothing = 0.53*), MLP (tuned hyperparameters: *activation = tanh*, *alpha = 0.05*, *hidden_layer_sizes = (50,100,50)*, *learning_rate = adaptive*, *max_iter = 30*, *solver = adam*) and SVC (tuned hyperparameters: *C = 10*, *gamma = 0.01*, *kernel = sigmoid*): moderate to vigorous physical activity (MVPA) of wake time, sedentary bouts, years since T2D diagnosis, diet score, triglycerides, BMI, baseline MCS, baseline PCS and baseline HbA1c. Additionally, age (GNB, SVC) and sedentary wake minutes (MLP, SVC) were selected by two models. Finally, SBP, DBP, Big 5 extraversion and insulin use were selected only in the SVC algorithm.

In the prediction of deterioration in physical QoL (population QoL), the following features were selected for both GNB (tuned hyperparameters: *var_smoothing = 1*), MLP (tuned hyperparameters: *activation = relu*, *alpha = 0.05*, *hidden_layer_sizes = (50,50,50)*, *learning_rate = constant*, *max_iter = 30*, *solver = adam*) and SVC (tuned hyperparameters: *C = 1*, *gamma = 0.0001*, *kernel = sigmoid*): MVPA of wake time, sedentary bouts, sedentary wake minutes, diet score, triglycerides, BMI, DBP, baseline MCS, baseline PCS, Big 5 conscientiousness. Additionally, age, years since T2D diagnosis, SBP, Big 5 extraversion and Big 5 openness were selected by both MLP and

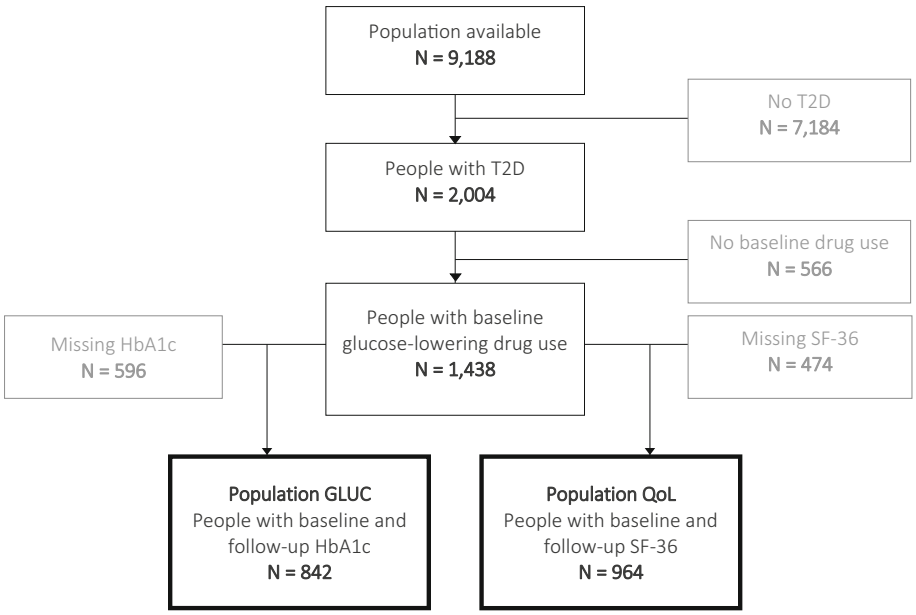

FIGURE 1 Flowchart showing the selection of people with T2D and baseline drug use into two groups of people with no missing values in outcome parameters in the baseline and follow-up outcomes (Population GLUC and QoL). Overlap was allowed in the final two populations (bold boxes) in order to maximise the number of people. HbA1c, glycated haemoglobin 1Ac; SF-36, short form 36; T2D, type 2 diabetes.

TABLE 1 Baseline characteristics and output values of the population used in the prediction of glycaemic control (Population GLUC) and in the prediction of deterioration in QoL (Population QoL). Data are given in *n* (%) unless specified otherwise.

	Population GLUC		Population QoL	
	N (%)	Missing (%)	*N* (%)	Missing (%)
Number of people	842	n/a	964	n/a
General				
Number of women	252 (29.9)	0 (0.0)	275 (28.5)	0 (0.0)
Age, mean (SD)	63.3 (7.5)	0 (0.0)	63.1 (7.5)	0 (0.0)
Socio-economic				
Low education level	409 (49.9)	23 (2.7)	448 (46.9)	8 (0.8)
Income lower than median	575 (68.3)	0 (0.0)	625 (64.8)	0 (0.0)
Lifestyle				
Diet score [0–100], mean (SD)	79.9 (14.6)	72 (8.6)	80.1 (14.7)	44 (4.6)
High alcohol consumption	128 (15.5)	16 (1.9)	157 (16.3)	<5 (0.1)
Current smoker	156 (18.9)	16 (1.9)	163 (16.9)	<5 (0.2)
Fitness				
Sedentary wake minutes per day, mean (SD)	607.1 (109.5)	138 (16.4)	602.7 (106.0)	154 (15.9)
Number of sedentary bouts, mean (SD)	318.4 (112.0)	138 (16.4)	319.8 (111.4)	154 (15.9)
Percentage MVPA of wake time, mean (SD)	4.2 (2.5)	138 (16.4)	4.4 (2.5)	154 (15.9)
Wmax in lowest tertile	380 (61.5)	224 (26.6)	459 (61.2)	214 (22.2)
Diabetes-related				
Years since T2D diagnosis, mean (SD)	9.7 (7.6)	210 (24.9)	9.3 (7.6)	237 (24.6)
HbA1c at baseline in mmol/mol, mean (SD)	54.0 (11.8)	<5 (0.1)	53.1 (11.3)	<5 (0.1)
Inadequate glycaemic control at baseline	364 (43.2)	0 (0.0)	n/a	
BMI in kg/m^2, mean (SD)	30.0 (4.9)	<5 (0.2)	29.9 (5.0)	<5 (0.1)
Questionnaire				
EQ-5D 3L score [−0.330–1.000], mean (SD)	0.8 (0.2)	33 (3.9)	0.9 (0.2)	<5 (0.3)
EQ-5D health score [0–100], mean (SD)	69.8 (21.0)	32 (3.8)	71.1 (20.4)	<5 (0.3)
EQ-5D health worse than 12 months ago	94 (11.6)	34 (4.0)	105 (10.9)	<5 (0.3)
SF-36 MCS, mean (SD)	53.0 (8.7)	35 (4.2)	53.1 (8.5)	0 (0.0)
SF-36 PCS, mean (SD)	45.8 (10.3)	35 (4.2)	46.9 (9.6)	0 (0.0)
Big5 extraversion, mean (SD)	4.8 (1.2)	154 (18.3)	4.9 (1.2)	146 (15.2)
Big5 conscientiousness, mean (SD)	5.2 (1.0)	154 (18.3)	5.2 (1.0)	146 (15.2)
Big5 agreeableness, mean (SD)	5.6 (0.8)	153 (18.2)	5.6 (0.8)	145 (15.0)
Big5 emotional stability, mean (SD)	4.9 (1.1)	153 (18.2)	5.0 (1.1)	145 (15.0)
Big5 openness, mean (SD)	4.6 (1.1)	154 (18.3)	4.6 (1.1)	146 (15.2)
Comorbidities				
Depression	43 (5.5)	57 (6.8)	48 (5.3)	53 (5.5)
Anxiety	45 (6.2)	116 (13.8)	43 (5.1)	118 (12.2)
Abnormal albumin excretion	186 (22.5)	17 (2.0)	205 (21.4)	8 (0.8)
Impaired renal function	364 (43.2)	0 (0.0)	48 (5.0)	0 (0.0)
Cardiovascular disease	277 (32.9)	0 (0.0)	286 (29.7)	0 (0.0)
Laboratory values				
HDL <1 mmol/L	96 (11.4)	0 (0.0)	95 (9.9)	0 (0.0)
LDL >3 mmol/L	112 (13.3)	0 (0.0)	132 (13.7)	0 (0.0)
Triglycerides in mmol/L, mean (SD)	1.7 (1.1)	0 (0.0)	1.7 (1.1)	0 (0.0)
SBP in mmHg, mean (SD)	142.0 (17.7)	0 (0.0)	141.4 (17.5)	<5 (0.1)
DBP in mmHg, mean (SD)	76.3 (9.6)	0 (0.0)	76.4 (9.4)	<5 (0.1)

