Development of a Machine Learning Model to Identify Nonadherence in People With Type 2 Diabetes using Connected Insulin Pen Data

Master Thesis

Hst-24-st-10-10413

Participants:

Line Højer
Mads W. Nielsen
Maja R. Leensbak

Supervisors:

Thomas Kronborg Larsen
Jannie Toft Damsgaard Nørlev

AALBORG UNIVERSITY
STUDENT REPORT

Aalborg University
Biomedical Engineering

AALBORG UNIVERSITY
STUDENT REPORT

Department of Health Science and Technology
Biomedical Engineering and Informatics
Selma Lagerløfs vej 249
9260, Gistrup
https://www.hst.aau.dk

Project:
Master Thesis

Title:
Development of a Machine Learning Model to Identify Nonadherence in People With Type 2 Diabetes using Connected Insulin Pen Data

Project period:
1. February 2024 - 31. May 2024

Project group:
Hst-24-st-10-10413

Participants:
Line Højer
Mads W. Nielsen
Maja R. Leensbak

Supervisors:
Thomas Kronborg Larsen
Jannie Toft Damsgaard Nørlev

Total pages: 59

Abstract:

Background: Type 2 diabetes (T2D) is a growing healthcare issue, where adherence to anti-diabetic medicine is a prominent factor. Nonadherence causes inefficient treatment which affects glycemic control and increases the financial burden. Early detection of nonadherence is crucial for effective interventions and is seen as being more beneficial than developing new treatment methods. Hence, this study aimed to develop a machine learning model to identify nonadherence in people with T2D based on data from a connected insulin pen. **Methods:** Data from 331 people with T2D were extracted from the DiaMonT trial (NCT04981808) to develop eight supervised machine learning models. Features were selected based on sequential forward feature selection and all models were trained and validated using 5-fold cross-validation and Receiver Operator Characteristics Area Under Curve (ROC-AUC). All models were optimized using grid search, however, only the results from the best-performing model are presented. **Results:** 43 features were extracted and used in the feature selection. Random Forest (ROC-AUC: 0.749) was the best-performing model using time in range of continuous glucose monitor values, HbA1c, if patients were telemonitored, systolic blood pressure, health status, and insulin type (basal+bolus or basal) as features ranked by feature importance. **Conclusion:** This study provides a model that can identify insulin nonadherence in people with T2D. Furthermore, the findings indicate that telemonitored patients are more likely to be adherent.

AALBORG UNIVERSITY
STUDENT REPORT

Institut for Medicin og Sundhedsteknologi
Sundhedsteknologi
Selma Lagerløfs vej 249
9260, Gistrup
https://www.hst.aau.dk

Projekt:
Kandidatspeciale

Titel:
Udvikling af en maskinlæringsmodel til identificering af nonadhærens i personer med type 2 diabetes ved brug af data fra en insulin smartpen

Projektperiode:
1. Februar 2024 - 31. Maj 2024

Projektgruppe:
Hst-24-st-10-10413

Gruppemedlemmer:
Line Højer
Mads W. Nielsen
Maja R. Leensbak

Vejledere:
Thomas Kronborg Larsen
Jannie Toft Damsgaard Nørlev

Total antal sider: 59

Abstrakt:

Baggrund og formål: Type 2 diabetes (T2D) er et stigende sundhedsproblem, hvor adhærens til anti-diabetisk medicin er en fremtrædende faktor. Nonadhærens forårsager ineffektiv behandling, hvilket påvirker glykæmisk kontrol og øger den finansielle byrde. Ved at opdage nonadhærens tidligt er det muligt at implementere effektive interventioner, hvilket også ses som værende mere fordelagtigt end at udvikle nye behandlingsmuligheder. Formålet med dette studie er at udvikle en maskinlæringsmodel, der kan identificere nonadhærens hos personer med T2D baseret på data fra en insulin smartpen. **Metoder:** Data fra 331 personer med T2D blev udtrukket fra det kliniske forsøg DiaMonT (NCT04981808) til udvikling af otte superviserede maskinlæringsmodeller. Features blev valgt på baggrund af sequential forward feature selection, hvor modellerne blev trænet og valideret ved brug af 5-fold krydsvalidering og Receiver Operator Characteristics Area Under Curve (ROC-AUC). Alle modellerne blev optimeret ved brug af grid search, men kun modellen, der præsterer bedst, bliver præsenteret i dette studie. **Resultater:** 43 features blev udtrukket og brugt i feature selection. Random Forest (ROC-AUC: 0.749) var bedst til at klassificere nonadhærente personer med T2D ved at bruge tid med kontinuerlige glukoseværdier i normalområdet, HbA1C, om patienterne var telemonitorerede, systolisk blodtryk, generel sundhedsstatus og insulin type (basal+bolus eller basal) som features rangeret ud fra feature importance. **Konklusion:** Modellen, der er udarbejdet i dette studie, er en model, der kan identificere nonadhærente personer med T2D. Desuden indikerer resultaterne, at telemonitorerede patienter er mere tilbøjelige til at være adhærente.

Preface

This project is a Master's thesis in Biomedical Engineering and Informatics at Aalborg University, made by group 10413. The project period was from the 1st of February 2024 to the 31st of May 2024. A huge thanks to Thomas Kronborg Larsen and Jannie Toft Damsgaard Nørlev for supervision and feedback during the project period.

Reading guide

The following report includes a scientific article and associated worksheets. The worksheets are supplementary material and contain chapters concerning different parts of the project. Chapter 1 is the problem analysis, where the problems associated with type 2 diabetes and medication adherence are presented. This is explored with knowledge from the current literature described in Chapter 3. From this, the aim is specified in Chapter 2. The methods used to investigate the aim are presented in Chapter 4, whereas some methods are elaborated in detail with additional tests. Chapter 5 describes the results for all the machine learning models. Lastly, Chapter 6 presents a portfolio with reflections on time management and project planning, which are part of the learning objectives.

The Vancouver Referencing System presents the literature used in the project. All references can be found in the bibliography at the end of the report. To ensure structure in the report, figures and tables are referenced by type, chapter, and number, e.g. Table 1.4 refers to chapter one, the fourth table. Likewise, sections and subsections come in numerical order, based on which chapter they are in. Some words, such as technical terms, are presented in full form the first time mentioned, after which they are abbreviated in parenthesis. Afterwards, only the abbreviation is applied.

Line Højer
lhajer19@student.aau.dk

Maja Randa Leensbak
mleens19@student.aau.dk

Mads Weiss Nielsen
mwni19@student.aau.dk

Development of a Machine Learning Model to Identify Nonadherence in People With Type 2 Diabetes using Connected Insulin Pen Data

Line Højer[1], Mads Weiss Nielsen[1], and Maja Randa Leensbak[1]
[1]*Department of Health Science and Technology, Aalborg University, Selma Lagerløfs Vej 249, Gistrup 9260, Denmark*

***Abstract*—Background: Type 2 diabetes (T2D) is a growing healthcare issue, where adherence to anti-diabetic medicine is a prominent factor. Nonadherence causes inefficient treatment which affects glycemic control and increases the financial burden. Early detection of nonadherence is crucial for effective interventions and is seen as being more beneficial than developing new treatment methods. Hence, this study aimed to develop a machine learning model to identify nonadherence in people with T2D based on data from a connected insulin pen. Methods: Data from 331 people with T2D were extracted from the DiaMonT trial (NCT04981808) to develop eight supervised machine learning models. Features were selected based on sequential forward feature selection and all models were trained and validated using 5-fold cross-validation and Receiver Operator Characteristics Area Under Curve (ROC-AUC). All models were optimized using grid search, however, only the results from the best-performing model are presented. Results: 43 features were extracted and used in the feature selection. Random Forest (ROC-AUC: 0.749) was the best-performing model using time in range of continuous glucose monitor values, HbA1c, if patients were telemonitored, systolic blood pressure, health status, and insulin type (basal+bolus or basal) as features ranked by feature importance. Conclusion: This study provides a model that can identify insulin nonadherence in people with T2D. Furthermore, the findings indicate that telemonitored patients are more likely to be adherent.**

Keywords: insulin adherence, type 2 diabetes, machine learning, connected insulin pen, insulin dosage data

I. Introduction

The World Health Organization (WHO) estimated that 108 million people had diabetes in 1980. This increased to 422 million people in 2014, which corresponds to 8.5% of all adults living with diabetes worldwide [1]. Additionally, the International Diabetes Federation (IDF) estimated that 537 million people lived with diabetes in 2021. This number is expected to increase to 643 million in 2030. In 2021, IDF also estimated that 6.7 million people died from diabetes-related causes before the age of 79 [2]. Type 2 diabetes (T2D) is responsible for approximately 90% of all diabetes-related cases and is a rapidly increasing chronic disease that affects how well glucose is used in the human body. Insulin is a blood-regulating hormone produced by the pancreas and is needed for glucose to enter the cells. In T2D either the pancreas does not produce enough insulin or the body cannot use insulin effectively, which results in elevated blood glucose levels [3].

Sufficient management of T2D is important for the delay or prevention of diabetes-related complications that can arise due to continuous high blood glucose levels. T2D can be controlled through lifestyle modifications, though additional anti-diabetic medication is often needed. People respond differently to the anti-diabetic treatment, resulting in personalized treatment plans [3]. Even though every patient has a personal treatment plan, diabetes management is complex, and several studies have shown that adherence to anti-diabetic medication for people with T2D is not optimal [4], [5]. WHO has defined adherence as *"the extent to which a person's behavior – taking medication, following a diet, and/or executing lifestyle changes, corresponds with agreed recommendations from a health care provider"* [6]. One of the reasons for nonadherence may be that people with T2D often do not experience notable symptoms in the case of high blood glucose levels, which causes them to deviate from their treatment plan, making them nonadherent to their medication. Other known reasons are fear of hypoglycemia, fear of injection pain, forgetfulness, etc. [7]–[11].

Nonadherence has a reducing effect on treatment, resulting in an increase in the financial burden due to the need for additional medications and hospital visits [12]–[14]. Therefore, improving adherence to anti-diabetic medication is more beneficial than developing new treatment methods [11], [14], where early detection of nonadherence is crucial for effective interventions [15], [16]. To aid clinicians in being more proactive and conducting more personalized treatment plans, thereby increasing adherence, it is essential to identify factors that could distinguish adherent and nonadherent patients [15]–[17]. Several studies [4], [5], [7]–[53] found associated factors with medication nonadherence in people with T2D. Age was found to be a factor of importance in 17 studies [4], [8], [10], [13], [14], [23]–[26], [33]–[37], [39], [42], [48], where [8], [33], [36], [48] stated that a younger age is correlated with nonadherence in people with T2D. HbA1c was also discovered to be a factor of importance in seven studies [8], [13], [15], [25], [27], [30], [34]. Furthermore, studies found that the attitude and knowledge regarding diabetes [5], [9], [14], [18], [28], [38], [39], [49] and the level of education [4], [14], [31], [40] showed a positive correlation with adherence, whereas poor mental health [18], [22], [29], [38], [53] showed a negative correlation with adherence.

To the best of our knowledge, only six studies [12], [16], [17], [25], [28], [51] have used machine learning to identify and differentiate between nonadherent and adherent people with T2D with promising results. 300 different machine learning models were developed in the study by Wu et al. [25] to screen for the risk of nonadherence. The best model used nine variables and had a Receiver Operator Characteristics Area Under Curve (ROC-AUC) of 0.866. Similar results were found

in studies by Li et al. [16] and Chen et al. [17], who also aimed to identify patients at risk of nonadherence using machine learning. Li et al. [16] tested 1080 different models with a ROC-AUC of 0.8369 on the best-performing model. Chen et al. [17] used Extreme Gradient Boosting to develop a model that could predict nonadherence with a ROC-AUC of 0.771. These results indicate that machine learning models can be used for the early identification of nonadherent patients.

Although several studies have been conducted to investigate the factors associated with adherence, no current studies have used data from actual insulin dosage to measure adherence. In the current studies, the measure of adherence is often self-reported through questionnaires [4], [5], [7], [9]–[11], [13], [14], [18], [19], [22], [23], [28]–[32], [38]–[40], [43]–[46], [49], [53] or based on pharmacy claims [12], [16], [17], [24]–[27], [33]–[37], [42], [48], [50]. Potential errors in using self-reported measures to assess medication adherence are common limitations throughout the literature [5], [7], [9]–[11], [13], [16], [18], [19], [21], [22], [25], [28]–[31], [38]–[40], [44], [45], [49], [53]. Self-reported measures can introduce recall bias and social desirability bias. This potential error is especially plausible with sensitive questions where problems may be over- or underestimated. Bias introduces errors that distort the image of medication adherence, resulting in a need to define adherence based on objective measures to gain more precise knowledge. [10], [18], [39] Therefore, this study aims to develop a machine learning model based on data from a connected insulin pen to identify insulin nonadherence in people with T2D. The connected insulin pen measures a more precise administration of injected insulin, thereby reducing bias [31], [39]. This identification model may assist clinicians in identifying nonadherence earlier and helping people with T2D establish better glycemic control.

II. Methods

A. Data acquisition

Data used in this study was acquired from the 3-month open-label randomized controlled trial Diabetes teleMonitoring of patients in insulin Therapy (DiaMonT) trial (ClinicalTrials.gov Identifier: NCT04981808). The trial aims to investigate the effect of telemonitoring in people with T2D on insulin therapy and collect data for developing dose guidance algorithms and algorithms for predicting adverse events in people with T2D. 331 participants were randomly assigned to either the intervention group or the control group. The intervention group received telemonitoring, whereas the control group continued with their usual care. The intervention group received a continuous glucose monitor (CGM) (Dexcom G6) to monitor their interstitial glucose and a connected insulin pen (Novopen 6) to monitor their administered insulin. Likewise, the control group received a Novopen 6 and a Dexcom G6, although it was blinded. Interstitial glucose was acquired every five minutes and insulin administration was acquired at every injection. The prescribed insulin was only acquired at baseline or when a clinician modified the dosage. Furthermore, health status, additional medications, comorbidities, and basic information such as age, duration of diabetes, smoking, etc. were obtained through questionnaires [54].

B. Preprocessing of data

As there is a need for early detection of nonadherence to improve patient outcomes [15], [16], data from day one to day 21 in the DiaMonT trial was acquired. Each day was separated at 03:00, as it was the hourly period with the fewest injections. This was done to reduce the number of misplaced injections and assign the injections to the correct day. Features were extracted from baseline characteristics, questionnaires, laboratory data, and the CGM within the first week. In total, 43 features were extracted based on literature and knowledge

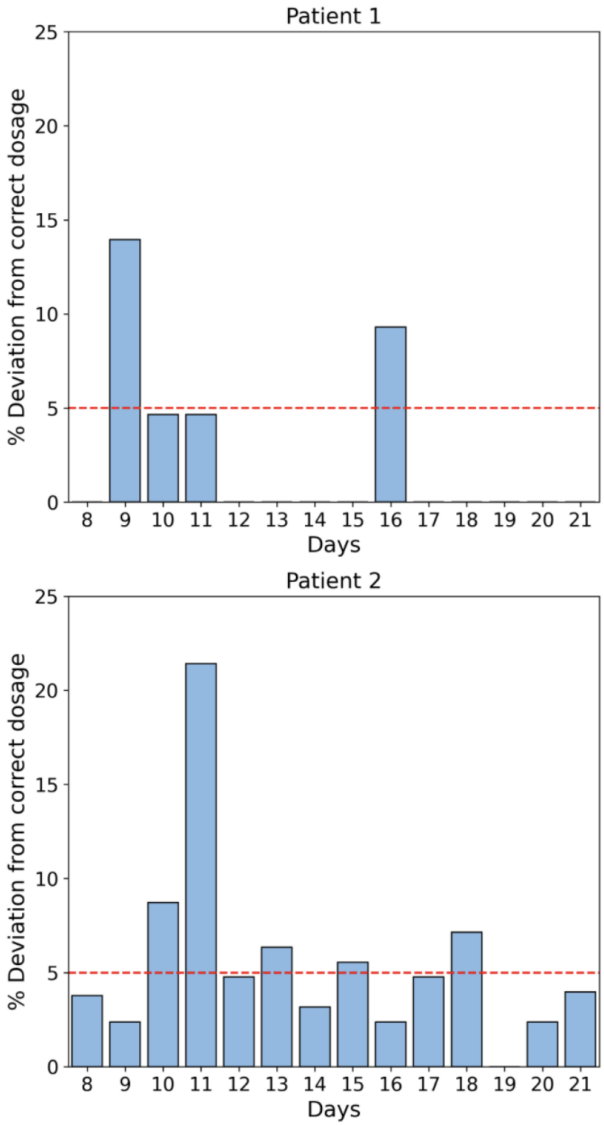


Fig. 1: Visualization of the percentage deviation from the correct dosage for two patients. *Patient 1* is adherent as the patient has <20% of nonadherent days above 5% deviation, whereas *Patient 2* is nonadherent as the patient has >20% of nonadherent days above 5% deviation. The 5% deviation line is represented in red.

regarding the topic [4], [5], [7]–[53]. A description of the selected features can be found in Appendix 1.

