 Open Access Full Text Article

ORIGINAL RESEARCH

Data-Driven Identification of Long-Term Glycemia Clusters and Their Individualized Predictors in Finnish Patients with Type 2 Diabetes

Piia Lavikainen [1],*, Gunjan Chandra [2],*, Pekka Siirtola [2], Satu Tamminen [2], Anusha T Ihalapathirana [2], Juha Röning [2], Tiina Laatikainen [3–5], Janne Martikainen [1]

[1]School of Pharmacy, University of Eastern Finland, Kuopio, Finland; [2]Biomimetics and Intelligent Systems Group, Faculty of ITEE, University of Oulu, Oulu, Finland; [3]Joint Municipal Authority for North Karelia Social and Health Services (Siun Sote), Joensuu, Finland; [4]Department of Public Health and Social Welfare, Finnish Institute for Health and Welfare, Helsinki, Finland; [5]Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland

*These authors contributed equally to this work

Correspondence: Piia Lavikainen, School of Pharmacy C/O Faculty of Health Sciences, University of Eastern Finland, P.O. Box 1627, Kuopio, FI-70211, Finland, Tel +358 40 7024682, Email piia.lavikainen@uef.fi

Purpose: To gain an understanding of the heterogeneous group of type 2 diabetes (T2D) patients, we aimed to identify patients with the homogenous long-term HbA1c trajectories and to predict the trajectory membership for each patient using explainable machine learning methods and different clinical-, treatment-, and socio-economic-related predictors.

Patients and Methods: Electronic health records data covering primary and specialized healthcare on 9631 patients having T2D diagnosis were extracted from the North Karelia region, Finland. Six-year HbA1c trajectories were examined with growth mixture models. Linear discriminant analysis and neural networks were applied to predict the trajectory membership individually.

Results: Three HbA1c trajectories were distinguished over six years: "stable, adequate" (86.5%), "improving, but inadequate" (7.3%), and "fluctuating, inadequate" (6.2%) glycemic control. Prior glucose levels, duration of T2D, use of insulin only, use of insulin together with some oral antidiabetic medications, and use of only metformin were the most important predictors for the long-term treatment balance. The prediction model had a balanced accuracy of 85% and a receiving operating characteristic area under the curve of 91%, indicating high performance. Moreover, the results based on SHAP (Shapley additive explanations) values show that it is possible to explain the outcomes of machine learning methods at the population and individual levels.

Conclusion: Heterogeneity in long-term glycemic control can be predicted with confidence by utilizing information from previous HbA1c levels, fasting plasma glucose, duration of T2D, and use of antidiabetic medications. In future, the expected development of HbA1c could be predicted based on the patient's unique risk factors offering a practical tool for clinicians to support treatment planning.

Keywords: type 2 diabetes, cluster, HbA1c, machine learning, SHAP

Introduction

Type 2 diabetes (T2D), which is a complex disease characterized by hyperglycemia, is a significant global health challenge, affecting about 537 million, ie, 1 in 10 adults worldwide (IDF 2021).[1] Globally, the economic burden associated with T2D was estimated to be as much as $966 billion in 2021.[1] Most of the health expenditures are caused by T2D-related micro- and macrovascular complications.[2] These unfavorable figures could be expected to increase even higher since the prevalence of T2D is rising due to ageing, increasing obesity, unhealthy diet, and physical inactivity.

Patients with T2D are a heterogeneous patient group with varying degrees of insulin resistance and impaired insulin secretion leading to differences in needs for treatment intensity and health care service use. In the treatment of T2D, numerous previous studies have demonstrated that high average blood glucose level, commonly estimated with

13

hemoglobin A1c (HbA1c), is significantly associated with T2D-related micro- and macrovascular complications.[3–7] However, aiming for normal or close-to-normal HbA1c levels is not always preferred due to pathogenic mechanisms related to the development of T2D-related complications.[8] Therefore, standardized mostly "one-size-fits-all" care is currently suggested by the treatment guidelines with more-stringent goals for younger T2D patients with short disease duration and less-rigorous treatment targets for older and the most comorbid patients.[9–12] The level of HbA1c is, however, often observed to deteriorate and fluctuate over time, eg, due to progressive nature of the disease, ageing, changes in lifestyle, other concomitant diseases, and different treatments, which is also known to be a significant risk factor for T2D-related morbidity and mortality.[13–15] Thus, there is an urgent need to identify patients with challenges in T2D care and to optimize the selection of treatments and the use of health care resources by enabling different models for care stream.

More specifically, individualized prediction models based on patient's characteristics are needed to aid physicians and health care professionals in identification of the patients with poor treatment balance and tailoring treatment and monitoring according to the needs of the patients. In the present study, which is a part of the HTx project (EU Horizon 2020 funded project 2019–2024), we applied artificial intelligence as a tool to examine individualized predictions by searching complex relationships from high-dimensional data. The primary aim of HTx is to create a framework for the Next Generation Health Technology Assessment (HTA) to support patient-centered, societally oriented, real-time decision-making on access to and reimbursement for health technologies throughout Europe. To achieve the goals, we characterized clusters of similar patients based on their HbA1c trajectories over six years and predicted the membership to the long-term glycemic control trajectories for each patient using explainable artificially intelligent methods and different clinical-, treatment-, and socioeconomic-related predictors to improve our understanding of sub-groups and their characteristics of this heterogeneous patient population. These developed methods may help to predict the expected development of longitudinal HbA1c trajectories when planning the treatment of patients based on their unique risk factors and other characteristics.

Materials and Methods
Study Design
In Finland, health care services are mainly produced as public services and are organized and funded by municipalities. Siun sote – The Joint Municipal Authority for North Karelia Social and Health Services was established in 2017 by joining healthcare organizations of 13 municipalities. Before that, all the municipalities had introduced identical electronic health register (EHR) called Mediatri in 2010–2011, and it has been in use since integration both in primary and specialized care, too.[16] EHR data were obtained from the Siun sote to identify a population-based cohort of patients with a diagnosed T2D at the end of 2012 as identified with the 10th revision of International Classification of Diseases (ICD-10) code E11 (N = 10139). EHRs provided us the data on utilization of both primary health care and specialized care for all patients living in the North Karelia region.

EHR data were compiled with medication purchase data for the years 1995–2019 from the Finnish Prescription Register (FPR) maintained by the Social Insurance Institution of Finland (SII). Data from the FPR include, for example, date of medication purchase, Anatomical Therapeutic Chemical (ATC) classification code,[17] and cost.

Data on entitlements to higher medication reimbursements before 2011 were obtained from the Special Reimbursement Register (SRR) also maintained by the SII. In Finland, all community-dwelling residents are eligible for reimbursement for prescription medications, and thus, all reimbursed medication purchases and entitlements to higher medication reimbursement are registered to the beforementioned registers maintained by the SII. Data from the FPR and SRR were utilized to ascertain the date of T2D diagnosis as the first record of either diabetes medication purchase, entitlement to special refund or T2D diagnosis in the EHRs in the case that a patient had T2D diagnosed in a healthcare provider outside of the Siun sote. The study cohort was restricted to those patients who were living in the region of North Karelia in 2014 and were alive at the baseline on Jan 1, 2014 (n = 9631).

14

Clinical Epidemiology 2023:15

Glycemic Control

Glycemic control was measured with HbA1c with the turbidimetric inhibition immunoassay method (TINIA) in the Eastern Finland laboratory (ISLAB, https://www.islab.fi). The method and its analysis were standardized across the North Karelia region. Values were standardized to the International Federation of Clinical Chemistry (IFCC) units. Yearly measures of HbA1c in 2014–2019 were utilized for data-driven clustering of long-term glycemic control. If several measurements occurred within a year, the yearly mean was calculated. If the HbA1c value was unmeasured within a certain year, it was considered as a missing value for that year. HbA1c values were measured until the date of death, moving outside of the study area, or 31st December 2019, whichever occurred first.

Antidiabetic Medication Use

Use of antidiabetic medications in 2014–2019 was examined among patients who were alive at the end of the examination year from the FPR and classified as follows: users of metformin (ATC code A10BA02) only, users of metformin and other oral antidiabetic medications (A10BA02 + other A10B), users of only other diabetes medications than insulin or metformin (A10B excluding A10BA02 and A10A), users of insulin and oral antidiabetic medication (including metformin) (A10A + A10B), users of insulin (A10A) only, and patients without antidiabetic medications.

Baseline Characteristics for Prediction

A detailed list of baseline characteristics measured prior to January 1st, 2014, with their definitions is presented in Supplementary Table 1. Three types of variables were available: clinical- and treatment-related factors as well as small-area level factors on socio-economic status (SES).