TABLE 1 (Continued)

	Population GLUC		Population QoL	
	N (%)	Missing (%)	*N* (%)	Missing (%)
History of drug use				
Sleep medication or hypnotics	23 (2.7)	0 (0.0)	24 (2.5)	0 (0.0)
Number of glucose-lowering drug classes	1.8 (0.8)	0 (0.0)	1.7 (0.8)	0 (0.0)
Biguanide	763 (90.6)	0 (0.0)	881 (91.4)	0 (0.0)
Sulphonylurea	315 (37.4)	0 (0.0)	348 (36.1)	0 (0.0)
DPP4-I	90 (10.7)	0 (0.0)	123 (12.8)	0 (0.0)
Other	61 (7.2)	0 (0.0)	255 (26.5)	0 (0.0)
Insulin	267 (31.7)	0 (0.0)	65 (6.7)	0 (0.0)
Output (after 1 year)				
HbA1c in mmol/mol, mean (SD)	54.4 (12.4)	0 (0.0)	n/a	
Inadequate glycaemic control	391 (46.4)	0 (0.0)	n/a	
Difference in MCS, mean (SD)	n/a		−1.0 (8.4)	0 (0.0)
Deterioration in MCS	n/a		282 (29.3)	0 (0.0)
Difference in PCS, mean (SD)	n/a		−1.8 (7.8)	0 (0.0)
Deterioration in PCS	n/a		331 (34.3)	0 (0.0)

Note: Elaborate feature definitions can be found in Table S1.
Abbreviations: BMI, body mass index; DBP, diastolic blood pressure; DPP4-I, dipeptidyl peptidase 4 inhibitor; HbA1c, glycated haemoglobin A1c; HDL, high-density lipoprotein; LDL, low-density lipoprotein; MCS, mental component summary ('mental QoL'); MVPA, moderate to vigorous physical activity; *N*, number; PCS, physical component summary ('physical QoL'); QoL, quality of life; SBP, systolic blood pressure; SD, standard deviation; SF, short form; T2D, type 2 diabetes.

SVC. Finally, Big 5 agreeableness and Big 5 emotional stability, EQ5D health Score, EQ5D 3 L score and baseline HbA1c were selected only in the SVC algorithm.

In the prediction of deterioration in mental QoL (population QoL), the following features were selected for both GNB (tuned hyperparameters: *var_smoothing* = *0.81*), MLP (tuned hyperparameters: *activation* = *tanh*, *alpha* = *0.0001*, *hidden_layer_sizes* = (*120,80,40*), *learning_rate* = *constant*, *max_iter* = *30*, *solver* = *adam*) and SVC (tuned hyperparameters: *C* = *0.1*, *gamma* = *0.001*, *kernel* = *sigmoid*): MVPA of wake time, sedentary wake minutes, diet score, triglycerides, BMI, SBP, baseline MCS, baseline PCS, baseline HbA1c and Big 5 conscientiousness. Additionally, sedentary bouts, age, Big 5 extraversion, Big 5 emotional stability and Big 5 openness were selected by both MLP and SVC. Finally, years since T2D diagnosis, DBP, Big 5 agreeableness, EQ5D health score and EQ5D 3 L score were selected only in the SVC algorithm.

3.3 | Model performance

Table 3 provides an overview of the model performance per outcome and per prediction algorithm used with the ensemble feature set. In the model to predict glycaemic control, SVC performed best in terms of accuracy (0.76), precision (0.79) and AUC-ROC. However, MLP performed best in terms of recall (0.72) and F1-score (0.73) and had slightly lower accuracy (0.75) and precision (0.74) scores, as well as a slightly lower AUC-ROC (0.83). Generally, SVC and MLP were the better performing models compared to DT, RF, KNN and GNB. LR performed similarly to SVC, but with slightly lower precision and recall.