A study by Sokol et al. [55] found that people with diabetes with adherence levels of >80% had a significantly lower total cost of care and were less likely to be hospitalized. Therefore, this study defines adherence as a <20% deviation between injected insulin and prescribed insulin. Furthermore, the overall adherence level was calculated with inspiration from Nørlev et al. [56], where correct and incorrect dosages were identified daily. Adherence was defined based on data from days eight to 21, due to potential insecurities from new treatment methods in the first week. If the patients' injected insulin varied by more than ±5% from the prescribed insulin, they were defined as nonadherent on that specific day. The 5% daily deviation was allowed due to the subjective administration of the insulin, which might cause small errors. To get the overall adherence level of each patient during the 14 days, the sum of the nonadherent days was calculated. If the sum exceeded 20% of the 14 days, the patients were classified as nonadherent throughout this study. An example of an adherent patient and a nonadherent patient can be seen in Figure 1. *Patient 1* is adherent as <20% of the days exceed the 5% deviation line, whereas *Patient 2* is nonadherent as >20% of the days exceed the 5% deviation line. The adherence and nonadherence classifications for each patient were used as labels in the machine learning models.

C. Model development

In this study, machine learning was used to predict if patients were nonadherent or adherent to their insulin. Eight supervised classification machine learning models were developed and compared based on Random Forest, K-Nearest Neighbour, Logistic Regression, Support Vector Machine, Linear Discriminant Analysis, Extreme Gradient Boosting, Multi-Layer Perceptron, and Naïve Bayes. All data analyses and model developments were performed in Python version 3.10.12 using relevant libraries (Scikit-learn 1.3.1, Mlxtend 0.23.1, Xgboost 2.0.3, Pandas 2.1.1, Numpy 1.26.0, Matplotlib 3.8.0, and Shap 0.45.1).

The models were trained and validated using cross-validation. Cross-validation is a method that has been shown to reduce bias and performance variability [57] by making multiple random splits, resulting in different subsets. Therefore, the cross-validation score is an average of numerous performances from all available data. This study used a 5-fold cross-validation. Furthermore, sequential forward feature selection was implemented in the training process to reduce dimensions and improve performance. A tolerance of 0.001 in ROC-AUC was added to the feature selection to reduce the number of selected features and minimize potential overfitting. This means that feature selection stops when the performance does not increase by >0.1% when adding the next feature. After feature selection, selected hyperparameters of each machine learning model were optimized to improve performance further. The selected hyperparameters and their attempted values can be found in Appendix 2. The hyperparameters were optimized using a grid search where all possible combinations were tested. The combination of hyperparameters that yielded the highest cross-validation score was chosen.

D. Model evaluation

To quantify the performance of each machine learning model, the mean ROC-AUC across all cross-validation folds was used as a scoring parameter. This article will further evaluate the best-performing model by calculating positive predictive value (PPV), negative predictive value (NPV), and specificity at different sensitivity thresholds. This was done to quantify the performances and evaluate the clinical relevance. Furthermore, permutation feature importance was calculated on each feature to investigate the impact on performance, whereas SHapley Additive exPlanations (SHAP) values were calculated to analyze the tendencies of these features.

III. Results

Patients were excluded if they did not finish the trial or had missing data in one or more features, which included 279 patients in the analysis. 97 patients were labeled as nonadherent,

Model	ROC-AUC	Feature names
Logistic Regression	0.683	Time below range, Time in range, Height, Telemonitored, HbA1c, Sadness
Support Vector Machine	0.636	CGM mean, Height
Random Forest	0.749	Time in range, Systolic, Health status, Insulin type, Telemonitored, HbA1c
Linear Discriminant Analysis	0.692	Time below range, Time in range, Sum of other medications, Hyperlipidaemia, Insulin type, Telemonitored, HbA1c
K-Nearest Neighbour	0.686	CGM mean, Presence of hypoglycemia, Time in range
Extreme Gradient Boosting	0.734	CGM min, Time below range, Time in range, Mean arterial pressure, Sum of diabetes-related complications, Minimum one diabetes complication
Multi-Layer Perceptron	0.684	CGM max, Number of hypoglycemic events, Minimum one diabetes complication, Telemonitored, HbA1c, Sadness
Naïve Bayes	0.707	CGM mean, Time below range, Time in range, Height, Insulin type, Telemonitored, Sadness

TABLE 1: Features that resulted in the highest cross-validation score from the feature selection. The chosen features are listed for all models.

whereas 182 patients were labeled as adherent, leading to a nonadherence ratio of ≈0.35. The features and characteristics of the total population, adherent, and nonadherent groups can be found in Appendix 1. A total of 43 features were extracted, with eight features being significantly different between the adherent and nonadherent groups.

Table 1 lists the ROC-AUC and selected features for each machine learning model. It can be seen that at least one CGM feature was chosen in all of the models. Based on the ROC-AUC score, Random Forest was best at classifying nonadherent patients, with an ROC-AUC of 0.749. Random Forest performed best using a maximum depth of 10, 100 trees, and a minimum of 1 sample in the leaf node. The ROC-AUC curve for Random Forest can be seen in Figure 2.

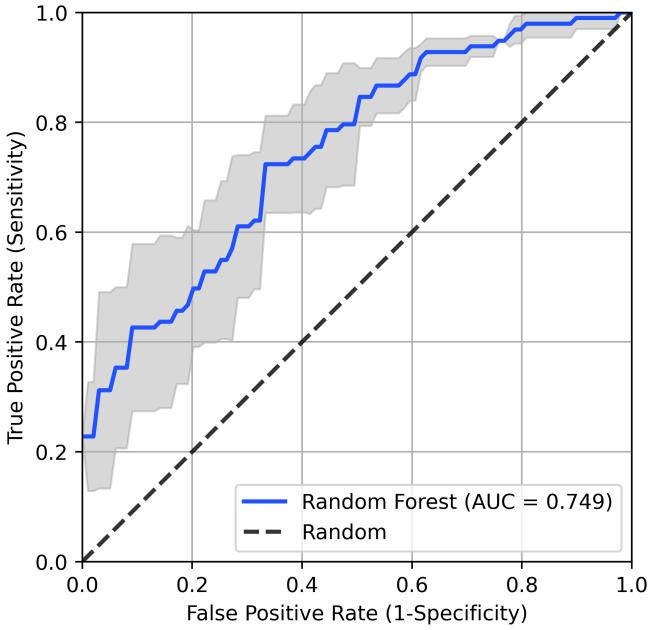


Fig. 2: The mean ROC-AUC of 5-fold cross-validation for Random Forest. The standard deviation is illustrated with a gray area.

Furthermore, Table 2 shows the calculated specificity, PPV, and NPV at different fixed sensitivity thresholds for Random Forest.

Sensitivity	Specificity	PPV	NPV
0.5	0.83	0.68	0.78
0.6	0.71	0.56	0.80
0.7	0.65	0.54	0.83
0.8	0.57	0.51	0.86
0.9	0.39	0.45	0.91

TABLE 2: Overview of specificity, PPV, and NPV at fixed sensitivities for Random Forest. If it was not possible to fix the sensitivity at the exact threshold, the closest point above the threshold was chosen.

Through feature selection, Random Forest selected *Time in range*, *HbA1c*, *Telemonitored*, *Systolic*, *Health status*, and *Insulin type* as the features that were best at classifying nonadherent patients with a tolerance of 0.001. The importance of these features is visualized in Figure 3.

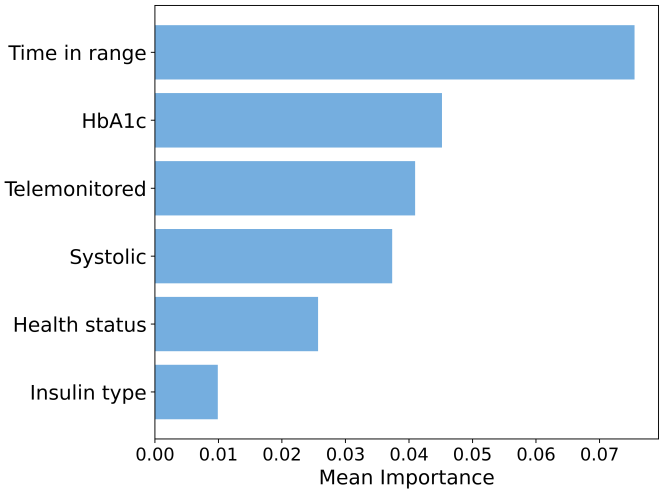


Fig. 3: Permutation feature importance for the features selected in the feature selection for Random Forest.

The SHAP values of the selected features can be seen in Figure 4. The SHAP values indicate that a nonadherent patient was more likely to have less time in the normal range with CGM values between 3.9 mmol/L and 10 mmol/L, a high HbA1c value, and a high systolic value. Furthermore, a nonadherent patient was more likely to use only basal insulin, have a poor health status, and not be telemonitored.

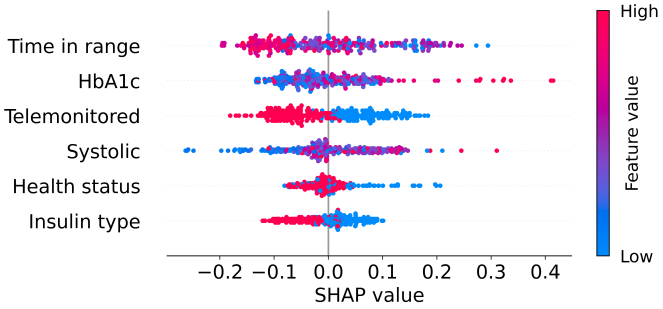


Fig. 4: SHAP values of the six features selected in the feature selection for Random Forest.

IV. Discussion

The purpose of this study was to develop a machine learning model to identify insulin nonadherence in people with T2D using data from a connected insulin pen. The study found that Random Forest was the best-performing model, with a ROC-AUC of 0.749 based on six features. This result indicates that the developed model was able to identify insulin nonadherence in people with T2D.

Random Forest has an embedded feature selection, which makes it possible to use all available features without overfitting the model. This means that a clinician should potentially measure and document data for 43 different features to identify

nonadherence using the developed model. However, as this is a time-consuming task, it may not be well implemented in clinical practice. An additional feature selection with a tolerance of 0.001 in the ROC-AUC was implemented to reduce the number of features and improve clinical relevance. However, the feature selection only investigated if the performance increased by adding one additional feature. This may have influenced the model's performance, as the tolerance implementation may have stopped the feature selection too early due to contradictory features. To avoid this potential problem, a moving window could be added to the feature selection to investigate whether the performance increased or if the current feature subset resulted in the best performance.

The acquired data consisted of data from a 3-month trial period. However, only 21 days of data were used in this study due to the desire for early detection of nonadherence in people with T2D. Further analysis showed that 81.4% of the people who were nonadherent in the first 21 days were also nonadherent during the entire 3-month trial. This indicates that 14 days of data can be a representative sample for predicting nonadherence.

In clinical practice, patients may either have a connected insulin pen and a CGM or not. If patients have these devices, adherence can be assessed immediately using the data already collected. However, if they do not have the devices initially, they would need to be provided before adherence assessment, which could introduce bias, as patients might change their behavior and become more adherent once they are aware that their data is being collected. This bias can also be found in this study, where the HbA1c values of each participant were investigated before and after the trial. This was done to see if the participants improved their glycemic control, indicating a behavior change. It was found that 78% of the participants had a lower HbA1c value at the end of the trial, of which 39.9% dropped by more than 10 mmol/mol. This indicates that the participants changed their behavior after entering the trial.

Before potential model implementation in clinical practice, threshold values for correct identification must be defined. To do so, the model's effectiveness was investigated using fixed sensitivities, PPV, and NPV. Uncovering the best threshold value would demand clinical involvement, as a higher NPV would result in a lower PPV and vice versa. Therefore, it is not straightforward to set a general threshold, as a clinician has to decide if it is more beneficial to identify too many patients where not all are nonadherent or to identify fewer people where all are nonadherent but not all nonadherent patients are identified. If a clinician prefers a sensitivity of 0.8, it results in $\approx$51% being true positives and $\approx$86% being true negatives. If this is applied to a population of 1000 people with a nonadherence prevalence of $\approx$35% it results in $\approx$179 of the nonadherent patients being identified, whereas 559 of the adherent people are identified. This means that $\approx$172 patients will receive a false negative answer, whereas 91 patients will receive a false positive answer. If the sensitivity is increased to 0.9, it results in $\approx$193 patients being false negatives and $\approx$59 patients being false positives. In both examples, more patients will wrongly be identified as nonadherent, resulting in more people coming to extra consultations. The consultations could involve additional screening, meetings with a dietitian, and education in diabetes management [4]. Being predicted as nonadherent is not a critical notice, and it would therefore be more beneficial to identify too many as nonadherent. This would potentially limit the overall financial burden in the future, as it lowers the chance of developing diabetes-related complications.

Studies by Chen et al. [17], Li et al. [16], and Wu et al. [25] have used machine learning to identify medication nonadherence in people with T2D. These studies yielded ROC-AUC values of 0.866, 0.837, and 0.771, which is better than the ROC-AUC in this study. However, the studies defined adherence based on self-reported measures or pharmacy claims, whereas this study defined adherence based on data from a connected insulin pen. Using self-reported measures or pharmacy claims may have introduced bias, as there is no guarantee that the patients speak the truth or have taken the prescribed medication. This means that the results from the recent studies may not give a clear view of medication nonadherence in people with T2D. However, using data from a connected insulin pen to identify medication nonadherence removes bias, as there is only a small chance of subjective error. This means that the results of this study provide a more exact identification of medication nonadherence than previous studies.

By using permutation feature importance and SHAP values, it was possible to investigate the performance of the selected features and find possible tendencies. *Time in range* is the feature with the highest importance, where low values have the highest impact on model performance, meaning that nonadherent patients were less likely to have CGM values between 3.9 mmol/L and 10 mmol/L. This corresponds with the fact that people with diabetes often fear hypoglycemia, which results in people not taking enough insulin, leading to higher glucose values [58]. This is supported in Appendix 1, where there are significantly higher values of *Time above range* than *Time below range*. *HbA1c* is the feature with the next highest feature importance, where high HbA1c values indicate nonadherence. This coheres with the literature [13], [15], [16], [25], [27], [30], [34] on the topic, which found that high HbA1c values indicate poor glycemic control and a low adherence level. *Health status* was also found to be a feature of high importance, where low general health indicates poor adherence. This corresponds with the study by Eby et al. [35], which found indicators that nonadherent patients have poorer general health. However, the study by Eby et al. found that nonadherent patients were likely to use both basal and bolus insulin. This does not correspond with the findings of this study, as nonadherent patients were more likely to use only basal insulin. A possible reason for this could be clinical inertia, where the clinician deviates from the treatment guidelines due to insufficient basal administration. Therefore, the clinician might not prescribe bolus insulin before the patient can manage basal insulin administration

correctly. The feature for systolic blood pressure was also found to have an impact, even though there are only minor noticeable differences between the adherent and nonadherent groups in Appendix 1. Similar results were found in the study by Vlacho et al. [27], which found only minor statistical differences in systolic blood pressure between the adherent and the nonadherent groups. However, they found a higher systolic value for the nonadherent group, which corresponds with the findings of this study. Lastly, the SHAP values indicate that nonadherent patients were more likely not to be telemonitored. Based on our knowledge, this study is the first to investigate telemonitoring as a feature of nonadherence. Due to the clear indication and importance, it might be beneficial to incorporate telemonitoring in clinical practice to increase adherence to insulin therapy.