Clinical factors included age and gender obtained from the EHRs. Duration of T2D was defined at the baseline based on the first record of either diabetes medication purchase, entitlement to a special refund for antidiabetic medications, or T2D diagnosis in the EHRs. For baseline HbA1c, fasting plasma glucose (mmol/l), body mass index (BMI; kg/m^2), total cholesterol (mmol/l), low-density lipoprotein cholesterol (mmol/l) and creatinine (µmol/l), the most recent measures before January 1st, 2014, were extracted from the EHRs.

Clinical factors included concordant (such as hypertension, coronary artery disease, atrial fibrillation, heart failure, peripheral arterial diseases, stroke, chronic kidney failure, neuropathies, and blindness) and discordant (such as cancers, asthma, gout, glaucoma, depression, dementia, mental diseases, chronic obstructive pulmonary disease, rheumatoid and other arthritis, osteoporosis, neuromuscular diseases, and liver diseases) diseases that occurred prior to the baseline in 2014 as defined from the EHRs. Micro- and macrovascular complications occurring prior to the baseline were identified based on definitions from the minimum data content of the Finnish Diabetes Quality Register.[18] In addition, 34 disease groups were formed based on categorizations of ICD-10 codes (Supplementary Table 1).

Treatment-related factors included the following subgroups of antidiabetic medication use in 2013 formed based on data from the FPR: users of metformin (ATC code A10BA02) only, users of metformin and other oral antidiabetic medications (OAD) (incl. metformin, sulfonylureas, combinations of oral blood glucose lowering drugs, glitazones, DPP-4 inhibitors, glinides, GLP-1 analogues, and SGLT2 inhibitors).

(A10BA02 + other A10B), users of only other antidiabetic medications than insulin or metformin (A10B excluding A10BA02 and A10A), users of insulin and oral antidiabetic medication (including metformin) (A10A + A10B), and users of insulin (A10A) only. Utilization of other than antidiabetic medications was identified based on the third level of ATC codes, and the total number of medications used in 2013 was calculated based on the number of different ATC codes. Based on HbA1c, LDL and BMI measurements taken before baseline, indicators of having HbA1c, LDL and BMI values measured before the baseline were created. In addition, variables on number of HbA1c measurements in 2013 and number of all laboratory measurements taken (incl. blood pressure) in 2013 were formed. Finally, use of health care services was measured as numbers of 1) visits and 2) phone calls for physicians and nurses for any reason separately and as a total during the year 2013.

Small-area level SES factors included, for example, income status and education status of patient's living area, which were obtained from the postal code area database[19] maintained by Statistics Finland. We used databases from 2016 to

2017 to gather information measured on December 31st, 2013. According to our previous study, small-area level variables provide similar effects to patient-level information and are a usable proxy in the absence of patient-level information.[20]

Statistical Methods

Trajectory Modelling

The growth mixture model (GMM) is a data-driven statistical method to examine the existence of homogeneous subpopulations within a heterogeneous population.[21–24] The method is suitable for longitudinal data to study evolution, ie, trajectories, over time and it assumes that each patient's trajectory over time results from a patient being a member of a latent unobserved subpopulations or class. Other trajectory models for longitudinal designs include latent class growth models (or alternatively, group-based trajectory models), latent transition analysis and sequence analysis (only for categorical data).[25] Of these, GMM is a parametric model and the only method that allows for heterogeneity within subpopulations. GMMs were utilized to cluster T2D patients into distinct trajectories of glycemic control according to their HbA1c patterns from 2014 to 2019. The models were fitted iteratively with 1–5 classes and varying shapes for the trajectories (linear, quadratic, cubic). GMMs were estimated with full-information maximum likelihood, and missing data were not imputed, but all available data were used. The selection of the final model was based on information from fit indices and classification performance of the model as well as clinical interpretation of the clusters. Bayesian Information Criteria (BIC) was utilized to compare models with lower values indicating better fit. Low-Mendel-Rubin likelihood ratio test with a P-value <0.05 was considered to indicate a better fit for an n-class model than for the n–1-class model. Posterior probabilities >0.8 are recommended and values close to 1 indicate better classification. In addition, entropy was used to guide the classification accuracy of the model with higher values indicating better classification. Small classes with <5% of the population were not accepted to end up with large enough groups of patients. GMM analyses were performed using Mplus Version 8.[26]

Differences in baseline characteristics between those included and excluded to the study were examined with standardized mean differences with values ≥10% considered to indicate meaningful differences between the groups.[27] These analyses were conducted with SAS 9.4 (SAS Institute Inc., Cary, North Carolina, USA).

Prediction Modelling

To predict the patient's trajectory membership, several machine learning algorithms using Python were implemented. Two machine learning models, a custom-built fully connected neural network (NN) using TensorFlow[28] and linear discriminant analysis (LDA) from the sklearn,[29] were selected as final models.

LDA is a method used to find a linear combination of features that characterize or discriminate two or more classes of events.[30] LDA is a classification method that assumes that distinct classes generate data based on different Gaussian distributions. To generate or train a classifier, the fitting function estimates the parameters of Gaussian distribution for each class; the model has the same covariance matrix for each class, only the means vary. However, predicting classes of new data using a trained discriminant classifier finds the class with the smallest misclassification cost.

The LDA instance used in this study can be derived from a simple probabilistic model, which models the class conditional distribution of the data for each class k. Prediction can be obtained using Bayes' rule for each training sample $x \in R^d$:

$$P(y = k|x) = \frac{P(x|y = k)P(y = k)}{P(x)} = \frac{P(x|y = k)P(y = k)}{\sum_l P(x|y = l) \cdot P(y = l)}$$

The class k, which maximizes the posterior probability, is selected. For LDA, $P(x|y)$ is modeled as a multivariate Gaussian distribution with density defined as:

$$P(x/y = k) = \frac{1}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1}(x - \mu_k)\right)$$
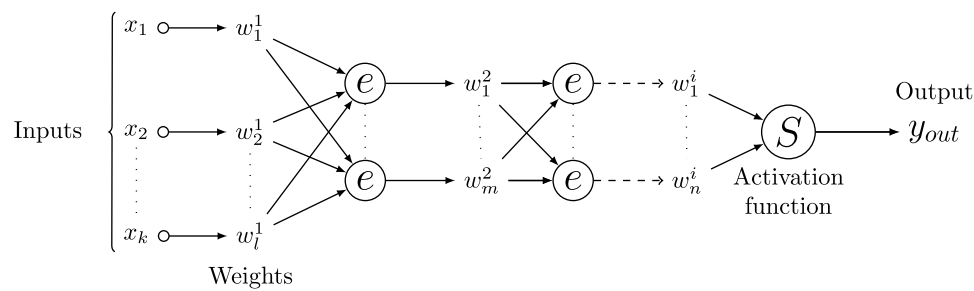
Figure 1 Structure of the fully connected neural network.

where d is the number of features, Σ_k is covariance matrix, and the term $(x - \mu_k)^T \Sigma_k^{-1}(x - \mu_k)$ corresponds to the Mahalanobis Distance between the sample x and the mean μ_k.

NNs, also known as artificial NNs, are systems inspired by biological neural networks consisting of artificial nodes. The primary goal of the NN was to solve problems as a human brain would. The NN determines the appropriate mathematical use to turn the input into output whether they have a linear or non-linear relationship. The NN used in this study is a fully connected multi-layered network with "elu" as an activation function in the hidden layer and a sigmoid as the activation function in the output layer.[27] The structure of NNs used in this study can be seen in Figure 1, for each input/observation, k is the number of features and i is the number of layers in the model and l, m and n represent the number of neurons in each layer. Further details on the selection of model's parameter can be found from the supplementary material under Appendix 1.

The feature selection based on the most important feature is a preliminary step used to select a minimum number of features from the data to train and test a model whilst maintaining the performance of models; the Shapley additive explanations (SHAP) plots,[31] on the other hand, explain how a model while training and testing used a particular predictor for creating a decision boundary and for constructing a prediction. SHAP is an approach based on game theory to explain the output of machine learning models. It allocates optimal credits with local explanations using Shapley values from the game theory and their connected extensions. SHAP explains the contribution of each feature to each individual predicted probability. In this study, SHAP is used to understand how diverse values of various features lead to the probability of belonging to a glycemic control class.

Data Pre-Processing

The data included 244 variables or features, out of which 9, consisting of a single value, were removed. Furthermore, features with equal and more than 40% of missing values were dropped. A total of 59 out of 66 SES features were converted into ratio variables to eliminate data bias. For each of the three predictor categories, the best k predictors were selected using the Python function called SelectKBest from sklearn.[30] Table 1 lists the categories of predictors with the total number and number of selected predictors. Samples with missing values in the training set were imputed with the most frequent value of the corresponding feature for that class, and samples with missing values in the testing set were dropped. The training set was under-sampled to match the number of samples for both classes using a random under-sampler. Later, datasets were normalized using the standard scaler method.