Model performance was lower in the prediction of deterioration in QoL, with accuracy ranging from 0.61 to 0.71 for mental QoL, and 0.58–0.67 for physical QoL. Precision scores ranged from 0.33 to 0.57 for mental QoL and 0.38–0.57 for physical QoL. Recall and F1-score were generally low, with most models not scoring over 0.30 for both parameters. These ranges do not include SVC for mental QoL, since this algorithm classified all cases as negative (i.e., no deterioration), leading to a precision, recall and F1-score of 0. AUC-ROC ranged between 0.53 and 0.64 for mental QoL and 0.54–0.63 for physical QoL. LR generally yielded the best accuracy, precision and AUC-ROC, whereas DT yielded higher recall and F1-scores.

Table 4 provides the results of the performance of the models with hyperparameter tuning for the three best performing models in the previous analysis. The classifiers included were GNB, SVC and MLP. The AUC-ROCs obtained for glycaemic control were 0.77, 0.83 and 0.83, respectively. For deterioration in physical QoL, the scores were 0.59, 0.58 and 0.57, respectively. For deterioration in mental QoL, the scores were 0.59, 0.61 and 0.58, respectively. The SVC for mental and physical QoL classified all cases as negative (i.e., no deterioration), leading to a precision, recall and F1-score of 0.

TABLE 2 Features selected in the different methods for the prediction of glycaemic control, deterioration in mental QoL (MCS) and deterioration in physical QoL (PCS). Green boxes show the total percentages over 50.0, that is, the features chosen for the final model.

	Glycaemic control				Deterioration in MCS				Deterioration in PCS			
	RFE	SFM	SFS	Total	RFE	SFM	SFS	Total	RFE	SFM	SFS	Total
General												
Number of women	70.0	33.3	76.7	60.0	50.0	0.0	33.3	27.8	40.0	0.0	23.3	21.1
Age	90.0	66.7	33.3	63.3	63.3	46.7	70.0	60.0	46.7	40.0	60.0	48.9
Socio-economic												
Low education level	26.7	13.3	33.3	24.4	50.0	33.3	13.3	32.2	66.7	33.3	20.0	40.0
Income lower than median	46.7	26.7	80.0	51.1	50.0	63.3	36.7	50.0	6.7	0.0	23.3	10.0
Lifestyle												
Diet score	66.7	46.7	63.3	58.9	36.7	26.7	66.7	43.3	73.3	60.0	63.3	65.6
High alcohol consumption	26.7	16.7	53.3	32.2	46.7	33.3	26.7	35.6	40.0	16.7	26.7	27.8
Current smoker	23.3	13.3	60.0	32.2	43.3	16.7	16.7	25.6	0.0	0.0	33.3	11.1
Fitness												
Sedentary wake minutes per day	50.0	20.0	26.7	32.2	40.0	33.3	73.3	48.9	33.3	26.7	66.7	42.2
Number of sedentary bouts	50.0	26.7	50.0	42.2	33.3	33.3	70.0	45.6	43.3	30.0	66.7	46.7
Percentage MVPA of wake time	60.0	40.0	46.7	48.9	40.0	26.7	63.3	43.3	36.7	33.3	70.0	46.7
Wmax in lowest tertile	3.3	0.0	63.3	22.2	40.0	30.0	26.7	32.2	63.3	46.7	20.0	43.3
Diabetes-related												
Years since T2D diagnosis	33.3	16.7	43.3	31.1	46.7	36.7	60.0	47.8	56.7	26.7	63.3	48.9
Baseline HbA1c in mmol/mol	100.0	100.0	100.0	100.0	53.3	36.7	56.7	48.9	40.0	30.0	50.0	40.0
BMI in kg/m^2	43.3	20.0	30.0	31.1	43.3	30.0	66.7	46.7	50.0	33.3	60.0	47.8
Questionnaire												
EQ5D 3L score	26.7	26.7	36.7	30.0	40.0	26.7	60.0	42.2	80.0	46.7	73.3	66.7
EQ5D health score (0–100)	30.0	10.0	50.0	30.0	33.3	20.0	63.3	38.9	33.3	23.3	60.0	38.9
EQ5D health worse than 12 months ago	60.0	20.0	60.0	46.7	50.0	20.0	40.0	36.7	10.0	13.3	40.0	21.1
SF-36 PCS	40.0	26.7	23.3	30.0	96.7	100.0	70.0	88.9	100.0	86.7	73.3	86.7
SF-36 MCS	30.0	30.0	26.7	28.9	96.7	100.0	73.3	90.0	100.0	90.0	63.3	84.4
Big5 extraversion	46.7	0.0	40.0	28.9	40.0	23.3	73.3	45.6	36.7	23.3	70.0	43.3
Big5 conscientiousness	73.3	30.0	16.7	40.0	50.0	33.3	60.0	47.8	86.7	40.0	70.0	65.6
Big5 agreeableness	33.3	10.0	56.7	33.3	56.7	23.3	63.3	47.8	36.7	16.7	63.3	38.9
Big5 emotional stability	46.7	13.3	53.3	37.8	76.7	53.3	66.7	65.6	40.0	26.7	63.3	43.3
Big5 openness	46.7	20.0	33.3	33.3	33.3	30.0	66.7	43.3	46.7	23.3	66.7	45.6
Comorbidities												
Depression	53.3	53.3	63.3	56.7	53.3	60.0	50.0	54.4	66.7	66.7	50.0	61.1
Anxiety	60.0	50.0	86.7	65.6	30.0	6.7	26.7	21.1	43.3	36.7	33.3	37.8
Abnormal albumin excretion	60.0	46.7	50.0	52.2	26.7	3.3	26.7	18.9	43.3	40.0	23.3	35.6
Impaired renal function	70.0	60.0	46.7	58.9	70.0	63.3	20.0	51.1	23.3	10.0	36.7	23.3
Cardiovascular disease	13.3	0.0	43.3	18.9	26.7	3.3	30.0	20.0	73.3	56.7	16.7	48.9
Laboratory values												
HDL <1 mmol/L	23.3	3.3	70.0	32.2	36.7	40.0	36.7	37.8	26.7	3.3	40.0	23.3
LDL >3 mmol/L	26.7	13.3	46.7	28.9	26.7	10.0	30.0	22.2	76.7	56.7	30.0	54.4
Triglycerides in mmol/L	83.3	20.0	16.7	40.0	66.7	36.7	66.7	56.7	63.3	46.7	63.3	57.8
SBP in mmHg	40.0	13.3	26.7	26.7	50.0	30.0	63.3	47.8	43.3	30.0	70.0	47.8
DBP in mmHg	73.3	33.3	40.0	48.9	63.3	30.0	73.3	55.6	80.0	63.3	70.0	71.1
Drug use												
Sleep medication or hypnotics	66.7	66.7	90.0	74.4	30.0	3.3	26.7	20.0	66.7	66.7	46.7	60.0