Limitations

The methods presented in this study provide a machine learning model to identify nonadherence. However, it includes some limitations when applied to data that focuses on the effect of telemonitoring. Firstly, some patients may have changed their behavior when entering the trial, resulting in misleading information regarding medication nonadherence in people with T2D. This especially occurs in the intervention group, where patients were telemonitored throughout the trial. The model would have to be applied to supplementary data to determine the effectiveness of identifying nonadherent patients and providing detailed information. This supplementary data should originate from people with T2D in an everyday setting to find the common nonadherent patient. Secondly, the data were obtained from only one hospital in Aalborg, Denmark, which may influence the model's generalizability and transferability to other populations or countries. Therefore, future studies should include more diverse data from different populations.

V. Conclusion

Current methods for investigating medication nonadherence are prone to bias as they define adherence using self-reported measures or pharmacy claims. This study provides a model where adherence is defined based on data from a connected insulin pen. The findings indicate that the developed model can identify insulin nonadherence in people with T2D. The model finds *Time in range*, *HbA1c*, *Telemonitored*, *Systolic*, *Health status*, and *Insulin type* as risk factors for nonadherence. Furthermore, the study found that telemonitored patients are more likely to be adherent, indicating that telemonitoring might be beneficial as a tool to increase insulin adherence in people with T2D.

Acknowledgment

The authors would like to give a special thanks to the consortium of ADAPT-T2D for the possibility of using data from the DiaMonT trial.

Abbreviations

T2D, Type 2 Diabetes; IDF, International Diabetes Federation; WHO, World Health Organization; ROC-AUC, Receiver Operator Characteristics Area Under Curve; DiaMonT, Diabetes Telemonitoring of Patients in Insulin Therapy; CGM, Continuous Glucose Monitor; PPV, Positive Predictive Value; NPV, Negative Predictive Value; SHAP, SHapley Additive exPlanations.

References

1 World Health Organization, "Diabetes," *World Health Organization: WHO*, Apr. 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/diabetes

2 "IDF Diabetes Atlas," May 2024, [Online; accessed 17. May 2024]. [Online]. Available: https://diabetesatlas.org

3 American Diabetes Association, "Living With Diabetes | ADA," Apr. 2024, [Online; accessed 16. Apr. 2024]. [Online]. Available: https://diabetes.org/living-with-diabetes

4 Krishnan V., S. and V., R., "A cross sectional study on drug adherence among type 2 diabetes mellitus patients attending a tertiary care hospital in Chennai," *Indian Journal of Public Health Research and Development*, vol. 11, no. 3, pp. 964–969, Mar. 2020. [Online]. Available: https://medicopublication.com/index.php/ijphrd/article/view/1497/1380

5 Nazir, S. U. R., Hassali, M. A., Saleem, F., Bashir, S., and Aljadhey, H., "Association Between Diabetes-related Knowledge and Medication Adherence: Results From Cross-sectional Analysis," *Altern. Ther. Health Med.*, vol. 22, no. 6, pp. 8–13, Nov. 2016. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/27866175

6 Organization, W. H., *Adherence to long-term therapies : evidence for action*. World Health Organization, 2003. [Online]. Available: https://iris.who.int/handle/10665/42682

7 Azri, N., Norsa'adah, B., Hassan, N. B., and Naing, N. N., "Insulin Adherence and Associated Factors in Patients with Type 2 Diabetes Mellitus Treated in Klang Primary Health Care Centres," *Malays. J. Med. Sci.*, vol. 28, no. 6, pp. 76–87, Dec. 2021.

8 Lee, D. S. U. and Lee, H., "Adherence and persistence rates of major antidiabetic medications: a review," *Diabetology and Metabolic Syndrome*, vol. 14, no. 1, Jan. 2022. [Online]. Available: http://dx.doi.org/10.1186/s13098-022-00785-1

9 Yong, S. Y., Goh, G. M., and Loh, H. H., "Insulin adherence and the associated factors among patients with type 2 diabetes mellitus at the hospital queen elizabeth ii, sabah," *Journal of Public Health*, vol. 30, no. 5, p. 1319–1327, Nov. 2020. [Online]. Available: http://dx.doi.org/10.1007/s10389-020-01409-6

10 Nelson, L. A., Wallston, K. A., Kripalani, S., LeStourgeon, L. M., Williamson, S. E., and Mayberry, L. S., "Assessing barriers to diabetes medication adherence using the information-motivation-behavioral skills model," *Diabetes Research and Clinical Practice*, vol. 142, p. 374–384, Aug. 2018. [Online]. Available: http://dx.doi.org/10.1016/j.diabres.2018.05.046

11 Jimmy, B., Jose, J., Al-Hinai, Z. A., Wadair, I. K., and Al-Amri, G. H., "Adherence to Medications among Type 2 Diabetes Mellitus Patients in Three Districts of Al Dakhliyah Governorate, Oman: A cross-sectional pilot study," *Sultan Qaboos University Medical Journal*, vol. 14, no. 2, p. e231, May 2014. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3997541

12 Saundankar, V., Peng, X., Fu, H., Ascher-Svanum, H., Rodriguez, A., Ali, A., Slabaugh, L., Young, P., and Louder, A., "Predictors of change in adherence status from 1 year to the next among patients with type 2 diabetes mellitus on oral antidiabetes drugs," *Journal of Managed Care and Specialty Pharmacy*, vol. 22, no. 5, pp. 467–482, May 2016. [Online]. Available: http://dx.doi.org/10.18553/jmcp.2016.22.5.467

13 Lee, C. S., Tan, J. H. M., Sankari, U., Koh, Y. L. E., and Tan, N. C., "Assessing oral medication adherence among patients with type 2 diabetes mellitus treated with polytherapy in a developed asian community: a cross-sectional study," *BMJ Open*, vol. 7, no. 9, Sep. 2017. [Online]. Available: http://dx.doi.org/10.1136/bmjopen-2017-016317

14 Iqbal, Q., Bashir, S., Iqbal, J., Iftikhar, S., and Godman, B., "Assessment of medication adherence among type 2 diabetic patients in quetta city,

pakistan," *Postgraduate Medicine*, vol. 129, no. 6, p. 637–643, May 2017. [Online]. Available: http://dx.doi.org/10.1080/00325481.2017.1328251

15 McClintock, H. F., Edmonds, S. E., and Bogner, H. R., "Adherence patterns to oral hypoglycemic agents among primary care patients with type 2 diabetes," *Primary Care Diabetes*, vol. 17, no. 2, p. 180–184, Apr. 2023. [Online]. Available: http://dx.doi.org/10.1016/j.pcd.2023.01.014

16 Li, M., Lu, X., Yang, H., Yuan, R., Yang, Y., Tong, R., and Wu, X., "Development and assessment of novel machine learning models to predict medication non-adherence risks in type 2 diabetics," *Frontiers in Public Health*, vol. 10, Nov. 2022. [Online]. Available: http://dx.doi.org/10.3389/fpubh.2022.1000622

17 Chen, Y.-L., Nguyen, P.-A., Chien, C.-H., Hsu, M.-H., Liou, D.-M., and Yang, H.-C., "Machine learning-based prediction of medication refill adherence among first-time insulin users with type 2 diabetes," *Diabetes Research and Clinical Practice*, vol. 207, Jan. 2024. [Online]. Available: http://dx.doi.org/10.1016/j.diabres.2023.111033

18 Huang, J., Ding, S., Xiong, S., and Liu, Z., "Medication adherence and associated factors in patients with type 2 diabetes: A structural equation model," *Frontiers in Public Health*, vol. 9, Nov. 2021. [Online]. Available: http://dx.doi.org/10.3389/fpubh.2021.730845

19 Syafhan, N. F., Donnelly, R., Harper, R., Harding, J., Mulligan, C., Hogg, A., Scott, M., Fleming, G., Scullin, C., Hawwa, A. F., Chen, G., Parsons, C., and McElnay, J. C., "Adherence to metformin in adults with type 2 diabetes: a combined method approach," *Journal of Pharmaceutical Policy and Practice*, vol. 15, no. 1, Oct. 2022. [Online]. Available: http://dx.doi.org/10.1186/s40545-022-00457-5

20 Masaba, B. B. and Mmusi-Phetoe, R. M., "Determinants of non-adherence to treatment among patients with type 2 diabetes in kenya: A systematic review," *Journal of Multidisciplinary Healthcare*, vol. Volume 13, p. 2069–2076, Jan. 2021. [Online]. Available: http://dx.doi.org/10.2147/JMDH.S270137

21 Eze, U. I., Akhumi, T. F., Iheanacho, C. O., and Saka, S. A., "Drug therapy and medication adherence in type 2 diabetes in a care facility: A cross sectional survey," *Exploratory Research in Clinical and Social Pharmacy*, vol. 8, Dec. 2022. [Online]. Available: http://dx.doi.org/10.1016/j.rcsop.2022.100200

22 Jackson, I. L., Adibe, M. O., Okonta, M. J., and Ukwe, C. V., "Medication adherence in type 2 diabetes patients in nigeria," *Diabetes Technology and Therapeutics*, vol. 17, no. 6, p. 398–404, Jun. 2015. [Online]. Available: http://dx.doi.org/10.1089/dia.2014.0279

23 Wong, M. C., Wu, C. H., Wang, H. H., Li, H. W., Hui, E. M., Lam, A. T., Chung, R. Y., Yip, B. H., and Morisky, D. E., "Association between the 8-item morisky medication adherence scale (mmas-8) score and glycaemic control among chinese diabetes patients," *The Journal of Clinical Pharmacology*, vol. 55, no. 3, p. 279–287, Nov. 2014. [Online]. Available: http://dx.doi.org/10.1002/jcph.408

24 Horsburgh, S., Barson, D., Zeng, J., Sharples, K., and Parkin, L., "Adherence to metformin monotherapy in people with type 2 diabetes mellitus in new zealand," *Diabetes Research and Clinical Practice*, vol. 158, Dec. 2019. [Online]. Available: http://dx.doi.org/10.1016/j.diabres.2019.107902

25 Wu, X.-W., Yang, H.-B., Yuan, R., Long, E.-W., and Tong, R.-S., "Predictive models of medication non-adherence risks of patients with t2d based on multiple machine learning algorithms," *BMJ Open Diabetes Research and Care*, vol. 8, no. 1, Mar. 2020. [Online]. Available: http://dx.doi.org/10.1136/bmjdrc-2019-001055

26 Lee, D. S. U. and Lee, H., "Clinical characteristics associated with adherence and persistence in patients with type 2 diabetes mellitus treated with dulaglutide," *Journal of Diabetes Research*, vol. 2023, p. 1–12, Jun. 2023. [Online]. Available: http://dx.doi.org/10.1155/2023/7917641

27 Vlacho, B., Simarro, F. L., Mata-Cases, M., Miravet, S., Escribano-Serrano, J., Asensio, D., Cortes, X., and Franch-Nadal, J., "Adherence to antidiabetic treatment among patients managed in primary care centres in spain: the intense study," *Primary Care Diabetes*, vol. 16, no. 6, p. 760–767, Dec. 2022. [Online]. Available: http://dx.doi.org/10.1016/j.pcd.2022.10.004

28 Hashimoto, K., Urata, K., Yoshida, A., Horiuchi, R., Yamaaki, N., Yagi, K., and Arai, K., "The relationship between patients' perception of type 2 diabetes and medication adherence: a cross-sectional study in japan," *Journal of Pharmaceutical Health Care and Sciences*, vol. 5, no. 1, Jan. 2019. [Online]. Available: http://dx.doi.org/10.1186/s40780-019-0132-8

29 Parada, H., Horton, L. A., Cherrington, A., Ibarra, L., and Ayala, G. X., "Correlates of medication nonadherence among latinos with type 2 diabetes," *The Diabetes Educator*, vol. 38, no. 4, p. 552–561, Apr. 2012. [Online]. Available: http://dx.doi.org/10.1177/0145721712445215

30 Abdullah, N. F., Khuan, L., Theng, C. A., Sowtali, S. N., and Juni, M. H., "Effect of patient characteristics on medication adherence among patients with type 2 diabetes mellitus: a cross-sectional survey," *Contemporary Nurse*, vol. 55, no. 1, p. 27–37, Jan. 2019. [Online]. Available: http://dx.doi.org/10.1080/10376178.2019.1583067

31 Sánchez-Hernández, M. S., Rodríguez-Caldero, M. C., Martín-Pérez, M. P., Mira-Solves, J. J., Vitaller-Burillo, J., and Carratalá-Munuera, M. C., "Impact of adherence to mediterranean diet and/or drug treatment on glycaemic control in type 2 diabetes mellitus patients: Dm2-cumcyl study," *Primary Care Diabetes*, vol. 14, no. 6, p. 685–691, Dec. 2020. [Online]. Available: http://dx.doi.org/10.1016/j.pcd.2020.06.008

32 Nazir, S. U. R., Hassali, M. A., Saleem, F., Bashir, S., and Aljadhey, H., "Disease related knowledge, medication adherence and glycaemic control among patients with type 2 diabetes mellitus in pakistan," *Primary Care Diabetes*, vol. 10, no. 2, p. 136–141, Apr. 2016. [Online]. Available: http://dx.doi.org/10.1016/j.pcd.2015.09.004

33 Sun, P. and Lian, J., "Treatment adherence in newly diagnosed type 2 diabetes: patient characteristics and long-term impact of adherence on inpatient care utilization," *Postgraduate Medicine*, vol. 128, no. 4, p. 338–345, Feb. 2016. [Online]. Available: http://dx.doi.org/10.1080/00325481.2016.1151326

34 Buysman, E. K., Liu, F., Hammer, M., and Langer, J., "Impact of medication adherence and persistence on clinical and economic outcomes in patients with type 2 diabetes treated with liraglutide: A retrospective cohort study," *Advances in Therapy*, vol. 32, no. 4, p. 341–355, Apr. 2015. [Online]. Available: http://dx.doi.org/10.1007/s12325-015-0199-z

35 Eby, E. L., Bajpai, S., Faries, D. E., Haynes, V. S., and Lage, M. J., "The Association Between Adherence to Insulin Therapy and Health Care Costs for Adults with Type 2 Diabetes: Evidence from a U.S. Retrospective Claims Database," *Journal of Managed Care and Specialty Pharmacy*, vol. 26, no. 9, Sep. 2020.