Table 1 Types of Predictors with Total Number of Predictors and Number of Selected Predictors

Type of Predictors	Total Number of Predictors	Number of Selected Predictors
Clinical (C)	83	5
Clinical + Treatment (CT)	233	10
Clinical + Treatment + SES (CTS)	299	15

Ethics Statement

Use of the data was approved by the Ethics Committee of the Northern Savonia Hospital District (diary number 81/2012). The study protocol was also approved by the register administrator, the Siun sote. A separate permission to link data on medication purchases and special reimbursements was achieved from the SII (diary number 110/522/2018). Only register-based data were utilized and, in accordance with Finnish legislation, consent from the patients was not needed. The study complies with the Declaration of Helsinki.

Results

HbA1c Trajectories

Data consisted of 9631 T2D patients of whom 9048 had at least one HbA1c measurement during the follow-up in 2014–2019. Excluded patients without HbA1c measurement in 2014–2019 were older, had more discordant diseases and complications but less healthcare contacts and antidiabetic medication use than those included in further analyses (Supplementary Table 2). Mean (SD) age of the patients was 69.2 (11.6) years and 47.2% were female (Table 2). Average time since T2D diagnosis was 8.9 (6.4) years and a patient had utilized an average 8.6 (4.5) different medications in 2013. The most frequently used antidiabetic medication in 2013 was metformin (33.9%) followed by insulin together with oral antidiabetic medication (22.0%).

A three-class cubic GMM was identified as the final, best fitting model with good interpretation for HbA1c development in 2014–2019 (Supplementary Table 3, Figure 2). The selected model is graphically presented in Supplementary Figure 1. The model included variances for the intercept and linear slope growth factors but, however, intercept and linear slope growth factors were not allowed to correlate to improve model fit. For the selected trajectory

Table 2 Baseline Characteristics and Variables Describing Treatment Monitoring During the Follow-Up Overall and by the Estimated HbA1c Clusters

	Total Population	Stable, Adequate Glycemic Control	Fluctuating Inadequate Glycemic Control	Improving, But Inadequate Glycemic Control
n (%)	9048 (100.0)	7829 (86.5)	562 (6.2)	657 (7.3)
Baseline characteristics				
Female, n (%)	4271 (47.2)	3732 (47.7)	255 (45.4)	284 (43.2)
Mean age (SD), years	69.2 (11.6)	69.3 (11.3)	68.5 (13.1)	69.2 (13.0)
Mean time since T2D diagnosis, years (SD)	8.9 (6.4)	8.3 (6.0)	11.7 (7.1)	13.6 (7.5)
HbA1c measured, n (%)	8604 (95.1)	7445 (95.1)	530 (94.3)	629 (97.9)
Mean HbA1c (SD), mmol/mol	48.8 (13.1)	46.1 (10.2)	59.9 (13.8)	71.3 (16.8)
LDL measured, n (%)	8397 (92.8)	7299 (93.2)	507 (90.2)	591 (90.0)
Mean LDL (SD), mmol/l	2.5 (0.9)	2.5 (0.9)	2.5 (0.9)	2.5 (0.9)
Total cholesterol measured, n (%)	7727 (85.4)	6695 (85.5)	480 (85.4)	552 (84.0)
Mean total cholesterol (SD), mmol/l	4.3 (1.0)	4.3 (1.0)	4.3 (1.0)	4.4 (1.2)
HDL measured, n (%)	7229 (79.9)	6261 (80.0)	439 (78.1)	529 (80.5)
Mean HDL (SD), mmol/l	1.2 (0.4)	1.3 (0.4)	1.1 (0.3)	1.1 (0.3)
Triglycerides measured, n (%)	7234 (80.0)	6265 (80.0)	440 (78.3)	529 (80.5)
Mean triglycerides (SD), mmol/l	1.5 (1.0)	1.5 (0.9)	1.7 (1.7)	1.8 (2.0)

(*Continued*)

18

Clinical Epidemiology 2023:15

Table 2 (Continued).

	Total Population	Stable, Adequate Glycemic Control	Fluctuating Inadequate Glycemic Control	Improving, But Inadequate Glycemic Control
FPG measured, n (%)	7452 (82.4)	6468 (82.6)	447 (79.5)	537 (81.7)
Mean FPG (SD), mmol/l	7.9 (2.4)	7.6 (1.9)	9.2 (3.2)	10.5 (3.9)
Creatinine measured, n (%)	8845 (97.8)	7662 (97.9)	540 (96.1)	643 (97.9)
Mean creatinine (SD), μmol/l	78.3 (33.1)	77.5 (32.7)	83.2 (33.1)	82.9 (37.5)
BMI measured, n (%)	6855 (75.8)	5973 (76.3)	410 (73.0)	472 (71.8)
Mean BMI (SD), kg/m^2	31.5 (6.3)	31.2 (6.1)	33.3 (6.5)	33.5 (7.5)
Concordant diseases, n (%)	7215 (79.7)	6174 (78.9)	484 (86.1)	557 (84.8)
Discordant diseases, n (%)	4630 (51.2)	3939 (50.3)	307 (54.6)	384 (58.4)
Prevalence of complications, n (%)	4083 (45.1)	3394 (43.4)	306 (54.4)	383 (58.3)
Microvascular	1980 (21.9)	1579 (20.2)	171 (30.4)	230 (35.0)
Macrovascular	3054 (33.8)	2545 (32.5)	231 (41.1)	278 (42.3)
Diabetes medication use in 2013, n (%)				
Metformin only	3063 (33.9)	2950 (37.7)	76 (13.5)	37 (5.6)
Metformin + other OAD	1274 (14.1)	1122 (14.3)	82 (14.6)	70 (10.7)
Only other OAD than metformin or insulin	708 (7.8)	657 (8.4)	30 (5.3)	21 (3.2)
Insulin + OAD	1989 (22.0)	1423 (18.2)	253 (45.0)	313 (47.6)
Insulin only	826 (9.1)	556 (7.1)	97 (17.3)	173 (26.3)
No antidiabetics	1188 (13.1)	1121 (14.3)	24 (4.3)	43 (6.5)
Mean number of medications in 2013 (SD)	8.6 (4.5)	8.4 (4.4)	9.6 (4.8)	10.1 (5.2)
Mean number of comorbidities (SD)	6.6 (3.3)	6.6 (3.2)	6.9 (3.3)	7.1 (3.5)
Mean number of contacts[a] in 2013 (SD)	7.3 (8.7)	7.1 (8.4)	7.7 (9.0)	9.0 (11.6)
Mean number of HbA1c measurements in 2013 (SD)	1.7 (1.2)	1.6 (1.1)	1.9 (1.3)	2.0 (1.3)
Mean number of laboratory measurements in 2013 (SD)	11.7 (7.9)	11.4 (7.5)	12.9 (10.1)	14.0 (9.9)
Variables describing treatment monitoring during the follow-up in 2014–2019[b]				
No. of HbA1c measurements, mean (SD)	8.57 (5.38)	8.50 (5.29)	9.97 (5.78)	8.37 (5.89)
Mean HbA1c (SD)	55.0 (14.9)	52.1 (12.1)	73.7 (18.0)	72.2 (17.7)

Notes: [a]Including phone calls and visits to nurses and physicians. [b]Mean follow-up 1948 days (5.3 years).
Abbreviations: BMI, body mass index; FPG, fasting plasma glucose; HDL, high-density lipoprotein; LDL, low-density lipoprotein; OAD, oral antidiabetic drug or GLP-1 analogue (incl. metformin, sulfonylureas, combinations of oral blood glucose lowering drugs, glitazones, DPP-4 inhibitors, glinides, GLP-1 analogues, and SGLT2 inhibitors).