TABLE 2 (Continued)

	Glycaemic control				Deterioration in MCS				Deterioration in PCS			
	RFE	SFM	SFS	Total	RFE	SFM	SFS	Total	RFE	SFM	SFS	Total
Number of glucose-lowering drug classes	3.3	6.7	23.3	11.1	46.7	40.0	30.0	38.9	0.0	0.0	30.0	10.0
Biguanide	66.7	80.0	53.3	66.7	33.3	13.3	23.3	23.3	53.3	40.0	40.0	44.4
Sulphonylurea	73.3	53.3	46.7	57.8	36.7	33.3	33.3	34.4	66.7	53.3	20.0	46.7
DPP4-I	26.7	23.3	53.3	34.4	56.7	26.7	26.7	36.7	40.0	13.3	36.7	30.0
Insulin	76.7	43.3	30.0	50.0	63.3	66.7	43.3	57.8	16.7	3.3	26.7	15.6
Other	60.0	53.3	56.7	56.7	23.3	0.0	33.3	18.9	46.7	30.0	40.0	38.9

Note: Elaborate feature definitions can be found in Table S1.

Abbreviations: BMI, body mass index; DBP, diastolic blood pressure; DPP4-I, dipeptidyl peptidase 4 inhibitor; HbA1c, glycated haemoglobin A1c; HDL, high-density lipoprotein; LDL, low-density lipoprotein; MCS, mental component summary; MVPA, moderate to vigorous physical activity; PCS, physical component summary; RFE, recursive feature elimination; SBP, systolic blood pressure; SF, short form; SFM, select from model; SFS, sequential feature selection; T2D, type 2 diabetes.

TABLE 3 Scoring of the final models with the ensemble feature set. Data are given in score (95% CI).

	Accuracy	Precision	Recall	F1-score	AUC-ROC
Glycaemic control					
DT	0.68 (0.66–0.71)	0.68 (0.65–0.70)	0.60 (0.49–0.70)	0.63 (0.58–0.69)	0.69 (0.67–0.73)
RF	0.67 (0.65–0.69)	0.74 (0.71–0.79)	0.55 (0.35–0.54)	0.55 (0.48–0.62)	0.69 (0.63–0.73)
KNN	0.67 (0.63–0.72)	0.68 (0.61–0.72)	0.58 (0.49–0.68)	0.62 (0.58–0.69)	0.62 (0.58–0.69)
GNB	0.69 (0.62–0.76)	0.72 (0.61–0.81)	0.55 (0.49–0.63)	0.62 (0.55–0.70)	0.76 (0.69–0.84)
SVC	0.76 (0.71–0.79)	0.79 (0.71–0.83)	0.67 (0.60–0.72)	0.72 (0.68–0.76)	0.85 (0.80–0.89)
MLP	0.75 (0.72–0.79)	0.74 (0.69–0.77)	0.72 (0.64–0.79)	0.73 (0.64–0.79)	0.83 (0.78–0.88)
LR	0.76 (0.71–0.79)	0.78 (0.71–0.84)	0.66 (0.61–0.70)	0.72 (0.68–0.75)	0.85 (0.80–0.89)
Deterioration of mental QOL					
DT	0.61 (0.60–0.63)	0.33 (0.30–0.36)	0.33 (0.27–0.37)	0.33 (0.38–0.36)	0.53 (0.50–0.56)
RF	0.67 (0.66–0.70)	0.36 (0.32–0.46)	0.15 (0.12–0.18)	0.21 (0.18–0.24)	0.57 (0.56–0.59)
KNN	0.68 (0.67–0.70)	0.43 (0.39–0.48)	0.23 (0.19–0.27)	0.29 (0.26–0.33)	0.56 (0.52–0.60)
GNB	0.67 (0.64–0.71)	0.39 (0.34–0.49)	0.18 (0.09–0.26)	0.24 (0.15–0.30)	0.61 (0.59–0.65)
SVC	0.71 (0.70–0.71)	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.61 (0.59–0.63)
MLP	0.68 (0.66–0.71)	0.43 (0.36–0.53)	0.20 (0.14–0.024)	0.27 (0.23–0.31)	0.59 (0.55–0.63)
LR	0.71 (0.69–0.73)	0.57 (0.40–0.74)	0.15 (0.11–0.18)	0.23 (0.19–0.27)	0.64 (0.59–0.69)
Deterioration of physical QOL					
DT	0.58 (0.51–0.54)	0.38 (0.30–0.47)	0.38 (0.30–0.47)	0.38 (0.31–0.49)	0.54 (0.47–0.62)
RF	0.64 (0.60–0.67)	0.46 (0.33–0.53)	0.20 (0.16–0.24)	0.28 (0.21–0.32)	0.55 (0.50–0.60)
KNN	0.64 (0.62–0.66)	0.45 (0.40–0.50)	0.26 (0.21–0.31)	0.33 (0.28–0.38)	0.58 (0.53–0.62)
GNB	0.64 (0.60–0.67)	0.44 (0.34–0.52)	0.19 (0.14–0.27)	0.26 (0.21–0.35)	0.62 (0.56–0.65)
SVC	0.66 (0.64–0.67)	0.24 (0.00–0.92)	0.01 (0.00–0.03)	0.02 (0.00–0.06)	0.63 (0.56–0.70)
MLP	0.65 (0.64–0.66)	0.49 (0.46–0.51)	0.29 (0.23–0.36)	0.36 (0.31–0.42)	0.62 (0.58–0.63)
LR	0.67 (0.63–0.70)	0.57 (0.41–0.81)	0.16 (0.12–0.19)	0.24 (0.19–0.29)	0.63 (0.59–0.69)

Abbreviations: DT, decision tree classifier; GNB, Gaussian Naïve Bayes; KNN, K-nearest neighbour classifier; LR, logistic regression; MLP, multi-layer perceptron classifier; RF, random forest classifier; SVC, support vector classifier.