36 Tunceli, K., Zhao, C., Davies, M. J., Brodovicz, K. G., Alexander, C. M., Iglay, K., and Radican, L., "Factors associated with adherence to oral antihyperglycemic monotherapy in patients with type 2 diabetes," *Patient Preference and Adherence*, Jan. 2015. [Online]. Available: http://dx.doi.org/10.2147/PPA.S71346

37 Vervloet, M., Spreeuwenberg, P., Bouvy, M. L., Heerdink, E. R., de Bakker, D. H., and van Dijk, L., "Lazy sunday afternoons: the negative impact of interruptions in patients' daily routine on adherence to oral antidiabetic medication. a multilevel analysis of electronic monitoring data," *European Journal of Clinical Pharmacology*, vol. 69, no. 8, p. 1599–1606, Apr. 2013. [Online]. Available: http://dx.doi.org/10.1007/s00228-013-1511-y

38 Kretchy, I. A., Koduah, A., Ohene-Agyei, T., Boima, V., and Appiah, B., "The association between diabetes-related distress and medication adherence in adult patients with type 2 diabetes mellitus: A cross-sectional study," *Journal of Diabetes Research*, vol. 2020, p. 1–10, Mar. 2020. [Online]. Available: http://dx.doi.org/10.1155/2020/4760624

39 Kumar, P. S., Ranjan, S. S., Pratyush, M., and Snehashini, D., "Assessment of medication adherence and its predictors in Type 2 diabetes mellitus patients in tertiary care teaching hospital: A cross-sectional observational study," *Journal of Cardiovascular Disease Research*, vol. 14, no. 12, p. 2023, Dec. 2023. [Online]. Available: https://www.researchgate.net/publication/376349519_Assessment_of_medication_adherence_and_its_predictors_in_Type_2_diabetes_mellitus_patients_in_tertiary_care_teaching_hospital_A_cross-sectional_observational_study

40 Wang, F.-C., Chang, W., Nie, S.-L., Shen, B.-X., He, C.-Y., Zhao, W.-C., Liu, X.-Y., and Lu, J.-T., "Predicting medication nonadherence risk in the chinese type 2 diabetes mellitus population – establishment of a new risk nomogram model: a retrospective study," *Journal of International Medical Research*, vol. 49, no. 9, Sep. 2021. [Online]. Available: http://dx.doi.org/10.1177/03000605211042502

41 Sapkota, S., Brien, J.-A. E., Greenfield, J. R., and Aslani, P., "A Systematic Review of Interventions Addressing Adherence to Anti-Diabetic Medications in Patients with Type 2 Diabetes–Components of Interventions," *PLoS One*, vol. 10, no. 6, Jun. 2015.

42 Egede, L. E., Gebregziabher, M., Lynch, C. P., and Echols, C., "Longitudinal Effects of Medication Nonadherence on Glycemic Control," *Annals of Pharmacotherapy*, vol. 48, pp. 562–570, Feb. 2014.

43 Kang, Y. and Hur, Y., "Medication adherence and its associated factors in laotians with type 2 diabetes mellitus," *Clinical Nursing*

Research, vol. 29, no. 5, p. 331–338, May 2019. [Online]. Available: http://dx.doi.org/10.1177/1054773819849349

44 Mishra, A., Pradhan, S. K., Sahoo, B. K., Das, A., Singh, A. K., and Parida, S. P., "Assessment of medication adherence and associated factors among patients with diabetes attending a non-communicable disease clinic in a community health centre in eastern india," *Cureus*, Aug. 2023. [Online]. Available: http://dx.doi.org/10.7759/cureus.43779

45 Shiyanbola, O. O., Unni, E., Huang, Y.-M., and Lanier, C., "Using the extended self-regulatory model to characterise diabetes medication adherence: a cross-sectional study," *BMJ Open*, vol. 8, no. 11, Nov. 2018. [Online]. Available: http://dx.doi.org/10.1136/bmjopen-2018-022803

46 Zongo, A., Guénette, L., Moisan, J., and Grégoire, J.-P., "Predictive validity of self-reported measures of adherence to noninsulin antidiabetes medication against control of glycated hemoglobin levels," *Canadian Journal of Diabetes*, vol. 40, no. 1, p. 58–65, Feb. 2016. [Online]. Available: http://dx.doi.org/10.1016/j.jcjd.2015.06.008

47 Simon-Tuval, T., Shmueli, A., and Harman-Boehm, I., "Adherence of patients with type 2 diabetes mellitus to medications: the role of risk preferences," *Current Medical Research and Opinion*, vol. 34, no. 2, p. 345–351, Nov. 2017. [Online]. Available: http://dx.doi.org/10.1080/03007995.2017.1397506

48 Campbell, D. J. T., Campbell, D. B., Ogundeji, Y., Au, F., Beall, R., Ronksley, P. E., Quinn, A. E., Manns, B. J., Hemmelgarn, B. R., Tonelli, M., and Spackman, E., "First-line pharmacotherapy for incident type 2 diabetes: Prescription patterns, adherence and associated costs," *Diabetic Medicine*, vol. 38, no. 9, Jun. 2021. [Online]. Available: http://dx.doi.org/10.1111/dme.14622

49 Ranjbaran, S., Shojaeizadeh, D., Dehdari, T., Yaseri, M., and Shakibazadeh, E., "Determinants of medication adherence among iranian patients with type 2 diabetes: An application of health action process approach," *Heliyon*, vol. 6, no. 7, Jul. 2020. [Online]. Available: http://dx.doi.org/10.1016/j.heliyon.2020.e04442

50 Gatwood, J. D., Chisholm-Burns, M., Davis, R., Thomas, F., Potukuchi, P., Hung, A., and Kovesdy, C. P., "Differences in health outcomes associated with initial adherence to oral antidiabetes medications among veterans with uncomplicated type 2 diabetes: a 5-year survival analysis," *Diabetic Medicine*, vol. 35, no. 11, p. 1571–1579, Jul. 2018. [Online]. Available: http://dx.doi.org/10.1111/dme.13775

51 Thyde, D. N., Mohebbi, A., Bengtsson, H., Jensen, M. L., and Mørup, M., "Machine learning-based adherence detection of type 2 diabetes patients on once-daily basal insulin injections," *Journal of Diabetes Science and Technology*, vol. 15, no. 1, p. 98–108, Apr. 2020. [Online]. Available: http://dx.doi.org/10.1177/1932296820912411

52 Wulandari, N., Maifitrianti, M., Hasanah, F., Atika, S., and Dini Putri, R., "Medication adherence assessment among patients with type 2 diabetes mellitus treated polytherapy in indonesian community health center: A cross sectional-study," *Journal of Pharmacy And Bioallied Sciences*, vol. 12, no. 6, 2020. [Online]. Available: http://dx.doi.org/10.4103/jpbs.JPBS_257_19

53 Nazir, S. U. R., Hassali, M. A., Saleem, F., Bashir, S., and Aljadhey, H., "Does adherence to the therapeutic regimen associate with health related quality of life: Findings from an observational study of type 2 diabetes mellitus patients in Pakistan," *Pakistan Journal of Pharmaceutical Sciences*, vol. 30, no. 6, pp. 2159–2166, Nov. 2017. [Online]. Available: https://go.gale.com/ps/i.do?id=GALE%7CA520323093&sid=googleScholar&v=2.1%E2%81%A2=r&linkaccess=abs&issn=1011601X&p=AONE&sw=w&userGroupName=anon%7E3841bd79&aty=open-web-entry

54 Hangaard, S., Kronborg, T., Hejlesen, O., Aradóttir, T. B., Kaas, A., Bengtsson, H., Vestergaard, P., and Jensen, M. H., "The diabetes telemonitoring of patients in insulin therapy (diamont) trial: study protocol for a randomized controlled trial," *Trials*, vol. 23, no. 985, Dec. 2022.

55 Sokol, M. C., McGuigan, K. A., Verbrugge, R. R., and Epstein, R. S., "Impact of Medication Adherence on Hospitalization Risk and Healthcare Cost," *Med. Care*, vol. 43, no. 6, p. 521, Jun. 2005.

56 Nørlev, J. T. D., Kronborg, T., Jensen, M. H., Vestergaard, P., Hejlesen, O., and Hangaard, S., "A three-step data-driven methodology to assess adherence to basal insulin therapy in patients with insulin-treated type 2 diabetes," *Journal of Diabetes Science and Technology*, Dec. 2023. [Online]. Available: http://dx.doi.org/10.1177/19322968231222007

57 Collins, G. S., Dhiman, P., Ma, J., Schlussel, M. M., Archer, L., Van Calster, B., Harrell, F. E., Martin, G. P., Moons, K. G. M., van Smeden, M., Sperrin, M., Bullock, G. S., and Riley, R. D., "Evaluation of clinical prediction models (part 1): from development to external validation," *BMJ*, vol. 384, p. e074819, Jan. 2024.

58 Huang, J., Ding, S., Xiong, S., and Liu, Z., "The Mediating Effects of Diabetes Distress, Anxiety, and Cognitive Fusion on the Association Between Neuroticism and Fear of Hypoglycemia in Patients With Type 2 Diabetes," *Front. Psychol.*, vol. 12, 2021.

Feature	Total (n = 279)	Adherent (n = 182)	Nonadherent (n = 97)	p-value
Demographic data				
Age, years	57.20 ± 11.26	57.25 ± 11.13	57.11 ± 11.54	0.755
Height, cm	173.69 ± 8.97	172.65 ± 9.24	175.64 ± 8.13	0.008†*
Weight, kg	100.10 ± 21.18	99.18 ± 20.83	101.84 ± 21.84	0.400
BMI	33.19 ± 6.60	33.30 ± 6.61	32.98 ± 6.60	0.673
Living alone, n	78 (27.96%)	48 (26.37%)	30 (30.93%)	0.421
Handyman, n	92 (32.97%)	56 (30.77%)	36 (37.11%)	0.284
Primary school, n	35 (12.54%)	25 (13.74%)	10 (10.31%)	0.412
Highschool, n	20 (7.17%)	11 (6.04%)	9 (9.28%)	0.319
Medium education, n	112 (40.14%)	76 (41.76%)	36 (37.11%)	0.449
Long education, n	20 (7.17%)	14 (7.69%)	6 (6.19%)	0.644
CGM data				
CGM variance, mmol/L	7.25 ± 4.42	6.82 ± 4.18	8.03 ± 4.74	0.022*
CGM mean, mmol/L	10.04 ± 2.49	9.69 ± 2.11	10.69 ± 2.98	0.010*
CGM max, mmol/L	17.98 ± 3.20	17.63 ± 3.12	18.63 ± 3.27	0.009*
CGM min, mmol/L	4.64 ± 1.73	4.51 ± 1.50	4.89 ± 2.07	0.187
Glycemic variability, %	25.94 ± 6.14	25.85 ± 6.07	26.1 ± 6.29	0.800
Hypoglycemic events in the last seven days, n	0.86 ± 2.03	0.77 ± 1.75	1.02 ± 2.47	0.818
Presence of hypoglycemia during the last seven days, n	78 (28.05%)	53 (29.12%)	25 (25.77%)	0.554
Number of hyperglycemic events in the last seven days, n	14.38 ± 6.93	14.73 ± 6.90	13.72 ± 6.98	0.318
Presence of hyperglycemia during the last seven days, n	278 (99.6%)	181 (99.45%)	97 (10%)	0.469
Time above range (>10 mmol/L), minutes	3756.77 ± 2487.35	3457.66 ± 2366.0	4317.99 ± 2621.81	0.009*
Time in range (3.9-10 mmol/L), minutes	5077.22 ± 2489.03	5438.54 ± 2406.13	4399.28 ± 2512.57	0.001*
Time below range (<3.9 mmol/L), minutes	45.84 ± 122.52	37.58 ± 95.96	61.34 ± 160.42	0.304
Laboratory data				
Telemonitored, n	145 (51.9%)	108 (59.34%)	37 (38.14%)	0.001*
HbA1c, mmol/mol	64.04 ± 14.22	61.93 ± 12.0	67.99 ± 17.04	0.006*
Diastolic, mmHg	81.61 ± 11.08	81.35 ± 10.51	82.08 ± 12.12	0.600†
Systolic, mmHg	138.11 ± 17.35	137.67 ± 17.10	138.94 ± 17.85	0.519
Mean arterial pressure, mmHg	100.44 ± 11.51	100.12 ± 10.86	101.03 ± 12.70	0.531
Diabetes-related information				
Minimum one diabetes-related complication, n	174 (62.4%)	119 (65.38%)	55 (56.70%)	0.155
Sum of diabetes-related complications, n	1.08 ± 1.13	1.09 ± 1.1	1.04 ± 1.19	0.427
Insulin type (basal + bolus), n	117 (42.04%)	81 (45.05%)	36 (37.11%)	0.235
Sum of anti-diabetic medication, n	1.68 ± 0.93	1.69 ± 0.95	1.66 ± 0.88	0.691
Ever experienced hypoglycemia, n	191 (68.5%)	124 (68.13%)	67 (69.07%)	0.873
Minimum one additional medication, n	265 (95.08%)	173 (95.05%)	92 (94.85%)	0.951
Sum of other medications, n	2.32 ± 1.03	2.28 ± 1.04	2.38 ± 1.0	0.289
Comorbidities				
Number of comorbidities, n	2.68 ± 0.95	2.66 ± 0.94	2.72 ± 0.98	0.443
Overweight, n	226 (81%)	148 (81.32%)	78 (80.41%)	0.855
Hypertension, n	220 (78.9%)	145 (79.67%)	75 (77.32%)	0.648
Cardiovascular disease, n	86 (30.82%)	54 (29.67%)	32 (32.99%)	0.569
Hyperlipidaemia, n	216 (77.42%)	137 (75.27%)	79 (81.44%)	0.239
Physical and mental data				
>5 hours of exercise, n	89 (31.9%)	61 (33.52%)	28 (28.87%)	0.429
Overall good health status, n	202 (72.4%)	135 (74.18%)	67 (69.07%)	0.369
Emotional state, (1; negative - 6; positive)	3.23 ± 1.43	3.2 ± 1.45	3.28 ± 1.40	0.739
Sadness within last four weeks, (1; all time - 6; at no time)	5.07 ± 1.08	5.17 ± 0.98	4.88 ± 1.22	0.081

TABLE Appendix 1: Features and characteristics of all participants divided into total, adherent, and nonadherent groups. The binary features are shown with the number of positive answers and the percentage share of participants in the group. The integers and floats are shown with the mean ± standard deviation. Mann Whitney U test with $p<0.05$ is performed on all features except the ones listed with †, where a 2-sample t-test with $p<0.05$ is performed. All features that are significantly different between adherent and nonadherent groups are listed with an *.

Model	ROC-AUC	Parameters	Attempted values
Logistic Regression	0.683	solver	newton-cg, liblinear, newton-cholesky, sag, saga, **lbfgs**
		penalty	none, **l2**, l1, elasticnet
		C	0, 0.01, 0.1, 0.2, 0.4, 0.6, 0.8, **1.0**, 1.2, 1.4, 10, 100
Support Vector Machine	0.636	kernel	**linear**, rbf, sigmoid, poly
		gamma	auto, **scale**
		C	0.1, 0.2, **1.0**, 1.5, 2.0, 5.0
Random Forest	0.749	max_depth	5, 6, 7, 8, 9, **10**, 15, 30, 50
		min_samples_leaf	1-16 **(1)**
		n_estimators	50, **100**, 200, 300, 1000
Linear Discriminant Analysis	0.692	solver	**svd**, lsqr, eigen
K-Nearest Neighbour	0.686	n_neighbours	1-30 **(5)**
		weights	**uniform**, distance
		algorithm	**auto**, ball_tree, kd_tree, brute
		leaf size	20, **30**, 50
		metric	manhattan, euclidean, **minkowski**, mahalanobis
Extreme Gradient Boosting	0.734	max_depth	**2**, 3, 4, 5, 6
		learning_rate	0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, **1.0**, 1.1, 1.2
		gamma	**0**, 1.0, 1.1, 1.25, 1.5, 2
		scale_pos_weight	0, 0.25, 0.5, **1.0**, 3.0
Multi-Layer Perceptron	0.684	hidden_layer_sizes	(25,), (50,), (100), (200,), **(300,)**, (400,), (500,)
		activation	identity, logistic, tanh, **relu**
		learning_rate	**constant**, invscaling, adaptive
		learning_rate_init	0.0001, **0.001**, 0.005, 0.01, 0.05
Naïve Bayes	0.707	None	None

TABLE Appendix 2: Hyperparameter optimization for all models, where different settings of hyperparameters were tested in the algorithms using grid search. The ROC-AUC score and the best combination of hyperparameters are listed in the table. The selected hyperparameters are underlined and marked in bold.