model, average posterior probability was 0.842, entropy 0.892 and Low-Mendel-Rubin likelihood ratio test p = 0.024. To note, some models had lower BIC than the selected model, but some of the extracted classes were too small (<5%) to be used in further analyses. The identified trajectories were named as "stable, adequate" (86.5%), "improving, but inadequate" (7.3%), and "fluctuating, inadequate" (6.2%) glycemic control clusters.
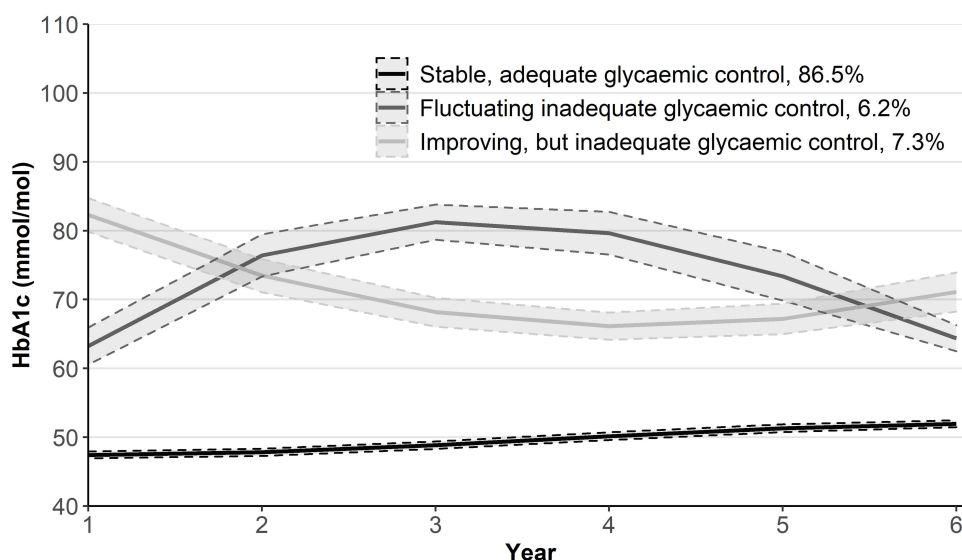
Figure 2 Estimated HbA1c trajectories. Solid line represents average HbA1c and dotted lines 95% confidence interval for the mean.

Patients with inadequate glycemic control had longer duration of T2D, higher BMI and creatinine level, had more frequently comorbidities and higher prevalence of micro- and macrovascular complications than patients with stable, adequate glycemic control (Table 2). Additionally, antidiabetic medication structure differed by the HbA1c trajectories during the follow-up (Figure 3). Patients belonging to "improving, but inadequate" and "fluctuating, inadequate" trajectories used more often insulin only or insulin together with oral antidiabetic medication, but less often metformin only than patients from the "stable, adequate" trajectory.

Prediction Modelling

For prediction of the trajectory membership, "improving, but inadequate" (7.3%) and "fluctuating, inadequate" (6.2%) glycemic control trajectories were combined into one "inadequate" class. The binary classification was performed using LDA and NN models to predict the class membership (adequate/inadequate). Both models were trained in cascading order with the type of predictors mentioned in Table 1 using 4-fold cross-validation. For each cross-validation run, data was divided into four parts; three parts were used for training and one part for testing. Figures 4 show the best predictors selected for model training and testing for each predictor type.
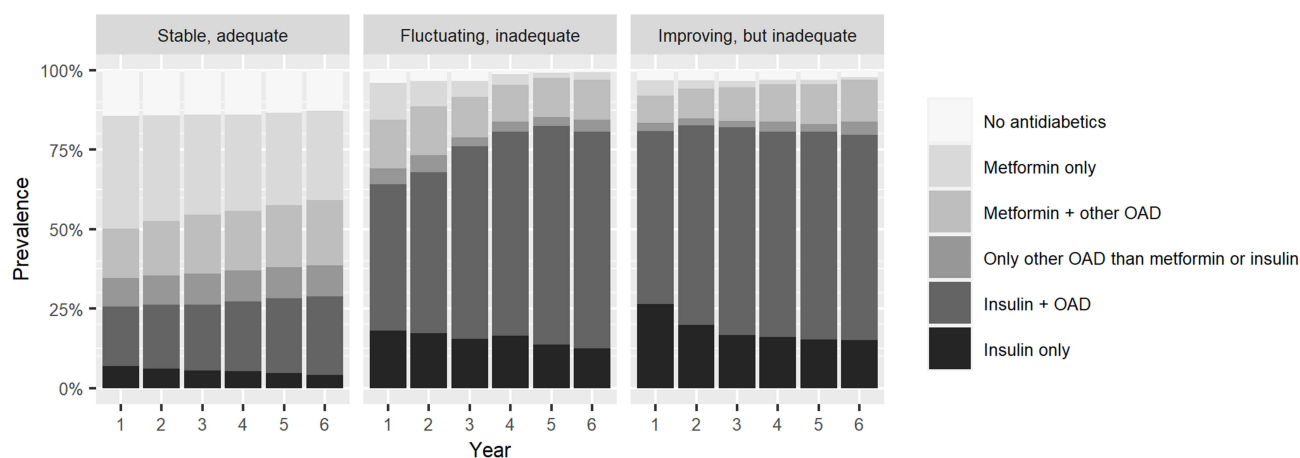

Figure 3 Prevalence of antidiabetic medication use by the estimated HbA1c trajectories. Abbreviations: OAD, oral antidiabetic drugs or GLP-1 analogues (incl. metformin, sulfonylureas, combinations of oral blood glucose lowering drugs, glitazones, DPP-4 inhibitors, glinides, GLP-1 analogues, and SGLT2 inhibitors).
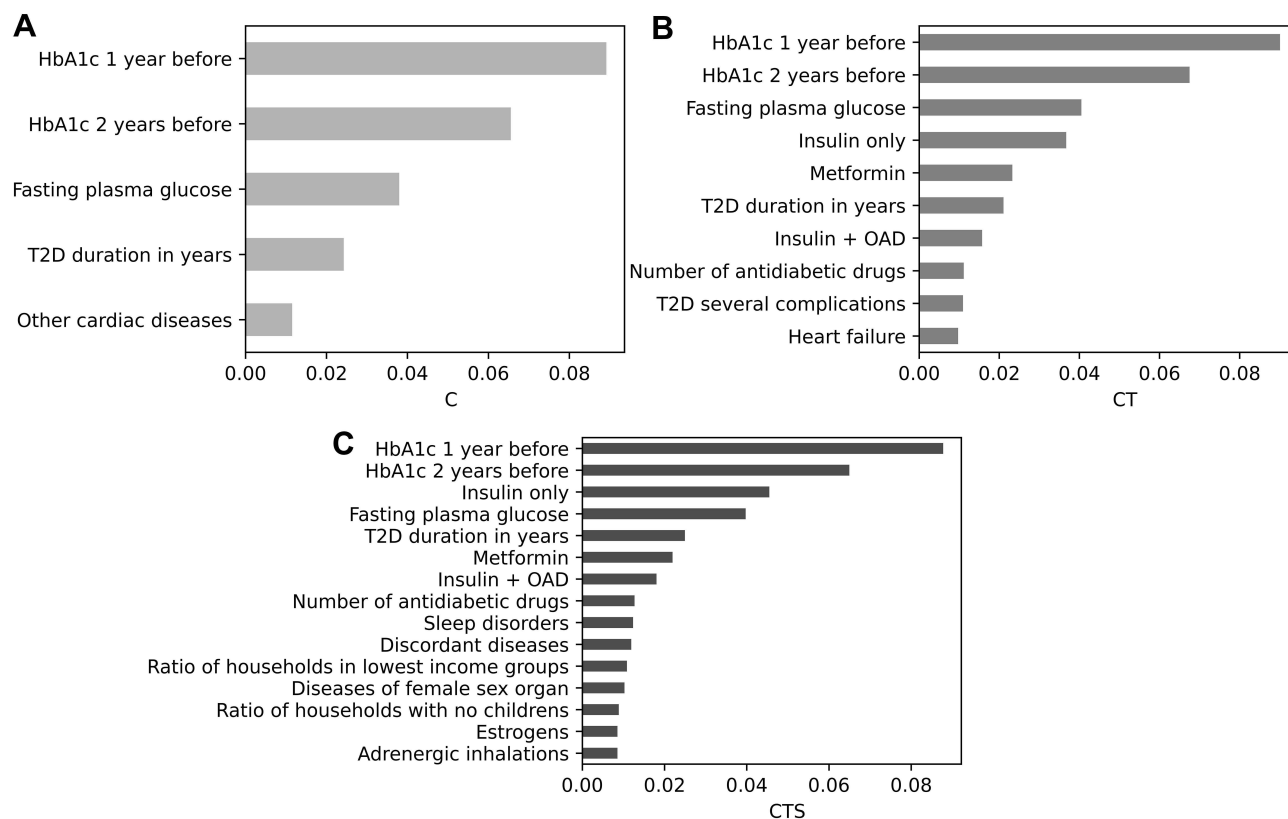
A

HbA1c 1 year before
HbA1c 2 years before
Fasting plasma glucose
T2D duration in years
Other cardiac diseases

0.00 0.02 0.04 0.06 0.08
C

B

HbA1c 1 year before
HbA1c 2 years before
Fasting plasma glucose
Insulin only
Metformin
T2D duration in years
Insulin + OAD
Number of antidiabetic drugs
T2D several complications
Heart failure

0.00 0.02 0.04 0.06 0.08
CT

C

HbA1c 1 year before
HbA1c 2 years before
Insulin only
Fasting plasma glucose
T2D duration in years
Metformin
Insulin + OAD
Number of antidiabetic drugs
Sleep disorders
Discordant diseases
Ratio of households in lowest income groups
Diseases of female sex organ
Ratio of households with no childrens
Estrogens
Adrenergic inhalations

0.00 0.02 0.04 0.06 0.08
CTS

Figure 4 Feature importance plot for (**A**) Clinical (C), (**B**) Clinical + Treatment (CT), (**C**) and Clinical + Treatment + SES (CTS) models. OAD, oral antidiabetic drugs or GLP-1 analogues (incl. metformin, sulfonylureas, combinations of oral blood glucose lowering drugs, glitazones, DPP-4 inhibitors, glinides, GLP-1 analogues, and SGLT2 inhibitors). Other cardiac diseases refer to other diseases of the heart and pulmonary circulation.