4 | DISCUSSION

In this study, we used various machine learning algorithms and LR to predict being under inadequate glycaemic control and experiencing deterioration in mental or physical QoL scores, all 1 year after baseline. We observed good performance of some models for predicting glycaemic control, but prediction of the deterioration in mental or physical QoL score does not perform well under the conditions described. LR performed similarly to the best performing machine learning algorithm in the prediction of both glycaemic control and deterioration in QoL.

TABLE 4 Scoring of the final models with hyperparameter tuning. Data are given in score (95%CI).

	Accuracy	Precision	Recall	F1-score	AUC-ROC
Glycaemic control					
GNB	0.70 (0.68–0.73)	0.82 (0.75–0.88)	0.47 (0.41–0.54)	0.59 (0.54–0.64)	0.77 (0.74–0.81)
SVC	0.75 (0.73–0.77)	0.79 (0.75–0.82)	0.64 (0.59–0.69)	0.70 (0.67–0.73)	0.83 (0.80–0.87)
MLP	0.76 (0.74–0.78)	0.78 (0.75–0.80)	0.68 (0.64–0.72)	0.72 (0.70–0.75)	0.83 (0.80–0.85)
Deterioration of mental QOL					
GNB	0.70 (0.69–0.71)	0.47 (0.22–0.72)	0.04 (0.02–0.05)	0.06 (0.04–0.09)	0.59 (0.55–0.63)
SVC	0.71 (0.71–0.71)	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.61 (0.58–0.65)
MLP	0.70 (0.69–0.71)	0.45 (0.39–0.51)	0.17 (0.11–0.22)	0.24 (0.18–0.30)	0.58 (0.55–0.62)
Deterioration of physical QOL					
GNB	0.65 (0.64–0.66)	0.42 (0.13–0.70)	0.02 (0.01–0.03)	0.05 (0.03–0.07)	0.59 (0.53–0.64)
SVC	0.66 (0.65–0.66)	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.58 (0.55–0.61)
MLP	0.63 (0.61–0.65)	0.43 (0.37–0.48)	0.26 (0.21–0.30)	0.32 (0.27–0.37)	0.57 (0.55–0.60)

Abbreviations: GNB, Gaussian Naïve Bayes; MLP, multi-layer perceptron classifier; SVC, support vector classifier.

4.1 | Features selected

The ensemble feature selection allowed us to reduce the model from 42 features to 14, 10 and 10 features for the prediction of glycaemic control, deterioration in mental QoL and deterioration in physical QoL, respectively. As expected, the baseline values of the outcomes were selected as predictors in all models. HbA1c was even selected in 100% of the feature selection folds. Mental QoL was selected in 90.0% of the folds and physical QoL in 86.7% of the folds in their respective feature selection models.

In the model for the prediction of glycaemic control, most selected features were comorbidities or categories of drug use. This is in accordance with previous reports where depression,[10] anxiety,[29] albuminuria[30,31] and renal disease[31] have been associated with glycaemic control. Sex and age have been known to have a relationship with glycaemic control as well,[32,33] and so does having a low income.[34] Although adherence to the Dutch Healthy Diet index has only been associated with a decrease in BMI rather than improved glycaemic control,[35] diet has been associated with glycaemic control before.[36] All glucose-lowering drug classes have been selected for the model except for DPP4-I use. Given that drug classes influence HbA1c levels to varying degrees,[4] it is a logical result that each of them is an important predictor. The class of DPP4-Is was not selected, possibly due to the low number of people using this drug at baseline, leading to limited information provided by this drug class compared to the other drug classes.

In the prediction models for the deterioration of physical and mental QoL, greater instability was observed in terms of the selected variables with greater variability within the folds of the different cross-validations. It does make sense that baseline mental and physical QoL were selected for their models, and that the EQ5D 3 L score was also chosen in two-thirds of the folds. Interestingly, the mean EQ-5D 3 L score was quite high at 0.9 considering the maximum score of 1.000, whereas the total health score of 71.1 was rather low compared to a mean value of 82.0 reported in Dutch community-dwelling elderly.[37] Baseline mental and physical QoL were similar to those reported previously in this population,[7] but lower compared to the scores reported in people without diabetes.[6] Although 10.9% indicated that they had experienced a decline in health over the past year at baseline, 29.3% (mental QoL) and 34.3% (physical QoL) reported a deterioration in QoL a year after that. These contradictive patterns show that there must be an intricate network of pathways underlying the change in QoL, not to mention the subjectiveness of these measures. This makes it difficult to incorporate these networks into a well-performing prediction model.

4.2 | Interpreting model performance

Generally, an AUC form 0.70 to 0.80 is considered acceptable, 0.80 to 0.90 is excellent, and over 0.90 is outstanding.[38] This definition implies that for the first analysis with the ensemble feature set, the SVC, MLP and LR were excellent in predicting glycaemic control and that GNB was acceptable in doing so. In practice, it would be important to identify as many people who are likely not to reach glycaemic control as possible in order to monitor those people closely and limit disease progression. In other words, we are interested in a model that maximises the number of TPs and minimises the number of FPs and FNs, so performance in terms of F1-score is also important. Again, the three algorithms that obtain good values in terms of F1-score above 0.70 are SVC, MLP and LR, with MLP slightly above. In the second analysis with hyperparameter tuning, very similar results are obtained, with the SVC and MLP algorithms obtaining excellent results above 0.80 in terms of AUC-ROC and good results above 0.70 in terms of F1-score.

Regarding the prediction of QoL, similar results were obtained in both analyses. The mental and physical QoL models did not reach an acceptable AUC-ROC, always below 0.65, as well as an F1-score below 0.40.

4.3 | Comparison with other models

The number of prediction models similar to the ones we attempted to create in the current study is limited. Patel et al.[14] used various techniques to predict glycaemic control after 6 months in 147 people with prediabetes. The ensemble machine learning method performed best, and predictions improved using (wrist-worn) wearable data. The AUC-ROC was 0.85 (95%CI 0.79–0.90), which is similar to our SVC and LR models. No other scoring parameters were reported.