Contents

1 Problem Analysis

1.1 What is diabetes?

Diabetes mellitus, more commonly known as diabetes, is a chronic disease that affects the use of glucose in the human body. Glucose is one of the body's main sources of fuel and the most important fuel for the brain. Glucose comes from food intake or is released by the liver. The liver holds excess glucose from previous meals which is released when the glucose levels are too low. When glucose enters the bloodstream, it circulates until it can enter the cells. For glucose to enter the cells, it needs insulin. Insulin is a hormone that is produced by the pancreas. In healthy people, insulin secretion will increase and decrease concurrently with the presence of glucose in the bloodstream. In people with diabetes, the pancreas does not produce enough or any insulin, which results in elevated glucose levels in the bloodstream that do not reach the cells. Continuously high glucose levels in the bloodstream can lead to serious health problems such as nerve damage, heart disease, kidney damage, or eye damage. [1], [2], [3]

In 2014, the World Health Organization (WHO) estimated that 422 million people lived with diabetes worldwide. This equates to 8.5% of all adults having diabetes. This is a rapid increase from the 108 million people living with diabetes in 1980. Furthermore, WHO estimated that 1.5 million deaths were directly caused by diabetes in 2019, with 48% of all deaths being before the age of 70. [4]

1.2 Type 1 and type 2 diabetes

There are two common types of diabetes. Type 1 diabetes (T1D) is an autoimmune disease where the immune system attacks the cells that produce insulin in the pancreas. This eventually destroys them, leading to no insulin being released into the bloodstream when there are elevated glucose levels. Type 2 diabetes (T2D) occurs due to insulin resistance, where the body's cells do not respond sufficiently to insulin. This causes the pancreas to produce extra insulin; thus, the insulin demand will increase and the pancreas will not be able to keep up, causing the blood sugar to rise. T1D cannot be prevented, whereas T2D can be prevented or delayed. Known factors for developing T2D are obesity, a lack of physical activity, and a family history of diabetes. T2D is the cause of 90-95% of all diabetes cases, whereas T1D only occurs in 5-10% of the cases. Depending on the glucose levels in the bloodstream, different symptoms can arise. People with T1D often have more severe symptoms, whereas people with T2D might not have any. This can be seen as 11.3% of the adult US population lived with diabetes in 2019, but almost 1 in 4 did not know they had it since they did not show any symptoms. Symptoms of diabetes can include more frequent urination, feeling more thirsty, unexplained weight loss, fatigue, blurry vision, slow healing sores, and mood swings or irritation. [2], [1], [5], [3]

1.3 Test for diabetes

Different tests can be made to check for diabetes. One is the glucose tolerance test, where the person fasts overnight and is measured regularly after drinking a liquid with sugar. If the glucose levels in the blood are more than 200 mg/dL, the person is considered to have diabetes. Another test can be made where the patient's fasting blood sugar is measured without the sugary drink and the person fasts for at least eight hours. If the fasting blood glucose levels are 126 mg/dL or higher, based on two separate tests, the patient is considered to have diabetes. Random blood glucose level tests can also be conducted regardless of the time since the last meal. A blood glucose level of more than 200 mg/dL suggests diabetes. Another common test that can also be used to monitor how a patient controls their diabetes is the HbA1C test. The test measures the average blood glucose levels from the previous two to three months by measuring the percentage of glucose attached to hemoglobin in the blood. Hemoglobins are proteins that carry oxygen, which glucose likes to attach to. An HbA1C level of more than 6.5% indicates that you have diabetes. [1], [6], [3]

1.4 Treatment

People with T1D require insulin to live since their pancreas does not produce any. People with T2D can sometimes control their glucose levels through lifestyle changes, but some need additional medication as well. This is either insulin or different kinds of oral medications. Insulin works either rapid-acting, short-acting, or long-acting. They all have different onsets and have different lengths of effect. Rapid-acting works within 15 minutes and has a duration of two to four hours. Short-acting has an onset of 30 minutes and a duration of three to six hours, whereas long-acting has an onset of two hours and a duration of 24 hours. Insulin can either be injected or inhaled. The most common are injections with syringes, an insulin pen, or an insulin pump. Oral medication varies depending on the desired effect. Some oral medications help the pancreas release more insulin, whereas others prevent glucose from being released by the liver. This means that the insulin requirement is reduced. Other oral medications block the enzymes that break down carbohydrates into glucose, which increases insulin sensitivity. [6], [1], [7]

Managing diabetes is a life-long process that requires routine check-ups with the health care team. These check-ups should be held at least twice a year to find and treat health issues early and possibly prevent them. Every person responds differently to an anti-diabetic treatment due to factors such as lifestyle, medication effectiveness, additional health conditions, etc. Therefore, the personalized treatment plan needs a review at each check-up concerning medication adjustment, lifestyle changes, etc. If the check-ups are neglected, it can have consequences for the patient's general health due to the progression of diabetes. [8]

1.5 State-of-the-art

A structured literature search was conducted to cover the current research on medication adherence in patients with T2D. The findings can be seen in Chapter 3 Literature Search. All included articles are used to investigate the state-of-the-art, whereas only some results are presented to specify the aim of this study.

1.5.1 Medication adherence

Only about 50% of people living with chronic diseases are likely to adhere to their medication. WHO defines adherence as *"the extent to which a person's behavior – taking medication, following a diet, and/or executing lifestyle changes, corresponds with agreed recommendations from a health care provider"* [9]. This lack of adherence and reduced treatment was expected to expand the worldwide burden of chronic diseases, leading to an increase from 54% in 2001 to 65% in 2020. [9] Diabetes is a chronic disease where sufficient management is important to delay or prevent diabetes-related complications. Some complications associated with uncontrolled T2D are poor glycemic control, higher HbA1c levels, microvascular diseases, macrovascular diseases, and other comorbidities. [10] [11] Besides anti-diabetic medication, management includes a healthy lifestyle with an increase in physical activity and low-calorie intake to keep the blood glucose levels within range, resulting in personalized treatment plans. Diabetes management can be complex, and studies have shown that adherence to anti-diabetic medication for T2D is not optimal. [12], [13]

The study by Lee and Lee [14] found that adherence to anti-diabetic medication was one of the keys to adequately managing T2D. Medication nonadherence in T2D patients is a growing healthcare issue that has a reducing effect on treatment. This reduction results in additional medication and hospital visits, increasing the financial burden. [15], [10], [16] A study by Eby et al. [17] aimed to estimate the association between adherence and health care costs in American T2D patients. They found that adherent patients generally had better health and had significantly fewer hospital visits. Even though adherent patients have significantly higher diabetes-related and drug-related costs, they have significantly lower all-cause total costs as well as acute costs. Furthermore, a study by Shiyanbola et al. [18] estimated that if medication adherence is improved by 10% it would result in a 6.6% reduction in hospitalizations.

1.5.2 Machine learning models to predict nonadherence

Improving medication nonadherence in T2D patients is more beneficial than developing new treatments [16]. Patients with T2D often do not have noticeable symptoms in the event of high blood glucose, causing them to deviate from the treatment plan [19]. Early detection of nonadherence is the premise for effective interventions. Therefore, it is essential to identify the associated factors that could characterize adherent patients and nonadherent patients. This would help clinicians to conduct a more personalized treatment plan, increasing the adherence level of T2D patients. [20], [21], [22] Few studies have used machine learning to develop a model to predict nonadherence in T2D patients and differentiate between groups, with promising results. A study by Wu et al. [23] developed 300 models to screen for the risk of nonadherence. The models were based on 30 machine learning algorithms using 16 different variables. The study found that the best model used nine variables and had a Reciever Operator Characteristics Area Under Curve (ROC-AUC) of 0.866.

Furthermore, a study by Li et al. [22] tested 1080 different models, and the best-performing model could predict nonadherence with a ROC-AUC of 0.8369. A study by Chen et al. [21] used a model based on XGBoost to predict nonadherence in T2D patients with an ROC-AUC of 0.771. These results indicate that using machine learning models for early identification of nonadherent patients has the potential to improve patient care and reduce workload.

1.5.3 Factors associated with nonadherence

Several studies have found factors associated with medication nonadherence. As mentioned previously, the study by Wu et al. [23] developed machine learning models to screen for the risk of nonadherence in patients. Based on a feature selection, the study found age, BMI, working status, last HbA1c value, fasting glucose, weight, and disease duration among the best factors to distinguish patients. This is followed by Li et al. [22], who also found age, BMI, and HbA1c levels to be significant factors in differentiating between adherence and nonadherence. Additionally, studies by Lee and Lee [14], Sun and Lian [24], Iglay et al. [25], and Campbell et al. [26] found that a younger age had a negative correlation with adherence. Another study by Huang et al. [27] found that neuroticism had a negative correlation with adherence ($p<0.001$), whereas social support and self-efficacy had a positive correlation ($p<0.001$). This is followed by Ranjbaran et al. [28] which found neuroticism to have a significant impact on adherence, whereas studies by Kretchy et al. [29] and Jackson et al. [30] found anti-diabetic medication to have an impact on adherence. Furthermore, studies by Masaba and Mmusi-Phetoe [31] and Kumar et al. [32] found poor knowledge of diabetes as an important factor for nonadherence.

1.5.4 Definition of medication adherence

Adherence is a measure that is often self-reported or derived from pharmacy claims. By using self-reported measures to assess medication adherence, potential errors with recall bias and social desirability bias can occur, as stated in the study by Nelson et al. [33]. Recall or social desirability bias occurs when answers are potentially over-reported or under-reported. This especially happens with sensitive questions where the problem may be underestimated. Bias is prone to errors that misrepresent the image of medication adherence, resulting in wrongful factors being identified. These limitations were also stated in the studies by Azri et al. [34], Hashimoto et al. [35], Abdullah et al. [36], and Lee et al. [15]. Therefore, adherence must be defined based on objective measures to obtain precise answers on medication adherence.

A method to define adherence objectively is to use insulin injection data, as it gives precise answers on dosage and time of injection. To the best of our knowledge, a study by Nørlev et al. [37] is the only current study that has used insulin injection data to quantify adherence. They classified patients as adherent if $\geq$80% of the doses were administered correctly and as nonadherent if $<$80% of the doses were administered correctly based on injection data from 12 weeks. Additionally, they calculate an overall adherence level for each week to identify adherence patterns in T2D patients. The method developed in the study provided detailed information on insulin administration and identified certain nonadherence behaviors. As a result, the study found that 50.5% of the participants were considered overall adherent. Furthermore, the study found that 49% of the incorrect doses were increased doses, 32% were reduced doses, and 19% were missed doses. These results were based on 103 participants and indicate that injection data might be beneficial as an objective measure to explore medication adherence.

Aim of this study 2

2.1 Our investigation

T2D is a chronic disease that affects the insulin production in the pancreas or the insulin sensitivity of the cells. In the last decades, the prevalence of T2D has increased rapidly, making it a worldwide healthcare issue. T2D can be managed through lifestyle changes and anti-diabetic medications such as insulin or oral medications. Management of T2D is complex, and every patient has a personalized treatment plan. Thus, studies have shown that adherence to anti-diabetic medication is not optimal, resulting in diabetes-related complications. Improving medication adherence is a key factor in treating T2D and is seen as more beneficial than developing new medications. As people with T2D often do not experience symptoms of high or low blood glucose, it can result in patients deviating from their treatment plan. To improve patient outcomes, there is a need for early detection of nonadherence and finding related factors that can aid in identifying and characterizing patients who are prone to being nonadherent. Several studies have found factors associated with nonadherence, whereas only a few have used them with machine learning models to identify nonadherent patients. Most current studies have used self-reported measures or pharmacy claims to define adherence, which are prone to bias. Introducing bias in the definition of adherence could potentially lead to the identification of incorrect factors, which gives clinicians invalid information. Therefore, this study aims to develop a machine learning model based on data from a connected insulin pen to identify nonadherence in people with T2D. The connected insulin pen measures a more precise administration of injected insulin, thereby reducing bias [38; 32]. This identification model may assist clinicians in identifying nonadherence earlier and establishing better glycemic control.

Literature Search 3

3.1 Unstructured Literature Search

An unstructured literature search was made to gain knowledge about the topic of the project. This was done to identify relevant search terms which could be used in the structured literature search. The following questions were made to concertize the topic:

- What is T2D and how many people does it affect?
- Which types of treatment are available for T2D patients?
- How common is medication adherence among T2D patients?
- Can machine learning or artificial intelligence be used to optimize the treatment of T2D?

It was found that T2D is a disease that affects many people worldwide. Treatment of T2D can be a variety of different medications, including insulin injections and oral hypoglycemic medications. However, many of the patients do not take their medications as prescribed, which can have serious consequences for their general health. Many studies aimed to identify factors associated with medication adherence in T2D patients. Still, only a few have used machine learning or artificial intelligence to classify adherence and identify the related factors. Based on these findings, the initial research question for this project is:

- Which characteristics are associated with medication adherence in T2D patients and is it possible to identify them using machine learning or artificial intelligence?

A structured literature search was made to cover the current research in the field of medication adherence in patients with T2D. The search was formed as a block search, and the search terms were based on the previous unstructured literature search findings. The block search was done with inspiration from well-known search tools, such as the *Population-Intervention-Comparison-Outcome* model and the *Population-Exposure-Outcome* model. The boolean operators AND/OR were used to combine the search terms into search strings. Table 3.1 shows the included search terms, the different blocks of search terms, and the boolean operators AND/OR.

	AND			
OR	Type 2 diabetes mellitus T2D Type 2 diabetes	Insulin Medication	Artificial intelligence Machine learning Characteristic*	Adherence

Table 3.1. The block search for the structured literature search.

The searches were performed in both PubMed and Embase. To meet their respective requirements for a search, different search strings were made. The first structured literature

search was made without the term *Characteristic**, due to the main focus on machine learning and artificial intelligence as methods. This search returned a total of 49 articles across the two databases. For that reason, the search term *Characteristic** was included to gain more articles about associated characteristics with adherence. The inclusion of this search term increased the number of found articles to a total of 912 articles. The final search strings for PubMed and Embase were as follows:

- **PubMed:** ("T2D"[Title/Abstract] OR "Type 2 Diabetes Mellitus"[Title/Abstract] OR "Type 2 diabetes"[Title/Abstract]) AND ("Adherence"[Title/Abstract]) AND ("Machine Learning"[Title/Abstract] OR "Artificial Intelligence"[Title/Abstract] OR "Characteristic*"[Title/Abstract]) AND ("Insulin"[Title/Abstract] OR "Medication*"[Title/Abstract]) AND (2010:2024[pdat])
- **Embase:** ('t2d':ab,ti OR 'type 2 diabetes mellitus':ab,ti OR 'type 2 diabetes':ab,ti) AND 'adherence':ab,ti AND ('machine learning':ab,ti OR 'artificial intelligence':ab,ti OR 'characteristic*':ab,ti) AND ('insulin':ab,ti OR 'medication*':ab,ti) AND ([danish]/lim OR [english]/lim) AND [2010-2024]/py

A set of inclusion and exclusion criteria was defined to limit the number of included articles. The criteria can be seen in Table 3.2 and were defined according to the knowledge of the topic. The purpose of the criteria is to ensure that only studies relevant to this project are included.

Inclusion criteria	Exclusion criteria
Publication date between 2010-2024	Type 2 diabetes is not the primary disease
Available in English or Danish	Adherence to diabetic medication is not the main focus
Full text is available	Animal studies
	Not usable publication type
	Investigating different types of medication

Table 3.2. Inclusion and exclusion criteria for the structured literature search. The right column contains the inclusion criteria and the left column contains the exclusion criteria.

To visualize the screening process, a PRISMA diagram was made. The screening process was divided into three parts: Identification, screening, and inclusion. In the identification part, the searches were conducted in the scientific databases using the search strings. The results were summarized into a total of 912 articles. In the screening part, all articles were screened for duplicates, which excluded 290 articles. The abstracts and titles for the remaining 622 articles were screened based on the inclusion and exclusion criteria. This screening excluded additionally 522 articles. The last part of the screening was full-text screening for eligibility, which excluded 51 articles. The inclusion part summarizes the total number of articles included in the analysis. A total of 49 articles were found eligible in the structured literature search. The PRISMA diagram can be seen in Figure 3.1.