After each model was trained and tested for each predictor type for all four folds, the testing results were concatenated to make the final evaluation of the LDA and NN models. Table 3 shows the confusion matrix and three other evaluation parameter results for the complete analysis. Figure 5 shows the variation in the performance of both LDA and NN models over different folds for all predictor types. Figures 6 and 7 show the SHAP plots for the LDA and NN models for the Clinical + Treatment + SES (CTS) predictor type from the last split of the four-folds.

Every model performed well with a balanced accuracy of 85% and receiver-operating characteristic area under the curve (ROC AUC) ranging from 91% to 92% (Table 3). The kappa coefficient for all models ranged between 0.35 and 0.4. The prediction of the inadequate class was more critical than the adequate class; therefore, the F1 macro score was used for the evaluation, which showed promising results with a score of 69% for LDA and 66% for NN. As more

Table 3 Models with Their Overall Performance After 4-Fold Cross-Validation

Predictors	Model		Confusion Matrix			F1 Score	Balanced Accuracy	ROC AUC
Clinical	**LDA**	True	Predicted		Class	0.69	0.84	0.92
			4394	726	Adequate			
			68	329	Inadequate			
	NN	True	Predicted		Class	0.66	0.85	0.91
			4191	929	Adequate			
			47	350	Inadequate			

(*Continued*)

Table 3 (Continued).

Predictors	Model	Confusion Matrix				F1 Score	Balanced Accuracy	ROC AUC
Clinical + Treatment	LDA	True	**Predicted**		**Class**	0.69	0.85	0.92
			4405	715	Adequate			
			67	330	Inadequate			
	NN	True	**Predicted**		**Class**	0.66	0.85	0.91
			4202	918	Adequate			
			48	349	Inadequate			
Clinical + Treatment + SES	LDA	True	**Predicted**		**Class**	0.69	0.84	0.92
			4402	714	Adequate			
			70	326	Inadequate			
	NN	True	**Predicted**		**Class**	0.66	0.84	0.91
			4249	867	Adequate			
			57	339	Inadequate			

Abbreviations: LDA, linear discriminant analysis; NN, neural network; ROC AUC; receiver operating characteristic area under the curve; SES, socio-economic status.

predictors were added for prediction, the model's mean performance was improved along with the model's confidence as variance decreased (Figure 5). Of all the models, NN with CTS predictors was the best performing model for the inadequate class with the lowest number of misclassified samples and showed higher overall confidence.
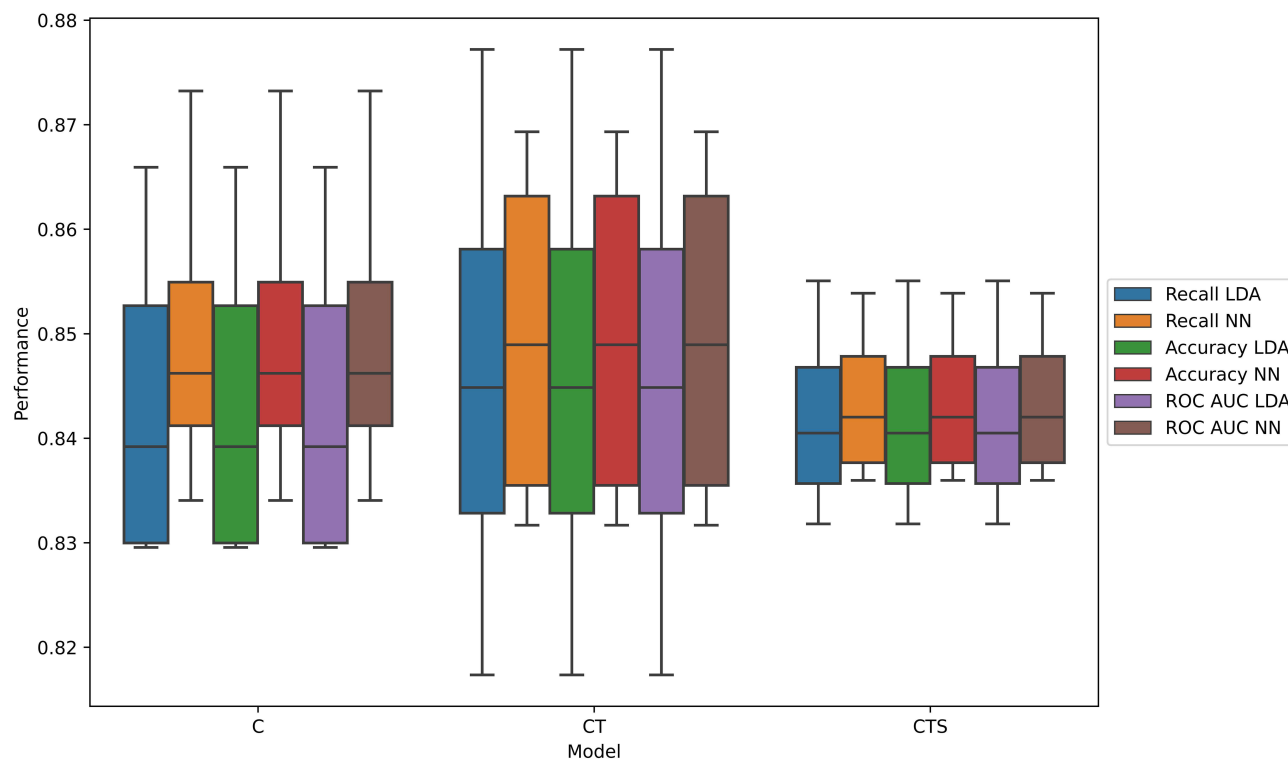
Figure 5 Performance of Clinical (C), Clinical + Treatment (CT), and Clinical + Treatment + Socioeconomic (CTS) models over different splits in 4-fold cross-validation. LDA, linear discriminant analysis; NN, neural networks; ROC AUC; receiver operating characteristic area under the curve.
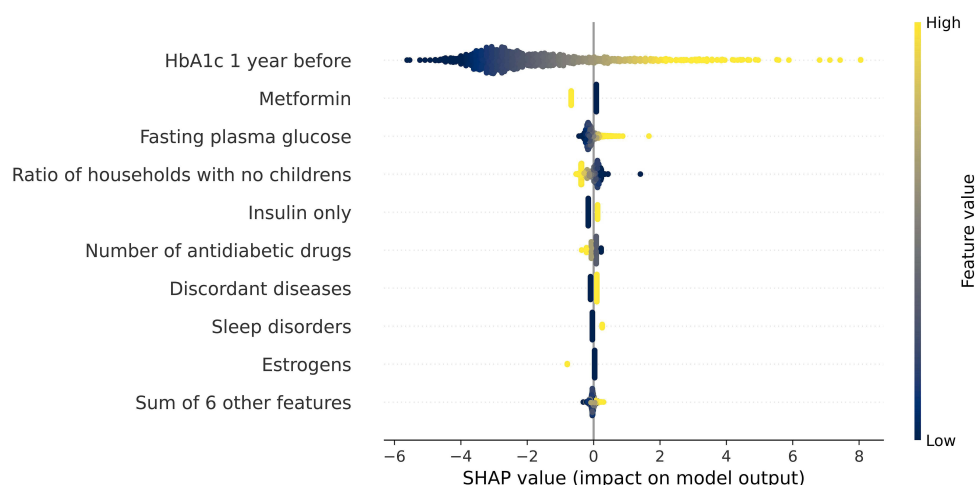
Figure 6 SHAP summary plot for one of the linear discriminant analysis models for clinical + treatment + socioeconomic predictors. Six other features consist of estrogens, diseases of female sex organ, adrenergic inhalations, discordant diseases, insulin + OAD, and T2D duration in years. OAD, oral antidiabetic drugs or GLP-1 analogues (incl. metformin, sulfonylureas, combinations of oral blood glucose lowering drugs, glitazones, DPP-4 inhibitors, glinides, GLP-1 analogues, and SGLT2 inhibitors).
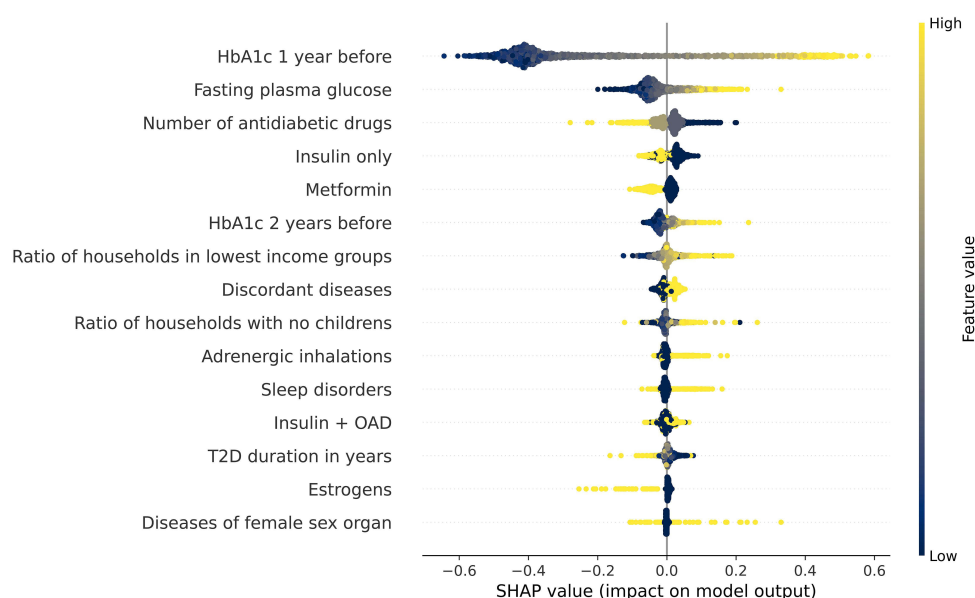