A Finnish study[13] reported a prediction model for HbA1c trajectories over 6 years, based on 9631 individuals with T2D. The trajectories were grouped into adequate and inadequate glycaemic control, and the final neural network model gave correct predictions for 86.6% of the individuals with inadequate glycaemic control, which is higher than what we obtained in our 1-year prediction model. This is possibly due to our use of less advanced models, different features and a smaller dataset.

Fan et al.[39] studied a large range of prediction models in 165 non-adherent people with T2D. The Bayesian network reached the highest AUC-ROC of 0.82. This model score is slightly lower than the score of our best performing model.

Fu et al.[40] created a prediction model for glycaemic control after 52 weeks with BMI, pulse and several biochemical blood measurements. The XGBoost algorithm performed best with an AUC-ROC of 0.68, which is lower compared to our models. The differences in the definition of glycaemic control (cut-off of 48 mmol/mol), as well as the use of various biochemical blood measurements and different prediction algorithms, could explain these results.

Overall, these model scores for the prediction of HbA1c do not perform excellently (AUC-ROC over 0.9), showing that accurate prediction remains a challenge. This could be due to the intricate pathways involved in glycaemic control,[41] as well as limitations in data availability and follow-up measurements.

The prediction of QoL remains a challenge. To our knowledge, there is no literature available on prediction models for QoL in diabetes. In different diseases and settings, prediction models for QoL have been created using the SF-36 like we did,[15] EQ-5D[42] or disease-specific QoL scales.[16,17] The work of Khan et al.[15] obtained an AUC-ROC of 0.77 for mental QoL improvement and 0.78 for physical QoL improvement 1 year after surgery for mild degenerative cervical myelopathy. In contrast to our work, this model predicts an improvement in QoL (defined as an increase of at least 4 points on the SF-36). Furthermore, the study population was small and no cross-validation was performed.

The work of de Jonge et al.[42] obtained an R^2 of 0.52 using EQ-5D for QoL in patients 1 year after intensive care admission. Although the sample in this work is larger and more robustly evaluated by cross-validation, the prediction target is continuous, so a direct comparison between the results obtained is not possible. The works of Candel-Parra et al.[16] and Karri et al.[17] obtained an AUC-ROC of 0.80–0.90 in disease-specific scales for Parkinson's and cancer disease after 1 year respectively. These works used disease-specific scales with threshold cut-off values different from those used in our work, as well as small patient samples complicating the robustness and generalisability of the results. Moreover, several of these papers make no mention of the treatment of missing data[15,16] or perform simple imputation with the mean,[42] contrary to recommendations to use more advanced imputation techniques for bias reduction. All these results emphasise the need for further work in the field of QoL prediction, as well as in the design and use of disease-specific scales and in the use of larger samples of patients.

4.4 | Strengths and limitations

The strengths of this study include the wide range of features explored as well as the feature selection techniques employed to select the most important features. The sample of patients used, although not large for a machine learning study, is one of the largest used in the literature of QoL prediction. Data on drug use were complete and accurate due to the use of longitudinal pharmacy records. Moreover, we used advanced imputation techniques to handle missing data instead of analysing only complete cases or using simpler imputation techniques. In addition, we used different scoring parameters to obtain information on how well the models performed in terms of classifying positives and negatives correctly, obtaining a more complete and objective view of the real performance of the models.

However, there are some limitations to keep in mind. The use of a solely Dutch population could limit the external validity of the model, despite extensive cross-validation. Additionally, we excluded individuals with no glucose-lowering drug use (both populations) and those with no hospital records of HbA1c measurements (Population GLUC) which further limits the external validity. The large difference in insulin use between Population GLUC and Population QoL shows that there are mostly people with advanced diabetes in Population GLUC. Additionally, the use of newer glucose-lowering drug classes is underrepresented in the current population. Although our population is larger than in most of the studies described in the previous paragraph, this larger population sample has not resulted in better model performance. This might also be due to the limited follow-up time, as insufficient QoL follow-up was available beyond 1 year. The Maastricht Study does not measure the diabetes-specific QoL scale,[43] so we were unable to use a disease-specific scale for the prediction of deterioration in QoL. We had no information on some features previously associated with QoL, such as acceptance,[44] knowledge[44] and executive functioning.[45] The use of more sophisticated prediction algorithms remains open to further exploration for possible improvements of the results.

4.5 | Clinical implications

The current study not only presents a prediction model with potential clinical applications; it also provides insight into the features of importance in the prediction of inadequate glycaemic control and deterioration in QoL.

The implementation of a prediction model in practice would allow a clinician to evaluate newly diagnosed type 2 diabetes patients using a computerised assessment tool. By inputting the patient's profile, the system would generate a risk stratification for inadequate glycaemic control and/or QoL deterioration. If the model anticipates these adverse outcomes, the clinician could initiate a more rigorous follow-up protocol, incorporating enhanced monitoring and treatment intensification where necessary in order to prevent these outcomes from manifesting.

While awaiting further development into clinical implementation of this predictive model, the current study provides insight into the features of importance for these predictions. Our findings suggest that the presence of depression, anxiety, albuminuria and renal disease, along with low income and baseline HbA1c levels, are important characteristics for preliminary risk assessment regarding future inadequate glycaemic control. Risk factors for future deterioration in QoL are to be elucidated, as only baseline QoL demonstrated consistent significance across various algorithms tested.

While the current study focuses on the development and validation of predictive models, future work could benefit from the integration of explainability methods, such as SHapley Additive exPlanations (SHAP). These techniques can provide additional insights into model behaviour by highlighting the contribution of individual variables to specific predictions, thereby enhancing transparency and supporting clinical interpretability. Incorporating such approaches would further strengthen the robustness and practical applicability of the models in real-world clinical settings.

4.6 | Next steps and future research

The next steps in predicting glycaemic control include further research in different populations to assess the external validity of the model and its transferability. A final model could be adapted for use in clinical practice. Such a model would require the input of readily available patient characteristics and would output whether the patient is likely to experience inadequate glycaemic control within a year. This information could help target therapy and concentrate care on those people likely to suffer from inadequate glycaemic control. The adaptation of such a model for a deterioration in QoL in people with T2D remains a challenge, since the results of the models developed in this work would not allow for their use in clinical practice. Further research in the prediction of QoL in T2D should focus on models able to find the intricate pathways and potential subjectivity of this type of outcome. Perhaps the use of the Diabetes Specific Quality of Life Scale[43] would perform better for this goal. If the heterogeneity of T2D is indeed the cause of the ill-performing models for the prediction of QoL, a possible strategy would be to use the previously defined novel subgroups of T2D.[7,46] Using these subgroups might allow algorithms to find the patterns to use in prediction, rather than having to look for patterns in the highly heterogeneous total population of people with diabetes.