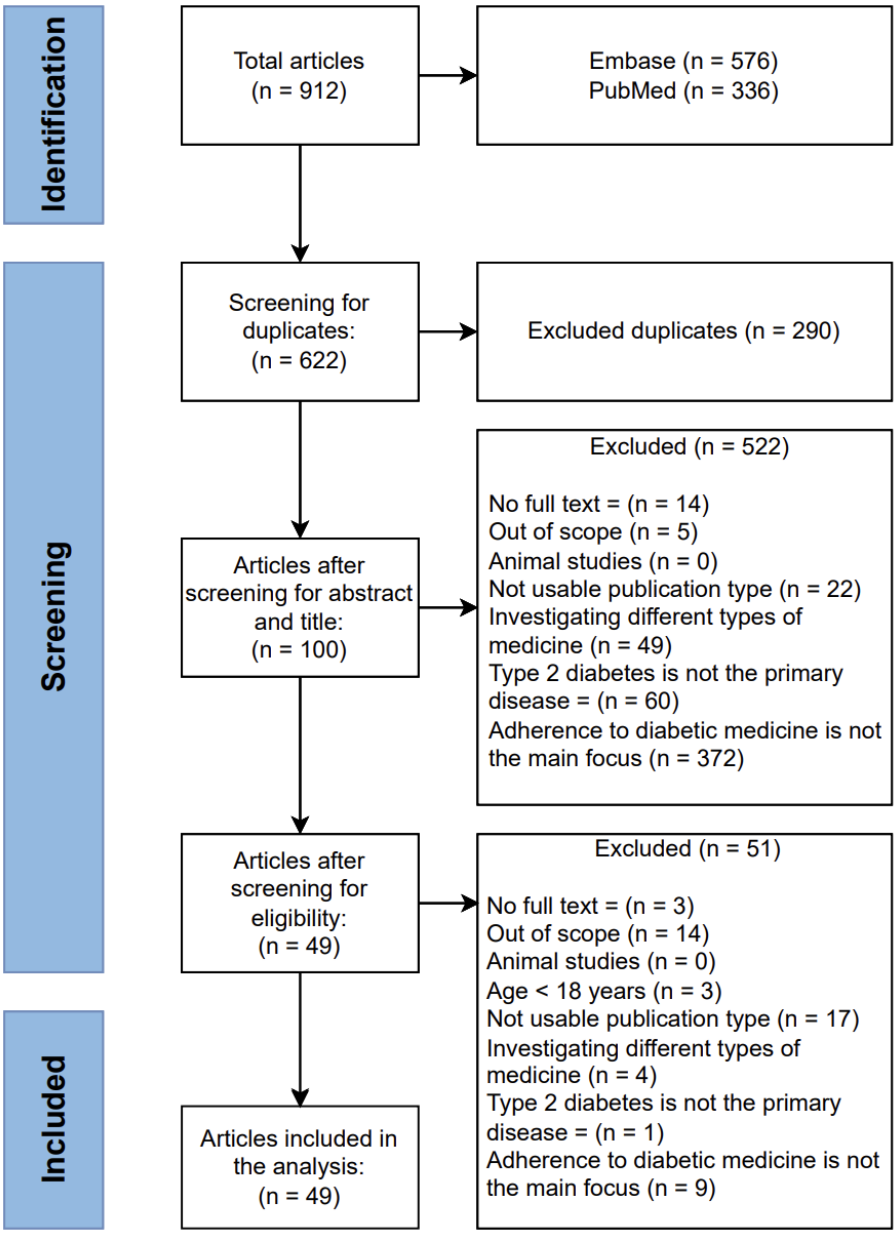


Figure 3.1. PRISMA diagram illustrating the sequence for the structured literature search. 912 articles were identified from the two databases, whereas 290 were found to be duplicates. After screening for abstract and title 100 articles were left for full-text screening. After full-text screening, 49 articles were found suitable based on the inclusion and exclusion criteria.

3.2 Synthesis

At the end of the structured literature search, 49 articles were chosen for full-text review. All articles used data related to the patient's demography, clinical characteristics, socioeconomic status, laboratory tests, and T2D-related information. The most frequently used method for data collection was questionnaires, which were used in 27 of the included 49 articles. Other methods used for data collection were data from existing databases, monitoring, interviews, and

simulations.

3.2.1 Adherence definition tools

There were different tools to define medication adherence, whereas the most common scale was Morisky Medication Adherence Scale-8 (MMAS-8). MMAS-8 is a validated assessment method to quantify nonadherence using short behavioral questions [39]. The scale was used in 13 of the included 49 articles. Other well-known tools are the Proportion of Days Covered (PDC) and the Medication Possession Ratio (MPR). PDC calculates the number of days on medication based on prescriptions divided by the number of days within a period, whereas MPR calculates the sum of days' supply in a period divided by the number of days in the period [40]. PDC was used in seven of the included studies, and MPR was used in five studies. The rest of the included studies used other tools or scales to define medication adherence. However, many of the smaller scales and questionnaires were inspired by the MMAS-8.

3.2.2 Methods used for data analysis

Different methods were used in the articles to identify factors associated with adherence or to classify whether or not a patient was adherent. The main method was statistical analysis which was used in 43 out of the 49 articles. The most common type of statistical analysis was logistic regression. Among other methods used were machine learning models which were used in six out of the 49 articles. Different machine learning algorithms were tested and some of the most commonly used were XGboost and SVM. The rest of the studies used methods such as nomogram models, linear models, and linear mixed models.

3.2.3 Features

Throughout the articles, different factors have been assessed by their ability to predict or characterize adherence and nonadherence. 17 articles found age to be a dominant factor in predicting adherence with young patients being more likely to be nonadherent in eight of the 17 articles. Furthermore, seven articles found gender to be a factor that could differentiate between adherence and nonadherence. Three of the seven articles found women to be more adherent, whereas two of the seven articles found men to be more adherent.

Six articles found that higher education impacted adherence whereas three articles found income level, and three articles found working status to be associated with adherence. Overall, the literature search showed a tendency between diabetes-related knowledge and adherence. 13 articles found that lack of knowledge about diabetes was negatively correlated with adherence. Besides knowledge, neuroticism was also found to be a factor that could predict adherence in eight of the articles. If patients expressed concerns, anxiety, or fear of hypoglycemia they were more likely to be nonadherent.

The type of medication, amount of medications as well as dose were also found to be associated with adherence. Seven articles found that the amount of medication was positively correlated with adherence, whereas six articles found that a higher dose was negatively correlated with adherence. Furthermore, four articles found that the type of medication used impacted the prediction of adherence. Through the literature search, it was also found that the duration of the disease and the amount of comorbidities were an associated factor in predicting nonadherence.

Six articles found the duration of disease to be significant whereas another six found comorbidity to be significant.

3.3 Overview of included studies

To give an overview of the 49 included articles, Table 3.3 was made. All articles are presented by citation, participants, methods, features for adherence, and main results.

Citation	Partipicants	Methods	Features for adherence	Main results
Huang et al. [27]	483 patients from China	MMAS-8, EPQ-RS, MSPSS, DMSES, statistical analysis	Neuroticism, social support, and self-efficacy had direct or indirect effects on adherence	Social support (P=0.029) and self-efficacy (P=0.023) directly influenced adherence. Neuroticism indirectly affected adherence through social support (P=0.023) and self-efficacy (P=0.014). Neuroticism was negatively associated with adherence (P<0.001), whereas social support (P<0.001) and self-efficacy (P<0.001) were positively associated with adherence.
Nelson et al. [33]	237 patients from Tennessee, USA	BHLS, ARMS-D, SDSCA-MS, statistical analysis	Nonadherence: Younger age and lower health literacy	Only 7% of participants reported no adherence barriers. The most frequent barriers were forgetting to take doses, pain when injecting insulin, disappointment, when medicine doesn't improve diabetes right away, and feeling burned out regarding taking diabetes medications
Lee and Lee [14]	48 articles published from 2017-2022	Systematic review	The highest adherence was observed in metformin users. The lowest rates was injectable therapies such as insulin.	Most studies reported adherence as a PDC>0.8. The most frequent cause across the studies for low adherence was the severity of adverse events. Baseline characteristics, demographic information, and comorbidity profiles have significant impacts on adherence.
Thyde et al. [41]	In-silico CGM	Classification CNN, data simulation using MVP model, logistic regression	-	The best-performing model could detect adherence and nonadherence 16 hours after the expected time of injection with a mean ensemble test accuracy of 79.8%. The simple feature-engineered logistic regression model performed almost as well as more complex deep learning models.
Chen et al. [21]	4,134 patients from Taipei	XGBoost models with 66 features, 5-fold cross-validation, statistical analysis, MPR	Adherence: Had a higher number of medications. Nonadherence: The dosage of index insulin was higher	40.14% of the patients were nonadherent. The average ROC-AUC from experiment 1 with internal testing was 0.782 and for experiment 2 with external testing 0.771.
Syafhan et al. [42]	121 patients from three hospitals in Ireland.	MARS, BMQ, CES-D, dried blood spot samples, patient interviews, logistic regression	Adherence: Metformin self-administration and use of purchased adherence pill box	61.2% of the patients were considered adherent, but from the questionnaire, 90.9% of the patients described themselves as adherent. Additionally, 102 patients had metformin exposure levels that fell within the therapeutic range. 17 patients had low exposure, and one person had undetectable metformin levels in their blood sample.

Masaba and Mmusi-Phetoe [31]	15 articles from Kenya	Systematic review	Adherence: Knowing the effects of non-adherence, knowledge of the disease process, family support and not taking excessive alcohol. Nonadherence: Multiple drugs, unsatisfactory health messages from health providers, unaffordable care, and indirect fees in health care	Three major domains: Cost, patient characteristics, and health system were associated with nonadherence
Eze et al. [43]	200 patients from Nigeria	A three-part structured questionnaire, statistical analysis	Moderate adherence were associated with self-glucose monitoring	Class of medicine and sociodemographics were not associated with adherence (P>0.05). Only two patients were observed as high adherent. 159 patients had poor glycemic control.
Azri et al. [34]	249 patients from Malaysia	Self-reported questionnaire, IAQDM, statistical analysis	More self-monitoring of blood glucose, exercise, more complementary medicine, and a higher number of insulin injections were associated with good adherence	228 patients were described as nonadherent. No significant association between socio-demographics or disease-related factors and adherence to insulin was found. Significant associations between the number of daily insulin injections, use of complementary medicine and self-monitored blood glucose, and adherence to insulin
Li et al. [22]	980 patients from Sichuan Hospital	Machine learning, questionnaires, 10-fold cross-validation, statistical analysis	Age, BMI, present fasting blood glucose, present HbA1C values, and random blood glucose values were the most significant factors associated with adherence.	184 patients were defined as nonadherent. A total of 1080 models were developed, whereas the best model scored AUC = 0.8369, accuracy = 0.9474, and recall = 0.6792
Jackson et al. [30]	303 patients from Nigeria	Statistical analysis, MMAS-8	Low literacy level, forgetfulness, cost of medication, lack of access to care, regimen-related factors, poor patient-provider communication, lack of trust in the provider, and depression had a significant impact on adherence.	19.8% of the patients were highly adherent to their medicine, whereas 50.2% were low adherent.
Wulandari et al. [44]	143 patients from Indonesia	Questionnaire, blood samples to measure HbA1c, statistical analysis	Patients who had T2D for less than five years tend to have low adherence	75.5% of the patients had low medication adherence. The duration of T2D was significantly related to the level of medication adherence

Horsburgh et al. [45]	85,066 patients from New Zealand	Data from national data collections, a linear mixed spline model, MPR, statistical analysis	Nonadherence: Time since initiating metformin, younger and Māori or Pacific ethnicity. Adherence: Receiving more nondiabetic medications, history of CVD, and cancer registration.	The number of patients with an MPR>0.8 was 63% in the first year and dropped in the following years. Māori and Pacific people had the lowest adherence.
Ranjbaran et al. [28]	734 patients from South Tehran, Iran	MMAS-8, HAPA	Adherence: behavior intention, task self-efficacy, coping planning, and copping self-efficacy, gender (women)	82.3% had low adherence and six features were found to be statistically significantly different between the adherence and nonadherence groups.
McClintock et al. [20]	72 patients from West Philadelphia, Pennsylvania.	PPP, multinomial logistic regression, sociodemographic characteristics, MMAS	Adherence: Intervention or not, HbA1C	Three patterns of adherence were identified, adherent, increasing adherent, and nonadherent. Patients in the intervention group were more likely to be adherent and mean HbA1c was significantly different between the groups.
Wu et al. [23]	401 patients from outpatient clinic of Sichuan Provincial People's Hospital	Questionnaires, machine learning, Wilcoxon rank-sum analysis, Kruskal-Wallis test	Last HbA1c value, fasting glucose, age, diet adjustment or not, weight, cost of hypoglycemic drugs, duration of current treatment regimen, BMI, working status, the duration since the prior blood glucose test, dyslipidemia	The best model was Ensemble and had a ROC-AUC at 0.87
Lee and Lee [46]	236 patients from Seoul National University Hospital, South Korea	Multivariate linear regression, multivariate logistic regression, PDC	Adherence: Increase in age, switching dose, and neuropathy at baseline	The study found clinical characteristics of dulaglutide users that could affect adherence. The findings can be used by clinicians treating T2D patients to optimize their adherence to dulaglutide.
Vlacho et al. [47]	1,205 patients from Spain	T-test, Mann-Whitney tests, logistic regression, PDC	HbA1c, triglycerides, and total cholesterol can be used as adherence indicators	The results showed a statistically significant difference between the groups in HbA1c, triglycerides, and total cholesterol.
Hashimot et al. [35]	157 patients from Japan	Questionnaires, PCA, cluster analyses	BMI, family history of diabetes, one factor of patient's perception, diabetes knowledge	The PCA found two components: 1) accessibility to medical treatment, and 2) status of taking medicines. The cluster analysis identified four groups of medication adherence using the PCA components.

Parada et al. [48]	302 patients from Imperial County, Southern California	MMAS-4-Item, multivariate logistic regression	Nonadherence: Males, with a lower frequency of engaging in personal actions, have depression	The study classified 60% of the patients as nonadherent. Patients with a high school education or higher and who positively rated their health were likely to be classified as nonadherent.
Abdullah et al. [36]	232 patients from Selangor, Malaysia	Questionnaires, electronic medical records data, Multivariate logistic regression, MCQ	Ethnicity, marital status, income level, employment status, duration of T2D diagnosis, HbA1c level, number of drugs taken, type of medications taken	The multiple logistic regression analysis showed that ethnicity and HbA1c were the only significant factors
Eby et al. [17]	23,365 patients from USA	Patient characteristics, t-statistics, A generalized linear model, PDC, A separate model estimated the cost of adherence and the cost of nonadherence	Adherence: older, male, and had higher BMI, better general health, fewer comorbidities, visits to ER/hospitals, and were prone to receive oral medicine.	41.4% of basal patients were adherent and 19.9% of basal-bolus patients were adherent. Multivariable analysis showed that adherent patients treated with basal insulin had a significantly lower total cost compared to the nonadherent group ($30,127 vs. $37,049). The same can be seen in the basal-bolus group where the cost was ($36,603 vs. $44,702). Furthermore, adherent people generally had higher drug costs.
Sánchez-Hernández et al. [38]	3,536 patients from Castilla y León, Spain	MMAS-4-Item, questionnaires, 14-point MEDAS, bivariable analysis, multivariable analysis	Nonadherence: lower educational level, sedentarism	38.8% were nonadherent and had lower educational levels, sedentarism was found to be the main factor associated. 33.7% had poor glycaemic control and younger age, rural residence, tobacco use, time since diagnosis and polypharmacy were the factors associated.
Cheng et al. [49]	7,728 patients from Taiwan	Generalized estimating equations, MPR	Adherence: Less hospitalization, higher healthcare expenses in the first five years, and better medical outcomes.	The results showed that the nonadherent group was at a higher risk of hospitalization, but the total healthcare expenses were higher for the adherent group in the first 5 years with T2D.
Nazir et al. [50]	392 patients from Pakistan	Questionnaire, MMAS-8, MDKT-U	-	The results showed that HbA1c had a non-significant association with diabetes-related knowledge and medication adherence. 71.94% had low medication adherence.
Sun and Lian [24]	192,717 patients from US National Healthcare Claims Databases	MPR, Logistic regression	Adherence: Older age, type of insurance	The average MPR was 0.74. The patients with adherence MPR0.8 had significantly fewer mean annualized inpatient admissions compared to nonadherent patients with MPR 0.8.