Figure 7 SHAP summary plot for one of the neural network models for clinical + treatment + socioeconomic predictors. OAD, oral antidiabetic drugs or GLP-1 analogues (incl. metformin, sulfonylureas, combinations of oral blood glucose lowering drugs, glitazones, DPP-4 inhibitors, glinides, GLP-1 analogues, and SGLT2 inhibitors).

The prediction of the membership to a trajectory pathway is not enough; it is equally important to be able to explain the reasons behind the prediction. While previous HbA1c levels, use of insulin only, fasting plasma glucose, duration of T2D, use of only metformin, use of insulin together with some OAD, and number of antidiabetic drugs were the most important predictors for CTS models training (Figure 4C), HbA1c level one year before showed the most significant influence in the prediction in both LDA and NN models (Figures 6 and 7). The SHAP summary plot (Figures 6 and 7) shows the overall impact of each feature in the order of importance, top being of the highest importance, with the ranging feature values. For any feature value, if the SHAP value is higher than 0, it indicates that the sample will belong to the inadequate class. For example, based on the SHAP summary plot (Figure 7), the probability of belonging to the inadequate class increased if any discordant disease or sleep disorder was present. In contrast, the total number of antidiabetic drugs had a negative effect on the prediction, ie, a higher number of diabetic medications decreased the
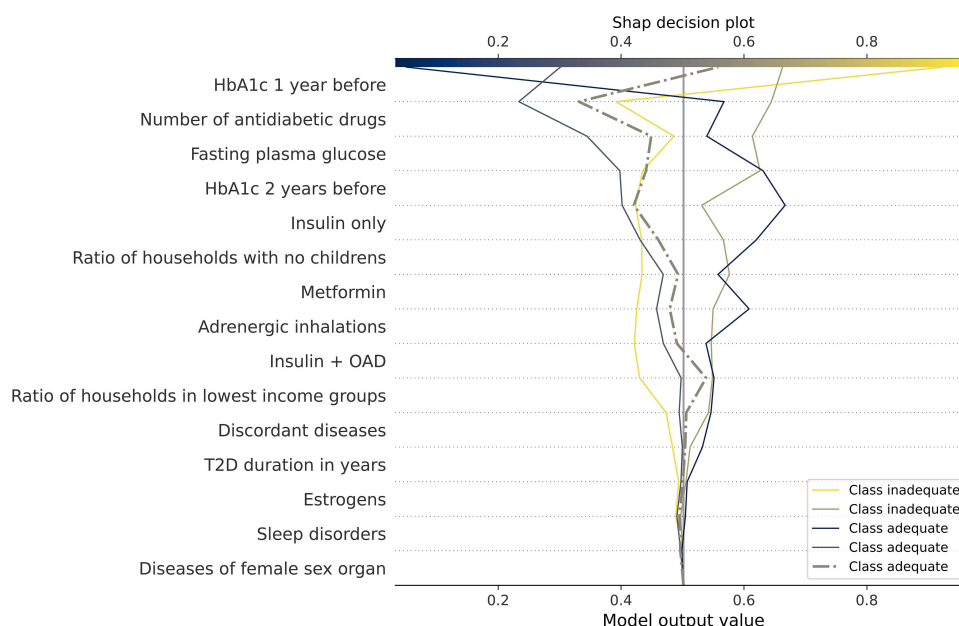
Figure 8 SHAP decision plot of five samples from one of the neural network models for clinical + treatment + socioeconomic predictors. Four correctly and one incorrectly predicted sample. Samples plotted with solid lines are correctly predicted samples, and the sample plotted with a dash-dot line is the incorrectly predicted sample. Legend of the plot marks the true classes. OAD, oral antidiabetic drugs or GLP-1 analogues (incl. metformin, sulfonylureas, combinations of oral blood glucose lowering drugs, glitazones, DPP-4 inhibitors, glinides, GLP-1 analogues, and SGLT2 inhibitors).

probability of inadequate glycemic control. Furthermore, the SHAP decision plot (Figure 8) illustrates the decision process of the NN model for the given five samples. The decision process for prediction starts at the base value, around 51% in the NN model for CTS predictors, and then SHAP computes the Shapley values to estimate the effect of each predictor. Once the base and Shapley values are estimated, the decision plot illustrates the effect of each predictor in the order of impact to arrive at the final prediction probability aka, the model output value (from bottom to top in Figure 8). Samples plotted with solid lines are correctly predicted samples, and the sample plotted with a dash-dot line is the incorrectly predicted sample. The incorrectly predicted sample was initially estimated to belong to the adequate class established by the GMM model; however, NN anticipated it to belong to the inadequate class with a probability of around 58%, while the base is at around 51%.

Discussion

This study captured variability in HbA1c by identifying three trajectories in heterogeneous T2D patients with similarities in HbA1c development over a 6-year follow-up. Prior glucose levels, time since T2D diagnosis, use of insulin only, use of insulin with some oral antidiabetic medication, use of only metformin, and the number of antidiabetic medications were the most important predictors for estimating the membership of the glycemic control class. We used binary classification to predict the trajectory membership to distinguish those with optimal, adequate treatment balance (86.5%) from those with inadequate treatment balance (13.5%). Overall, the prediction model was of high performance. In addition, the study showed that by using SHAP values, we could explain the reasons behind the predictions at both population and individual levels.

   As expected, the extracted three trajectories resemble those identified in previous studies.[32–38] Typically, two to six HbA1c trajectories have been identified, of which the biggest one has been similar to the "stable, adequate" observed in our study, with sizes varying between 44.7% and 88.7%. However, studies have differences in the study designs, follow-up times and patient groups; therefore, comparing the study results is complicated. In a Dutch study conducted among newly diagnosed T2D patients, three HbA1c trajectories were identified, and trajectory membership was predicted with machine learning methods.[39] A "stable, adequate" (76.5%) trajectory similar to ours was extracted in that study. The second identified trajectory was "improved glycemic control" (21.3%), which had the same shape as our "improving,

but inadequate" trajectory but with lower HbA1c values. The third trajectory, "deteriorated glycemic control" (2.2%), was also of the same shape as the "improving, but inadequate" of ours, but with higher HbA1c values. BMI, HbA1c, and triglyceride levels were the most important predictors of trajectory membership found in that study. Moreover, when we used our data to test the trained model of the Dutch study, we found that the model failed to predict the correct association between glycemic control classes. It predicted only 33 samples to be adequate, while 4902 samples in our data belonged to the adequate class and predicted 4933 samples to belong in the inadequate class while our data contained only 298 samples in that class. The most significant difference between our study and theirs was that their cohort was more diverse than ours, being only Finnish citizens, and all patients were newly diagnosed. In contrast, our data represented one region in Finland, and patients were diagnosed with T2D in the range of 2 to 47 years consisting of a very heterogeneous group of T2D patients in different phases of the disease.