5 | CONCLUSIONS

The results of the current study show that prediction of inadequate glycaemic control after 1 year in T2D is feasible, in particular with SVC and MLP algorithms. Good results have been obtained in terms of discriminability with excellent AUC-ROCs and good F1-score metrics. A model using LR performed similarly to these two machine learning algorithms. Ensemble feature selection yielded better prediction results than using a single feature selection technique followed by hyperparameter fitting. Prediction of deterioration in QoL after 1 year was not feasible in the current population under the conditions used. Therefore, further research is required to elucidate the intricate network of pathways leading to changes in QoL in T2D.

ACKNOWLEDGEMENTS

The authors have nothing to report.

FUNDING INFORMATION

This study was supported by the European Regional Development Fund via OP-Zuid, the Province of Limburg, the Dutch Ministry of Economic Affairs (grant 31O.041), Stichting De Weijerhorst (Maastricht, The Netherlands), the Pearl String Initiative Diabetes (Amsterdam, The Netherlands), the Cardiovascular Center (CVC, Maastricht, the Netherlands), CARIM School for Cardiovascular Diseases (Maastricht, The Netherlands), CAPHRI Care and Public Health Research Institute (Maastricht, The Netherlands), NUTRIM School for Nutrition and Translational Research in Metabolism (Maastricht, the Netherlands), Stichting Annadal (Maastricht, The Netherlands), Health Foundation Limburg (Maastricht, The Netherlands), and by unrestricted grants from Janssen-Cilag B.V. (Tilburg, The Netherlands), Novo Nordisk Farma B.V. (Alphen aan den Rijn, the Netherlands), and Sanofi-Aventis Netherlands B.V. (Gouda, the Netherlands).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest relevant to the current work.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/dom.16598.

DATA AVAILABILITY STATEMENT

The data of this study derive from The Maastricht Study, but restrictions apply to the availability of these data, which were used under license for the current study. Data are, however, available from the

authors upon reasonable request and with permission of The Maastricht Study management team.

ORCID

Nikki C. C. Werkman https://orcid.org/0000-0002-3878-8302
Johannes T. H. Nielen https://orcid.org/0000-0002-3633-2504
Coen D. A. Stehouwer https://orcid.org/0000-0001-8752-3223