Gatwood et al. [51]	159,032 veterans from Veterans Affairs Corporate Data Warehouse	t-tests, chi-squared test, PDC	-	Patients nonadherent to their medications were more likely to experience bad health outcomes.
Buysman et al. [52]	1,321 patients from US health plan	PDC, MPR, multivariate analysis	Adherence: older, male, reduction in HbA1c	The mean HbA1C reduction from baseline to follow-up was greater in the adherent group compared with the nonadherent group.
Kristy et al. [25]	133,449 patients from a US-based database	PDC, logistic regression analysis	Nonadherence: Younger, new to therapy, on a twice-daily dose, female, on fewer than three concomitant medications.	59% were found to be adherent to their medication. The mean PDC was 75%.
Wong et al. [53]	565 patients from China	Patient interviews, MMAS-8, Spearman correlation test, linear regression model, backward stepwise algorithm	Age and exercise	Negative correlation between HbA1c and MMAS-8 scores. 67.8% were adherent to their medicine and had lower income and optimal glycemic control. The nonadherent patients used antidepressive medicine, and lipid-lowering agents, had good dietary compliance, had regular exercise, and were smokers and drinkers.
Saundankar et al. [10]	238,402 patients	Gradient boosting trees, sensitivity analysis, 1-tailed statistical test, PDC	Use of mail-order pharmacy at baseline, 90-day prescriptions, the longest gap in oral medication therapy (7-day increments), use of the sulfonylurea drug class, diabetes-related pill burden at baseline, the month-wise oscillation between adherence statuses	Had 91 predictors and found the five best. 21.7% of the adherent patients became nonadherent predicted by the model with 76% sensitivity and 57% specificity. 41.7% of the nonadherent patients changed to adherent patients with 53% sensitivity and 71% specificity.
Vervloet et al. [54]	104 patients from the Netherlands	A real-time medication monitoring system for six months, statistical analysis, multilevel analysis, multilevel logistic regression, bivariate analysis	Age, amount of intake per day, and interruptions in the daily routine have a negative influence on medication intake	In total 36,199 medication intake moments were analyzed. Medication taken in the evening and during the weekends was more likely to be incorrect or missed. 61% of correct intakes occurred on Monday and Tuesday mornings, whereas 33% were correct on Sunday evenings.

Kretchy et al. [29]	188 patients from Ghana	PCA, questionnaires, MARS, Shapiro-Francia test, chi-square test, fisher, Wilcoxon signed rank, binary logistic regression model	Distress, discouragement, uncomfortable social situations, anger, anxiety, guilt, loneliness, and burnout	The study found that 44.7% showed high distress due to diabetes. 33% were adherent to their medication. The study found that the patients with high distress were 68% less likely to be adherent. There were significant associations between adherence and distress (discouragement, uncomfortable social situations, anger, anxiety, guilt, and loneliness)
Kumar et al. [32]	118 patients from Berhampur, Odisha	Questionnaires, MCQ, statistical analysis, chi-square test, and multivariate logistic regression	Age, socioeconomic status, residency status, medication knowledge, and comorbidities	Older age and low socioeconomic status were negatively associated with adherence. Residency status, comorbidities, and higher medication knowledge were positively associated with adherence. The current adherence among T2D patients is very low.
Yong et al. [55]	360 patients from Malaysia	MCQ, PHQ-8, statistical analysis, t-test, ANOVA, multivariate linear regression	Forgetfulness, complicated regime, fear of hypoglycemia, work commitment	60.3% of the patients were adherent. Positive correlation between HBM model and insulin adherence (except perceived barriers - significant negative effect). The HBM model predicted a 40.9% variance in insulin adherence. Age, year of diagnosis, duration of insulin, comorbidities, and depressive symptoms significantly influenced the HMB model.
Campbell et al. [26]	17,932 persons from Alberta, Canada	A validated algorithm, statistical analysis, multivariable log binormal regression modeling, multivariable generalized linear regression, multivariable-adjusted model, PDC	Adherence: Older age, having comorbidities, high neighborhood income, other drugs than metformin	48% were nonadherent. Has results regarding prescription patterns and the likelihood of the type of drug based on age (not written due to relevance).
Wang et al. [56]	338 patients from China	Univariate and multivariate logistic regression, nomogram model, MMAS, HAM-A, HAM-D, statistical analysis	Educational level, monthly income, negative emotions, family members reminding of medication, drug affordability, number of drugs, daily doses, and time spent taking medicines	226 patients had good adherence and 112 had poor adherence. Based on the significant features for nonadherence the nomogram was developed with an accuracy of 0.749 (C-index)

Krishnan and Roselin [13]	93 patients from Chennai tertiary care hospital	A semi-structured interview and statistical analysis	Nonadherence: Forgetfulness, feeling worse, religion, had diabetes for more than five years, not regularly physically active, more than 50 years, lower educational level, no family member support, no proper diet	66% of the patients were found to be adherent. 49.5% reported forgetfulness and 18.3% reported feeling worse after taking medicine. Patients who were on mono-therapy were more likely to be nonadherent and patients who were not physically active were three times more likely to be nonadherent.
Nazir et al. [57]	392 patients from Sargodha, Pakistan	Questionnaires, MMAS, EQ-5D, and Spearman rank order correlation	61.22% reported forgetting and 48% reported carelessness	71.93% were categorized with poor adherence, 24.75% with medium, and 3.3% with high adherence. The study highlighted that the T2D patients had decreased HRQoL. There was a significant but weak positive correlation between HRQoL and treatment adherence.
Nazir et al. [12]	392 patients from Sargodha, Pakistan	Questionnaire, MMAS, statistical analysis	61.22% reported forgetting and 48% reported carelessness	71.93% were categorized with poor adherence, 24.75% with medium, and 3.3% with high adherence. 28% had poor knowledge, 62.5% had moderate, and 8.67% had adequate knowledge. No statistically significant correlation between knowledge and adherence was found, there was a positive correlation between good knowledge and adherence.
Sapkota et al. [58]	52 studies from 15 countries	Systematic review	-	The study found a significant increase in studies implementing and evaluating interventions to promote adherence. They found many interventions to improve self-care and were delivered by nurses or pharmacists. Only a few studies found significant improvement in adherence to medication, adherence, and HbA1c levels.
Egede et al. [59]	11,272 veterans from Southeastern United States	MPR, statistical analysis, generalized linear mixed model	Nonadherence: Poor glycemic control, irregular refill patterns, and ongoing use of diabetes medications	Approximately 97% of the participants were males. There was a 48% decrease in the odds of poor glycemic control for each percentage increase in MPR. The odds for poor glycemic control were 1.3 higher for employed veterans.
Kang and Hur [60]	175 patients from Laos	Questionnaires, MMAS-8, statistical analysis, ANOVA, t-test, Pearson coefficient	Employment status (having a job = adherence), duration of illness (short duration = adherence).	10.3% had high adherence, 59.4% had medium adherence. Adherence was significantly different in employment status, and duration of illness. There was a positive correlation between adherence and self-efficacy.

Iqbal et al. [16]	300 patients from Pakistan	Questionnaires, DAI-10, MDK, statistical analysis, KS test, Mann-Whitney U, Kruskal-Wallis test	Age, gender, diabetes knowledge, education	7.3% were low adherent, 37% were medium adherent and 55.6% were adherent. Significant associations between age, gender (male), education, diabetes knowledge, and adherence. An increase in knowledge score of 1 increased adherence by 2.232 points.
Mishra et al. [11]	207 patients from India	Questionnaires, 9-item Hill Bone scale, univariate and multivariate binary logistic regression	Type of hospital, physical activity, smoking, education, diet, anxiety	37.69% were found to be nonadherent. Subjects from nuclear families belonging to castes or tribes and having social insurance had a higher risk of nonadherence. People visiting only public hospitals were more adherent. Lack of physical activity, not eating properly and anxiety could contribute to nonadherence. Smoking increased nonadherence.
Shiyanbola et al. [18]	174 patients from Midwestern States in USA	Questionnaires, MMAS-8, cluster analysis	Adherence: Not concerned about their illness, believe that their medications are necessary, not concerned about the medicine, feel confident that they can take their medications correctly, understand health information	Four clusters were made to group the patients in adherent and nonadherent clusters.
Jimmy et al. [19]	192 patients from Al Dakhiliyah Governorate of Oman	Mann-Whitney U, Kruskal-Wallis, questionnaires	Nonadherence: Forgetfulness	Forgetfulness was the most frequent reason for nonadherence (36.4%). No significant difference was observed for gender, age, education, or employment status.
Zongo et al. [61]	153 patients from Diabète Québec	SR-4, MMAS-8, self-reported proportion of pills, linear regression analysis	-	Results from SR-4 and MMAS-8 were significantly associated with HbA1c levels.

Lee et al. [15]	382 patients from north-eastern, Singapore	Questionnaires, MARS-5, statistical analysis, logistic regression	Adherence: Older, married, or widowed, assisted by a family member or domestic helper in taking medication, taking five or more daily medications. Nonadherence: Poor glycaemic control, Chinese ethnicity, younger, taking medication on their own, taking fewer than four daily medications	The median MARS-5 score was 24. Logistic regression found that younger, Chinese patients with poorer glycemic control were associated with low medication adherence. 57.1% of the population had low medication adherence to at least one of their medications

Table 3.3. Description of included articles listed by citation, participants, methods, features for adherence, and main results. Abbreviations: ARMS-D: Adherence to Refills and Medications Scale for Diabetes, BHLS: Brief Health Literacy Screen, BMQ: Beliefs about Medicines Questionnaire, CES-D: Centre for Epidemiologic Studies Depression Scale, CNN: Convolutional Neural Network, CVD: Cardiovascular disease, DAI-10: Drug Attitude Inventory questionnaire, DMSES: Diabetes Management Self-efficacy Scale, EPQ-RS: Neuroticism subscale of the Eysenck Personality Questionnaire-Revised Short Scale, EQ-5D: EuroQol-5 Domain, HAM-A: Hamilton anxiety scale, HAM-D Hamilton depression scale, HAPA: Health Action Process Approach, HRQoL: Health-related Quality of Life, IAQDM: Insulin Adherence Questionnaire for Diabetes Mellitus, KS-test: Kolmogorov-Smirnov test, MARS: Medication Adherence Report Scale, MCQ: Medication compliance questionnaire, MDKT-U: Michigan Diabetes Knowledge Test, MEDAS: Mediterranean Diet Adherence Screener, MMAS: Morisky Medication Adherence Scale, MPR: Medication Possession Ratio, MSPSS: Multidimensional Scale of Perceived Social Support, MVP: Medtronic virtual patient, PCA: Principal Component Analysis, PDC: Proportion of Days Covered, PHQ-8: Patient Health Questionnaire, PPP: Patient Prioritized Planning, SDSCA-MS: Summary of Diabetes Self-Care Activities medications subscale, SR-4: 4-item self-report

Methods 4

To establish a machine learning model to explore medication adherence, data was acquired from The Diabetes teleMonitoring of Patients in Insulin Therapy (DiaMonT) trial. DiaMonT is a 3-month open-label randomized controlled trial that aims to explore the influence of telemonitoring in T2D patients on insulin therapy. The second objective of the DiaMonT trial is to collect data for developing different algorithms for dose guidance and prediction of adverse events in T2D patients. The trial is conducted at two sites: Steno Diabetes Center North Denmark at Aalborg University Hospital and Steno Diabetes Center Zealand at Nykøbing Falster Hospital, in collaboration with Novo Nordisk and Glookoo. The protocol for the trial can be found at ClinicalTrials.gov with the identifier NCT04981808. [62]

All participants in DiaMonT were diagnosed with T2D and were already on basal or basal and bolus insulin therapy. During the trial, all participants were treated with insulin products provided by Novo Nordisk, which were free of charge. The participants were selected for the trial if they met the following criteria: [62]

Inclusion criteria:

- Women and men 18 years or above
- Diagnosis of T2D for at least 12 months
- Patients from the North Denmark Region or Region Zealand in treatment with insulin
- Being able to use a smartphone along with the other devices used in the trial
- Ability to understand and read Danish

Exclusion criteria:

- Pregnancy or breastfeeding
- Major surgery planned during the trial period
- Cancer diagnosis within five years before inclusion
- Participation in other trials
- Terms that, in the opinion of the sub-investigator or investigator, render the participant unfit to conduct the trial, including a lack of understanding of the trial or a lack of physical or cognitive ability to participate

The participants in DiaMonT were divided into intervention and control groups, where the intervention group receives telemonitoring and the control group continues usual care. The participants and the clinical staff involved in the trial were not blinded, as they knew whether or not a patient received telemonitoring. However, the control group received a blinded Novopen 6 insulin smart pen, so they could not see their data during the trial. The control group was also provided with a blinded Dexcom G6 continuous glucose monitor (CGM), which should be worn for the first and last 20±3 days of the trial. Likewise, the intervention group received a

Dexcom G6, a Novovpen 6, a Fitbit Charge 4, and a smartphone. They used the devices to continuously collect, log, and transfer interstitial glucose levels, insulin administration, activity, and sleep during the trial period. The participants in the intervention group were contacted at least three times throughout the trial period to ensure that they followed the instructions. [62]

The interstitial glucose was measured every five minutes, the pulse was tracked every five seconds, and the number of steps within a minute was collected every minute. The insulin use of each participant was collected every time they used the smart pen. The clinicians could change the prescribed insulin for the intervention group if needed through telemonitoring whereas the control group followed their usual care. Therefore, the data for prescribed insulin was only collected when the clinician made an edit in dosage. Furthermore, health status, medications, comorbidities, and basic information such as age, duration of diabetes, smoking, etc. were obtained through questionnaires. This information was collected to ensure suitability concerning the inclusion criteria and for later data analysis. The trial strived to include 400 participants with 200 in each group. However, only 331 participants were included in the final trial with 165 in the control and 166 in the intervention group. [62]

4.1 Preprocessing of data

The acquired data was loaded into Python version 3.10.12. The files contained data from 331 patients, but only 279 were included in this study due to the exclusion criteria illustrated in Figure 4.1.

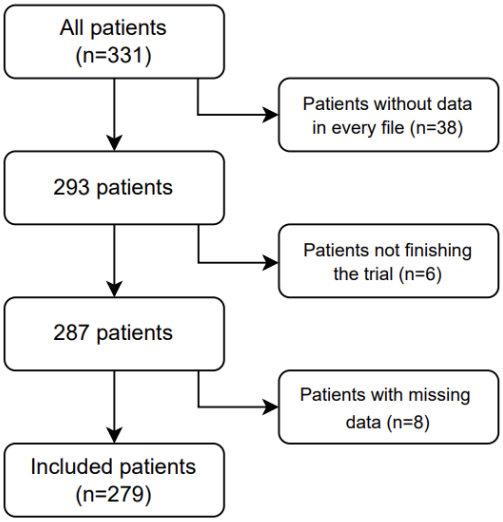


Figure 4.1. Segmentation of data with n being the number of patients.

The data was obtained with timestamps in seconds measured from the beginning of the trial period. The timestamps were converted to integer values, representing days from inclusion. This was done by dividing all the original timestamps by 86,400, equal to one day in seconds. As basal insulin was not prescribed at a specific time of injection, the participants took the medication at different times of the day. To ensure that the injected insulin was assigned to the correct day, the time of injection was explored among all participants. The distribution of injected basal insulin on an hourly basis from 00:00 to 23:59 can be seen in Figure 4.2.