Both LDA and NN models performed exceptionally well in predicting the membership into the glycemic control classes. All the final models reached a balanced accuracy of 85% with increasing confidence as more types of predictors were added. Three different predictor types were defined and used for classification in cascading order. The first type, clinical, was used as a baseline model to see how well the model predicted the association and later, more predictors of different types, such as treatment and SES factors, were added to potentially improve the prediction performance, increase the confidence, and lower the misclassification cost. Data imbalance in the class was significant and affected the model's performance. In our other experiments, when the model performed very well in the inadequate class, the misclassification for the adequate class increased quickly and vice versa. The trade-off between the two classes was very sensitive, and these models were optimized to find a perfect balance. It is reasonable to say that the misclassification cost of the inadequate class is higher than that of the adequate class, as the patients with an increased risk of deteriorating health are more important to be identified. The insufficiency of the data for deteriorating and improving trajectories also led to opting for binary classification instead of multi-class classification, similar to the related previous study.[39] Many experiments were performed; however, none of the models succeeded in distinguishing between deteriorating and improving trajectories; therefore, the trajectories were combined, and binary classification was performed against the adequate class.

Prior glucose levels, time since T2D diagnosis, use of antidiabetic drugs, and the number of antidiabetic drugs in use were the most important predictors for the treatment balance, which are also readily available for clinicians from the EHRs. Based on the SHAP summary plot, low blood glucose levels in previous years decreased, whereas the longer duration of T2D and use of insulin only increased the probability of belonging to the inadequate class. In addition, the probability of inadequate glycemic control decreased along with an increasing number of antidiabetic drugs. It is reasonable as antidiabetic medications are add-on treatments – treatments are intensified by adding treatments to the existing ones, which should also benefit HbA1c.[9–12] However, even though the target is to intensify the treatment following the HbA1c levels and add or change medication when needed, the use of drugs is not always optimal.[40] For the SHAP decision plot, three types of samples, two from each class and one incorrectly predicted sample, were selected to show the take of different predictors for diverse samples in estimating the model prediction. It is important to note that the interpretation of the SHAP plot cannot be read as changing one feature value will change the prediction probability directly; instead, the interaction of the features needs to be considered as features are either linearly or non-linearly correlated with each other. The SHAP plots assume the variables to be independent; however, in real-world data, especially in healthcare data, the assumption of absolute independence of variables is not adequate—this forces us to keep in mind that the findings from the SHAP plots should not be followed blindly. Furthermore, when the SHAP decision plot is plotted for multiple samples, the average impact of features is considered, meaning that the influence of each feature is individual to per sample's prediction, hence the order of features in plot changes as more samples are added. The SHAP decision plot for each selected sample individually with feature values can be found in Supplementary Figure 2.

Strengths of this study include the inclusion of all T2D patients from the North Karelia region, which allowed us to avoid selection bias, whereas using EHRs as a data source allowed us to avoid recall bias. In the study area, HbA1c and cholesterol, for example, are tested in the same regional laboratory with the same standardized methods. Generally, the Finnish health care system enables continuous monitoring of patients from the EHRs until death or moving out of the study area, with some exceptions in certain areas.

Our study has some limitations. First, we did not have external data to validate our study results. Therefore, the generalizability of the prediction model results and efficacy remain unknown. Although cross-validation was used in building the prediction models, it is essential to note that the prediction performance is susceptible to the data samples; the performance varies depending on which samples fall in the training or testing set. As more predictors were added, the models became more confident, yet variation was still present as expected for all real-world healthcare data. Second, a notable proportion of missing data on laboratory assays limited the use of these variables as predictors of trajectory membership. However, this is almost unavoidable when utilizing real-world data from EHRs. Third, the FPR does not contain data on medication use during hospital stays of over 90 days (about 3 months), and over-the-counter medication purchases are not registered in the FPR. In addition, even though we had FPR data on medication purchases, medication adherence remained unknown. Fourth, the two trajectories with inadequate treatment balance were combined for prediction purposes into one class as they were challenging to predict separately. However, as the trajectories had similarities in the fluctuation of HbA1c above good treatment balance and anti-diabetic medication use throughout the follow-up, combining the trajectories was justified. Fifth, we had only small-area level data on SES factors. For future work, personalized SES factors of the patient could help the prediction model to be more accurate and personalized. Sixth, we did not have information on specialist diabetes care and hospital admissions, which could have been important predictors of inadequate treatment balance.

The findings of this study are encouraging as patients with continuously inadequate treatment balance can be predicted with confidence based on patients' characteristics at any point of the disease course. Glycemic variability was captured by extracting three HbA1c trajectories, which were further classified into patients with stable, adequate glycemic control (86.5%) and those with inadequate, suboptimal glucose levels over six years (13.5%). The NN model with CTS predictors gave correct predictions for 82.5% of the patients in the adequate class and 86.6% in the inadequate class. In previous studies, HbA1c trajectories have been observed to be associated with micro- and macrovascular complications.[34,37] Identification of the clusters of patients with challenges in T2D care provides a tool that can support clinicians in further treatment planning, which also offers a possibility for optimizing the use of health care resources. More intense care could be allocated to patients with challenges in their treatment balance, whereas those with adequate, optimal HbA1c values could have less frequent monitoring of their HbA1c levels.

Conclusion

This study identified three HbA1c trajectories of which binary classification of adequate or inadequate long-term glycemic control was formed for prediction purposes. The extracted trajectories were in line with previous publications with most patients having stable, adequate long-term glycemia over 6-year follow-up. Inadequate long-term glycemic control could be predicted with confidence utilizing information from previous HbA1c levels and fasting plasma glucose, duration of T2D, and use of antidiabetic medications. Our findings suggest that heterogeneity in long-term treatment outcomes is predictable with patient's unique risk factors. This, in turn, offers a useful tool to support treatment planning in the future. However, future studies are needed to obtain even more accurate and personalized predictions.

Abbreviations

ATC, Anatomical Therapeutic Chemical; AUC, area under the curve; BIC, Bayesian Information Criteria; BMI, body mass index; CI, confidence interval; CTS, clinical, treatment and socio-economic status; EHR, electronic health records; FPR, Finnish Prescription Register; GMM, growth mixture model; HbA1c, glycated hemoglobin A1c; ICD, International Classification of Diseases; IFCC, International Federation of Clinical Chemistry; LDA, linear discriminant analysis; NN, neural network; OAD, oral antidiabetic drugs or GLP-1 analogues (incl. metformin, sulfonylureas, combinations of oral blood glucose lowering drugs, glitazones, DPP-4 inhibitors, glinides, GLP-1 analogues, and SGLT2 inhibitors); ROC, receiver operating characteristic; SES, socio-economic status; SHAP, Shapley additive explanations; SII, Social Insurance Institution of Finland; SRR, Special Reimbursement Register; T2D, type 2 diabetes; TINIA, turbidimetric inhibition immunoanalytical method.

Data Sharing Statement

Access to data is regulated by the European Union and Finnish laws, and therefore, sharing of sensitive data is not possible and data are not publicly available. An anonymized version of the data is available for researchers who meet the criteria as required by the European Union and Finnish laws for access to confidential data with a data permit of an appropriate authority. Contact information: aineistoneuvonta@siunsote.fi for data requests from the Siun sote – Joint municipal authority for North Karelia social and health services and tietoaineistot@kela.fi for data requests from the Social Insurance Institute.

Ethics Approval and Informed Consent

Use of the data was approved by the Ethics Committee of the Northern Savonia Hospital District (diary number 81/2012). The study protocol was also approved by the register administrator, the Siun sote. A separate permission to link data on medication purchases and special reimbursements was achieved from the SII (diary number 110/522/2018). Only register-based data were utilized and, in accordance with Finnish legislation, consent from the patients was not needed. The study complies with the Declaration of Helsinki.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

Disclosure

JM is a founding partner of ESiOR Oy. This company was not involved in carrying out this research. PL, GC, PS, ST, AI, JR, and TL declare no conflicts of interest.