REFERENCES

1. Khan MAB, Hashim MJ, King JK, Govender RD, Mustafa H, Al Kaabi J. Epidemiology of type 2 diabetes – global burden of disease and forecasted trends. *J Epidemiol Glob Health*. 2020;10(1):107-111.
2. Committee ADAPP. 6. Glycemic targets: standards of medical Care in Diabetes—2022. *Diabetes Care*. 2021;45(Suppl_1):S83-S96.
3. Fang HSA, Gao Q, Tan WY, Lee ML, Hsu W, Tan NC. The effect of oral diabetes medications on glycated haemoglobin (HbA1c) in Asians in primary care: a retrospective cohort real-world data study. *BMC Med*. 2022;20(1):22.
4. Feingold K. *Oral and Injectable (Non-Insulin) Pharmacoogical Agents for the Treatment of Type 2 Diabetes*. Endotext; 2000. Accessed August 26, 2022. https://www.ncbi.nlm.nih.gov/books/NBK279141/
5. Laiteerapong N, Ham SA, Gao Y, et al. The legacy effect in type 2 diabetes: impact of early glycemic control on future complications (the Diabetes & Aging Study). *Diabetes Care*. 2019;42(3):416-426.
6. de Ritter R, Sep SJS, van der Kallen CJH, et al. Sex comparisons in the association of prediabetes and type 2 diabetes with cognitive function, depression, and quality of life: the Maastricht study. *Diabet Med*. 2023;40:e15115.
7. Werkman NCC, García-Sáez G, Nielen JTH, et al. Disease severity-based subgrouping of type 2 diabetes does not parallel differences in quality of life: the Maastricht study. *Diabetologia*. 2024;67(4):690-702.
8. Farooqi A, Gillies C, Sathanapally H, et al. A systematic review and meta-analysis to compare the prevalence of depression between people with and without type 1 and type 2 diabetes. *Prim Care Diabetes*. 2022;16(1):1-10.
9. Sivertsen H, Bjørkløf GH, Engedal K, Selbæk G, Helvik AS. Depression and quality of life in older persons: a review. *Dement Geriatr Cogn Disord*. 2015;40(5–6):311-339.
10. Semenkovich K, Brown ME, Svrakic DM, Lustman PJ. Depression in type 2 diabetes mellitus: prevalence, impact, and treatment. *Drugs*. 2015;75(6):577-587.
11. Islam MS, Qaraqe MK, Belhaouari S, Petrovski G. Long term HbA1c prediction using multi-stage CGM data analysis. *IEEE Sens J*. 2021;21(13):15237-15247.
12. Zaitcev A, Eissa MR, Hui Z, Good T, Elliott J, Benaissa M. A deep neural network application for improved prediction of HbA_1c in type 1 diabetes. *IEEE J Biomed Health Inform*. 2020;24(10):2932-2941.
13. Lavikainen P, Chandra G, Siirtola P, et al. Data-driven identification of Long-term glycemia clusters and their individualized predictors in Finnish patients with type 2 diabetes. *Clin Epidemiol*. 2023;15:13-29.
14. Patel MS, Polsky D, Small DS, et al. Predicting changes in glycemic control among adults with prediabetes from activity patterns collected by wearable devices. *NPJ Digit Med*. 2021;4(1):172.
15. Khan O, Badhiwala JH, Witiw CD, Wilson JR, Fehlings MG. Machine learning algorithms for prediction of health-related quality-of-life after surgery for mild degenerative cervical myelopathy. *Spine J*. 2021;21(10):1659-1669.
16. Candel-Parra E, Córcoles-Jiménez MP, Delicado-Useros V, Ruiz-Grao MC, Hernández-Martínez A, Molina-Alarcón M. Predictive model of quality of life in patients with Parkinson's disease. *Int J Environ Res Public Health*. 2022;19(2):672.
17. Karri R, Chen YP, Drummond KJ. Using machine learning to predict health-related quality of life outcomes in patients with low grade glioma, meningioma, and acoustic neuroma. *PLoS One*. 2022;17(5):e0267931.
18. Schram MT, Sep SJS, van der Kallen CJ, et al. The Maastricht study: an extensive phenotyping study on determinants of type 2 diabetes, its complications and its comorbidities. *Eur J Epidemiol*. 2014;29(6):439-451.
19. World Health Organisation IDF. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation. 2006.
20. Ware JE, Kosinski M, Keller SK. *SF-36 Physical and Mental Health Summary Scales: A User's Manual*. The Health Institute; 1994.
21. Steyerberg WE. Clinical prediction models. 2019.
22. Jeon H, Oh S. Hybrid-recursive feature elimination for efficient feature selection. *Appl Sci*. 2020;10(9):3211.
23. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*. 2011;12:2825-2830.
24. Ferri FJ, Pudil P, Hatef M, Kittler J. *Comparative Study of Techniques for Large-Scale Feature Selection. Machine Intelligence and Pattern Recognition*. Vol 16. Elsevier; 1994:403-413.
25. Abeel T, Helleputte T, Peer DVY, Dupont P, Saeys Y. Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. *Bioinformatics*. 2010;26(3):392-398.
26. Kursa BM, Jankowski A, Rudnicki RW. Boruta—a system for feature selection. *Fundam Inform*. 2010;101(4):271-285.
27. Saeys Y, Abeel T, Peer DVY. Robust feature selection using ensemble feature selection techniques. *Machine Learning and Knowledge Discovery in Databases*. Springer; 2008:313-325.
28. Zinkevich M. Rules of machine learning: best practices for ML engineering. 2017. https://developers.google.com/machine-learning/guides/rules-of-ml
29. Bickett A, Tapp H. Anxiety and diabetes: innovative approaches to management in primary care. *Exp Biol Med (Maywood)*. 2016;241(15):1724-1731.
30. Zakkerkish M, Shahbazian HB, Shahbazian H, Latifi SM, Moravej Aleali A. Albuminuria and its correlates in type 2 diabetic patients. *Iran J Kidney Dis*. 2013;7(4):268-276.
31. Coca SG, Ismail-Beigi F, Haq N, Krumholz HM, Parikh CR. Role of intensive glucose control in development of renal end points in type 2 diabetes mellitus: systematic review and meta-analysis intensive glucose control in type 2 diabetes. *Arch Intern Med*. 2012;172(10):761-769.
32. Al-Qerem W, Jarab AS, Badinjki M, Hammad A, Ling J, Alasmari F. Factors associated with glycemic control among patients with type 2 diabetes: a cross-sectional study. *Eur Rev Med Pharmacol Sci*. 2022;26(7):2415-2421.
33. Jarab AS, Al-Qerem W, Alqudah S, et al. Glycemic control and its associated factors in hypertensive patients with type 2 diabetes. *Eur Rev Med Pharmacol Sci*. 2023;27(12):5775-5783.
34. Gomes MB, Tang F, Chen H, et al. Socioeconomic factors associated with glycemic measurement and poor HbA1c control in people with type 2 diabetes: the global DISCOVER study. *Front Endocrinol (Lausanne)*. 2022;13:831676.
35. Bartels ECM, den Braver NR, Borgonjen-van den Berg KJ, Rutters F, van der Heijden A, Beulens JWJ. Adherence to the Dutch healthy diet index and change in glycemic control and cardiometabolic markers in people with type 2 diabetes. *Eur J Nutr*. 2022;61(5):2761-2773.
36. Pan B, Wu Y, Yang Q, et al. The impact of major dietary patterns on glycemic control, cardiovascular risk factors, and weight loss in patients with type 2 diabetes: a network meta-analysis. *J Evid Based Med*. 2019;12(1):29-39.
37. Mangen M-JJ, Bolkenbaas M, Huijts SM, van Werkhoven CH, Bonten MJM, de Wit GA. Quality of life in community-dwelling Dutch

elderly measured by EQ-5D-3L. *Health Qual Life Outcomes*. 2017; 15(1):3.
38. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol*. 2010;5(9):1315-1316.
39. Fan Y, Long E, Cai L, Cao Q, Wu X, Tong R. Machine learning approaches to predict risks of diabetic complications and poor glycemic control in nonadherent type 2 diabetes. *Front Pharmacol*. 2021; 12:665951.
40. Fu X, Wang Y, Cates RS, et al. Implementation of five machine learning methods to predict the 52-week blood glucose level in patients with type 2 diabetes. *Front Endocrinol (Lausanne)*. 2022;13:1061507.
41. Pearson ER. Type 2 diabetes: a multifaceted disease. *Diabetologia*. 2019;62(7):1107-1112.
42. de Jonge M, Wubben N, van Kaam CR, et al. Optimizing an existing prediction model for quality of life one-year post-intensive care unit: an exploratory analysis. *Acta Anaesthesiol Scand*. 2022;66(10):1228-1236.
43. Bott U, Mühlhauser I, Overmann H, Berger M. Validation of a diabetes-specific quality-of-life scale for patients with type 1 diabetes. *Diabetes Care*. 1998;21(5):757-769.
44. Misra R, Lager J. Predictors of quality of life among adults with type 2 diabetes mellitus. *J Diabetes Complications*. 2008;22(3):217-223.
45. Ho HT, Lin SI, Guo NW, Yang YC, Lin MH, Wang CS. Executive function predict the quality of life and negative emotion in older adults with diabetes: a longitudinal study. *Prim Care Diabetes*. 2022;16(4): 537-542.
46. Ahlqvist E, Prasad RB, Groop L. Subtypes of type 2 diabetes determined from clinical parameters. *Diabetes*. 2020;69(10):2086-2093.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Werkman NCC, Nielen JTH, Tapia-Galisteo J, et al. Prediction of glycaemic control and quality of life in people with type 2 diabetes using glucose-lowering drugs with machine learning—The Maastricht study. *Diabetes Obes Metab*. 2025;1-14. doi:10.1111/dom.16598