References

1. International Diabetes Federation. IDF diabetes atlas. International Diabetes Federation; 2021. Availabe from: www.diabetesatlas.org. Accessed December 20, 2022.
2. American Diabetes Association. Economic cost of diabetes in the U.S. in 2017. *Diabetes Care*. 2018;41:917–928. doi:10.2337/dci18-0007
3. Colayco DC, Niu F, McCombs JS, Cheetham TC. A1C and cardiovascular outcomes in type 2 diabetes: a nested case-control study. *Diabetes Care*. 2011;34(1):77–83. doi:10.2337/dc10-1318
4. Zoungas S, Patel A, Chalmers J, et al. Severe hypoglycemia and risks of vascular events and death. *N Engl J Med*. 2010;363(15):1410–1418. doi:10.1056/NEJMoa1003795
5. Selvin E, Marinopoulos S, Berkenblit G, et al. Meta-analysis: glycosylated hemoglobin and cardiovascular disease in diabetes mellitus. *Ann Intern Med*. 2004;141:421–431. doi:10.7326/0003-4819-141-6-200409210-00007
6. Stratton IM, Adler AI, Neil HA, et al. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. *BMJ*. 2000;321:405–412. doi:10.1136/bmj.321.7258.405
7. UKPDS Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet*. 1998;352:837–853. doi:10.1016/S0140-6736(98)07019-6
8. Genuth S, Clinical I-BF. Implications of the ACCORD Trial. *J Clin Endocrinol Metab*. 2012;97(1):41–48. doi:10.1210/jc.2011-1679
9. Type 2 diabetes. Current care guidelines. working group set up by the Finnish medical society duodecim, the Finnish society of internal medicine and the medical advisory board of the Finnish diabetes society. Helsinki: The Finnish Medical Society Duodecim; 2020. Available from: https://www.kaypahoito.fi/hoi50056#K1. Accessed April 22, 2022.
10. Arnett DK, Blumenthal RS, Albert MA, et al. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: a report of the American college of cardiology/American Heart Association Task Force on clinical practice guidelines. *Circulation*. 2019;140:e596–e646. doi:10.1161/CIR.0000000000000678

11. Cosentino F, Grant PJ, Aboyans V, et al. ESC Guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD. *Eur Heart J.* 2019;2020(41):255–323. doi:10.1093/eurheartj/ehz486

12. Buse JB, Wexler DJ, Tsapas A, et al. update to: management of hyperglycaemia in type 2 diabetes, 2018. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetologia.* 2019;2020(63):221–228. doi:10.1007/s00125-019-05039-w

13. Ren X, Wang Z, Guo C. Long-term glycemic variability and risk of stroke in patients with diabetes: a meta-analysis. *Diabetol Metab Syndr.* 2022;14(1):6. doi:10.1186/s13098-021-00770-0

14. Cavalot F. Do data in the literature indicate that glycaemic variability is a clinical problem? Glycaemic variability and vascular complications of diabetes. *Diabetes Obes Metab.* 2013;15(Suppl. 2):3–8. doi:10.1111/dom.12140

15. Su G, Mi S, Tao H, et al. Association of glycemic variability and the presence and severity of coronary artery disease in patients with type 2 diabetes. *Cardiovasc Diabetol.* 2011;10:19. doi:10.1186/1475-2840-10-19

16. Wikström K, Lamidi M-L, Rautiainen P, Tirkkonen H, Kivinen P, Laatikainen T. The effect of the integration of health services on health care usage among patients with type 2 diabetes in North Karelia, Finland. *BMC Health Serv Res.* 2021;21(1):65. doi:10.1186/s12913-021-06059-2

17. WHO Collaborating Centre for Drug Statistics Methodology. ATC classification index with DDDs, 2019. Oslo, Norway; 2018. Available from: http://www.whocc.no/atc_ddd_index/. Accessed November 19, 2021.

18. Pikkujämsä S. Diabeteslaaturekisteripilotti – minimitietosisältöön tarvittava tietopohja. Rekisterin käynnistysvaihe 10.12.2019. (Minimum data content of the Finnish diabetes quality register). National Institute for Health and Welfare (THL). Available from: https://thl.fi/documents/2616650/4353715/12_2019_Minimitietosisa%CC%88lto%CC%88_alkuvaihe_lopullinen.pdf/8a9d733a-5019-4930-638f-7f374883bb01?t=1580828269938. Accessed June 22, 2022.

19. Statistics Finland. Paavo (open data by postal code area). Available from: https://pxnet2.stat.fi/PXWeb/pxweb/en/Postinumeroalueittainen_avoin_tieto/. Accessed June 22, 2022.

20. Toivakka M, Pihlapuro A, Tykkyläinen M, Mehtätalo L, Laatikainen T. The usefulness of small-area-based socioeconomic characteristics in assessing the treatment outcomes of type 2 diabetes patients: a register-based mixed-effect study. *BMC Public Health.* 2018;18(1):1258. doi:10.1186/s12889-018-6165-3

21. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics.* 1982;38:963–974. doi:10.2307/2529876

22. Muthén B, Shedden K. Finite mixture modelling with mixture outcomes using the EM algorithm. *Biometrics.* 1999;55:463–469. doi:10.1111/j.0006-341X.1999.00463.x

23. Jung T, Wickrama KAS. An introduction to latent class growth analysis and growth mixture modeling. *Soc Personal Psychol Compass.* 2008;2/1:302–317. doi:10.1111/j.1751-9004.2007.00054.x

24. Nagin DS, Odgers CL. Group-based trajectory modelling in clinical research. *Annu Rev Clin Psychol.* 2010;6:109–138. doi:10.1146/annurev.clinpsy.121208.131413

25. Nguena Nguefack HL, Pagé MG, Katz J, et al. Trajectory modelling techniques useful to epidemiological research: a comparative narrative review of approaches. *Clin Epidemiol.* 2020;12:1205–1222. doi:10.2147/CLEP.S265287

26. Muthén LK, Muthén BO. *Mplus User's Guide.* 8th ed. Los Angeles, CA: Muthén & Muthén; 1998.

27. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res.* 2011;46(3):399–424. doi:10.1080/00273171.2011.568786

28. Chollet F. *Deep Learning with Python.* 2nd ed. Shelter Island, NY: Manning; 2021.

29. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res.* 2011;12:2825–2830.

30. Fisher RA. The use of multiple measurements in taxonomic problems. *Ann Eugen.* 1936;7:179–188. doi:10.1111/j.1469-1809.1936.tb02137.x

31. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst.* 2017;30:4768–4777.

32. Luo M, Tan CS, Lim WY, et al. Association of diabetes treatment with long-term glycemic patterns in patients with type 2 diabetes mellitus: a prospective cohort study. *Diabetes Metab Res Rev.* 2019;35:e3122. doi:10.1002/dmrr.3122

33. Karpati T, Leventer-Roberts M, Feldman B, et al. Patient clusters based on HbA1c trajectories: a step toward individualized medicine in type 2 diabetes. *PLoS One.* 2018;13:e0207096. doi:10.1371/journal.pone.0207096

34. Luo M, Lim WY, Tan CS, et al. Longitudinal trends in HbA1c and associations with comorbidity and all-cause mortality in Asian patients with type 2 diabetes: a cohort study. *Diabetes Res Clin Pract.* 2017;133:69–77. doi:10.1016/j.diabres.2017.08.013

35. Mast MR, Walraven I, Hoekstra T, et al. Effectiveness of insulin therapy in people with type 2 diabetes in the Hoorn diabetes care system. *Diabet Med.* 2016;33(6):794–802. doi:10.1111/dme.13110

36. Ravona-Springer R, Heymann A, Schmeidler J, et al. Trajectories in glycemic control over time are associated with cognitive performance in elderly subjects with type 2 diabetes. *PLoS One.* 2014;9:e97384. doi:10.1371/journal.pone.0097384

37. Chang HY, Wahlqvist ML, Liu WL, et al. Management trajectories in the type 2 diabetes Integrated Delivery System project in Taiwan: accounting for behavioral therapy, nutrition education and therapeutics. *Asia Pac J Clin Nutr.* 2014;23(4):592–606. PMID: 25516317. doi:10.6133/apjcn.2014.23.4.06

38. Wang CP, Hazuda HP. Better glycemic control is associated with maintenance of lower-extremity function over time in Mexican American and European American older adults with diabetes. *Diabetes Care.* 2011;34(2):268–273. doi:10.2337/dc10-1405

39. Hertroijs DFL, Elissen AMJ, Brouwers MCGJ, et al. A risk score including body mass index, glycated haemoglobin and triglycerides predicts future glycaemic control in people with type 2 diabetes. *Diabetes Obes Metab.* 2018;20(3):681–688. doi:10.1111/dom.13148

40. Wikström K, Lamidi M-L, Rautiainen P, Tirkkonen H, Laatikainen T. Type 2 diabetes medication and HbA1c levels in North Karelia Finland, 2013–2019. *Diabet Med.* 2022;39. doi:10.1111/dme.14866