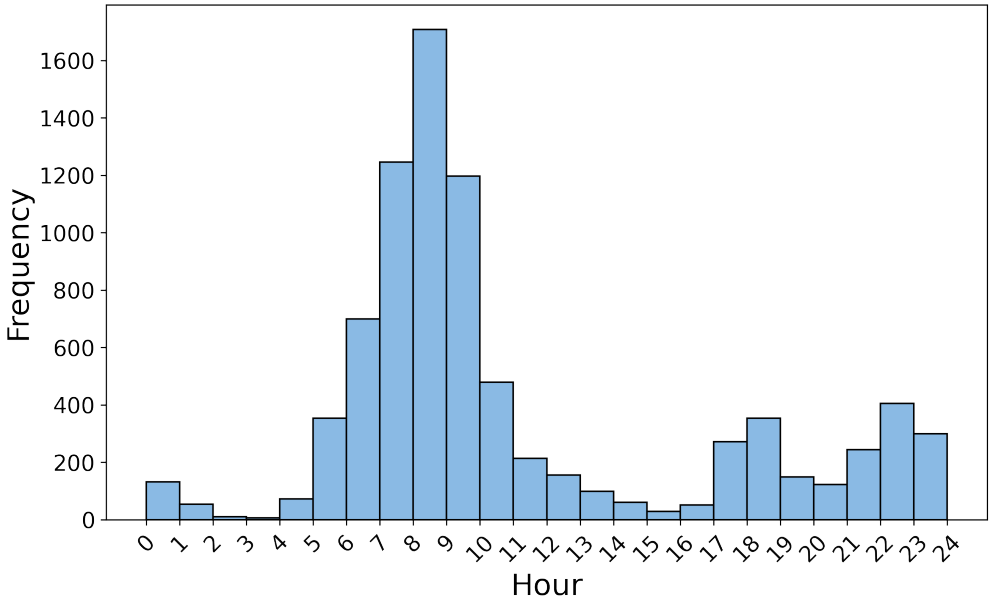


Figure 4.2. Histogram showing the distribution of taken basal insulin for all participants from day eight to day 21.

As seen on the histogram, the hourly period with the fewest injections was between 03:00 and 04:00. Therefore, the days were separated at 3:00 am to reduce the number of misplaced injections.

4.1.1 Features

As this project aimed to develop a machine learning model to classify adherent and nonadherent patients, different features were extracted from the data. The features were selected based on findings from relevant literature, described in Chapter 3, and knowledge regarding the topic. In total, 43 features were extracted, as seen in Table 4.1. The features are presented by feature name, description, and value type.

Feature	Description	Value type
Age	Age is the lowest number in the age intervals, e.g. 20-29	Integer: 20, 30, 40, 50, 60, 70, 80
Telemonitored	Is the patient telemonitored or not	Binary
HbA1c	HbA1c levels at inclusion date	Integer: 40-140
Glycemic variability%	Glycemic variability in percentage	Integer: 9.64 - 51.02
CGM variance	Variance of the CGM data in the first seven days	Float: 0.6-26.37
CGM mean	Mean of the CGM data in the first seven days	Float: 5.76-20.99
CGM max	Maximum value of the CGM data in the first seven days	Float: 9.6-22.26
CGM min	Minimum value of the CGM data in the first seven days	Float: 2.16-12.1
Number of hypoglycemic events	Number of hypoglycemic events in the last seven days	Integer: 0-17
Time below range	Minutes below 3.9 mmol/L	Integer: 0-1000
Presence of hypoglycemia	Occurrence of hypoglycemia in the last seven days	Binary
Number of hyperglycemic events	Number of hyperglycemic events in the last seven days	Integer: 0-35
Time above range	Minutes above 10 mmol/L	Integer: 0-9275
Presence of hyperglycemia	Occurrence of hyperglycemic events in the last seven days	Binary
Time in range	Minutes in normal range (3.9-10 mmol/L)	Integer: 0-9200
Height	Height (cm)	Float: 151.5-202.5
Weight	Weight (kg)	Float: 52.6-181.0
BMI	Body Mass Index	Float: 19.8-57.38
Mean arterial pressure	Mean arterial pressure (mmHg)	Float: 78.5-156.0
Diastolic	Diastolic pressure (mmHg)	Float: 51-117
Systolic	Systolic pressure (mmHg)	Float: 98-200
Emotional stage	What emotional stage is the patient in	Integer: 1-6
Hypoglycemia experience	Ever experienced a hypoglycemic event	Binary
Minimum one diabetes complication	Having one or more diabetes-related complications	Binary
Sum of diabetes complications	Diabetes related complications sum	Integer: 0-5
Insulin type	Basal or basal-bolus	Binary
Sum of anti-diabetic medication	Anti-diabetic medications sum	Integer: 0-4
Sum of comorbidities	Comorbidities sum	Integer: 0-5
Overweight	Is the patient overweight	Binary
Hypertension	Has the patient hypertension	Binary
Cardiovascular disease	Has the patient cardiovascular disease	Binary
Hyperlipidaemia	Has the patient hyperlipidaemia	Binary
Minimum one additional medication	Taking one or more additional non-diabetic medications	Binary
Sum of other medications	Non-diabetic medications sum	Integer: 0-6
Hours of exercise	Do the patient exercise for more than five hours a week	Binary
Marital status	Are the patient living alone or with a partner	Binary
Education	Five different education levels	One hot encoded
Health status	Self-reported overall health	Binary
Sadness	Time felt sad in the last four weeks	Integer: 1-6

Table 4.1. Description of the features used in the models and their value type. The range was added if the value type was an integer or a float.

The value type differed among the features as the data was acquired from questionnaires, demographics, and measuring equipment. The majority of value types were integers, binary values, and floats. The binary values represented 0 or 1, referring to yes or no. The feature for *Education* was one-hot encoded as it contained categorical parameters. In one-hot encoding, the categorical parameters are transformed into separate binary values, representing each parameter.

This ensures that the feature can be input for the machine learning models. Additionally, this will secure a more transparent model output.

4.1.2 Definition of adherence

Adherence was defined based on the injected basal insulin obtained from the smart pen and the prescribed insulin. As the definition of medication adherence remains unclear, different adherence levels and definitions were investigated. This was done with different machine learning algorithms to explore performances, as seen in Table 4.2.

	≥% deviation mean			**≥20% days of ≥% deviation**		
Models	**5%**	**10%**	**20%**	**5%**	**10%**	**20%**
n = nonadherent patients	**n = 158**	**n = 101**	**n = 52**	**n= 97**	**n = 71**	**n = 47**
Logistic Regression (Intervention + control)	0.66	0.69	0.68	0.69	0.68	0.74
Logistic Regression (Intervention)	0.74	0.79	0.70	0.76	0.80	0.83
Logistic Regression (Control)	0.63	0.71	0.65	0.59	0.66	0.68
Linear Discriminant Analysis (Intervention + control)	0.68	0.69	0.70	0.70	0.70	0.76
Linear Discriminant Analysis (Intervention)	0.76	0.78	0.78	0.77	0.84	0.90
Linear Discriminant Analysis (Control)	0.67	0.72	0.75	0.67	0.66	0.72
Random Forest (Intervention + control)	0.74	0.71	0.74	0.75	0.77	0.82
Random Forest (Intervention)	0.73	0.77	0.87	0.85	0.85	0.94
Random Forest (Control)	0.75	0.80	0.71	0.74	0.74	0.81
Support Vector Machine (Intervention+control)	0.64	0.63	0.44	0.68	0.61	0.68
Support Vector Machine (Intervention)	0.67	0.67	0.67	0.74	0.81	0.82
Support Vector Machine (Control)	0.54	0.67	0.65	0.64	0.43	0.53

Table 4.2. Machine learning models' performances scored on ROC-AUC using different adherence definitions. The first three columns represent machine learning models' performances from an adherence definition based on the mean deviation. The last three columns represent machine learning performances where adherence is defined each day after which the overall adherence level is calculated.

The best performances were seen at >20% days of ≥20% deviation, but this definition would only identify patients with severe nonadherence. However, this study aims to find all nonadherent patients. Therefore, the definition of >20% days with a deviation of ≥5% was chosen. The 5% daily deviation was allowed due to the subjective administration of the insulin, which might

cause small errors. The injected and the prescribed insulin were compared daily for each patient, and the percentage deviation was calculated. This was done from day eight to day 21, where an injected insulin deviation of more than ±5% from the prescribed insulin defined the patients as nonadherent on that specific day. The first week was not used due to insecurities as the patients had to get used to new treatment methods and equipment. The sum of the nonadherent days was calculated to get an overall adherence level for each patient. If the sum of the nonadherent days exceeded 20% of the period, they were defined as overall nonadherent. An example can be seen in Figure 4.3 where *Patient 1* is adherent and *Patient 2* is nonadherent. The machine learning algorithms used the adherence and nonadherence labels in the training.

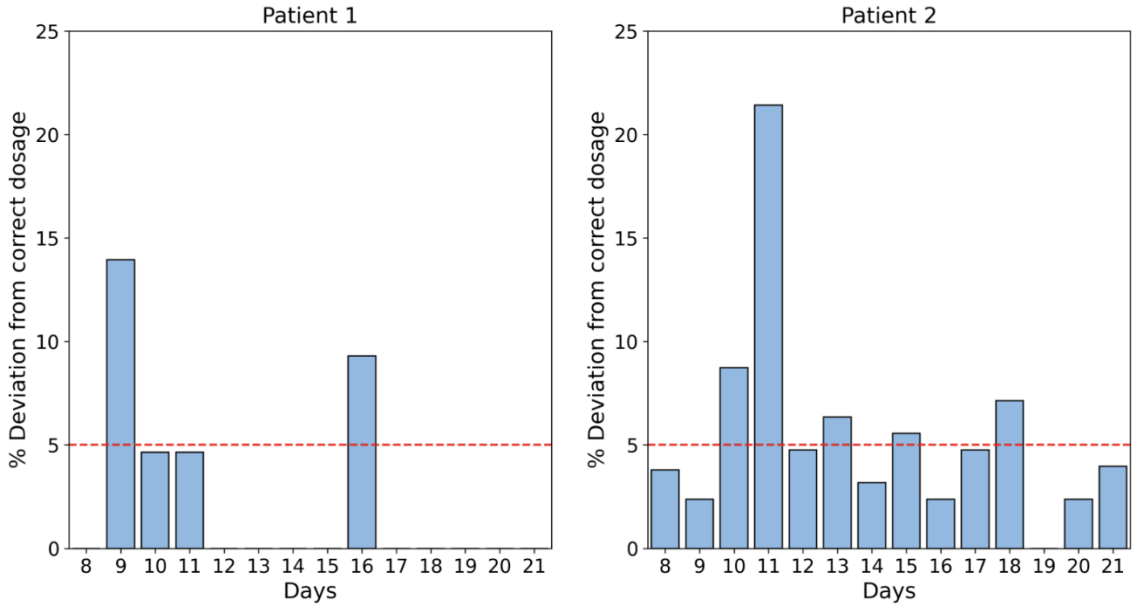


Figure 4.3. Percentage deviation from the correct dosage for two patients. *Patient 1* is adherent as the patient has <20% of the days above 5% deviation whereas *Patient 2* is nonadherent as the patient has >20% of the days above 5% deviation. The 5% deviation line is represented with red.

Adherence was defined based on data from day eight to day 21 to develop a machine learning algorithm that could detect nonadherence based on data from a short period. In clinical practice, this means that patients will have the opportunity to get used to their new medications and equipment for three weeks before a follow-up appointment. At this follow-up appointment, the clinician can use the data obtained in the last three weeks to decide whether the patient is adherent or nonadherent using the machine learning model.

4.1.3 Data extraction

Data used in this project was extracted from different time points during the trial period. Figure 4.4 visualizes the data extraction timeline. It illustrates the trial period of 90 days where data from the questionnaires, demographics, and HbA1C were collected on day one. As mentioned earlier, CGM was measured during the entire trial for the intervention group and the first and the last 20±3 days for the control group. However, this project only used CGM data from day one to day seven, as the intention was to characterize and classify patients at baseline.

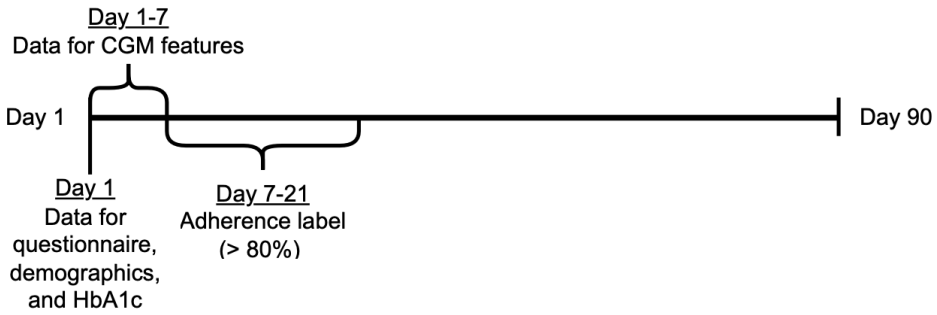


Figure 4.4. Timeline of data extraction to develop features and define adherence.

4.2 Machine learning

In this study, machine learning was used to identify adherent and nonadherent patients. Machine learning is a statistical tool that has shown promising results in finding relationships and patterns in simple and complex data. The main idea of machine learning is to transform input data into meaningful outputs. There are two main types of machine learning: Supervised and unsupervised. Supervised machine learning is when a model is trained on labeled data. In contrast, unsupervised machine learning is when the input data to the model does not have a label, which forces it to find hidden patterns in the data. Machine learning can be used for either regression or classification problems. Regression is used if the label is numeric, whereas classification is used to divide data into specific groups, e.g. sick and healthy. [63], [64], [65]

In this study, supervised machine learning was chosen as the data was labeled and aimed to classify if the patients were adherent or nonadherent to their basal insulin. Therefore, eight different types of supervised classification machine learning algorithms were tested in this study. The models were built using scikit-learn in Python. The machine learning algorithms were Linear Discriminant Analysis (LDA), Logistic Regression (LR), Support Vector Machine (SVM), Naïve Bayes (NB), K-Nearest Neighbour (KNN), Random Forest (RF), Extreme Gradient Boosting (XGB), and a Multi-layer Perceptron (MLP).

4.2.1 Cross-validation

Training and evaluation of machine learning models are important steps to achieving a model with good performance. Here, there is a need for reliable and unbiased approaches. This study used internal evaluation, where splitting and resampling methods are widely used. Cross-validation is a resampling method that trains and evaluates the entire dataset by making multiple random splits, resulting in different subsets. The model trains on K-1 subsets and tests on the last subset. This is done numerous times with different splits, resulting in an average cross-validation score that combines the various performances. Instead, data splitting is a method where the dataset is split into two groups. The models are trained and validated based on data in the first group and tested on data in the second group, resulting in one performance score. However, a recent study by Collins et al. [66] found that splitting might implement overfitting and bias due to limited data and the opportunity to affect data. Instead, cross-validation can reduce bias and performance variability while using all available data.

Therefore, this study used a 5-fold cross-validation to evaluate the performance of all machine

learning models. In 5-fold cross-validation, the dataset is split into five subsets, where the model is trained on four subsets and tested on one subset. This is done five times, and the model is tested on a different subset each time. The cross-validation score is therefore an average of five independent performances.

4.2.2 Feature selection

Feature selection was implemented to reduce input dimensions and remove redundant information. Implementing feature selection can reduce the computational cost and improve performance as additional non-relevant input variables are removed. In this study, mlxtend's sequential forward feature selection was used. Sequential forward feature selection is a wrapper method that creates numerous models with different subsets of features. It works by developing a model that tests the performance of all individual features, where the highest-performing feature based on cross-validation is selected. The model then uses the highest-performing feature and tests it in combination with all additional features individually. The model then chooses the subset of features that results in the highest cross-validation score. The addition of features continues until all features have been iterated through. [67]

To prevent overfitting, it can be beneficial to implement tolerance as a stopping criterion in feature selection. By implementing a tolerance, the feature selection will stop when the cross-validation score does not increase by a certain percentage. Therefore, different tolerances were tested to find an optimal tolerance. This was done by investigating the number of selected features and the ROC-AUC score. The test can be seen in Table 4.3.

	Tolerance = 0.01		**Tolerance = 0.001**		**No tolerance**	
Models	**ROC-AUC**	**Features, n**	**ROC-AUC**	**Features, n**	**ROC-AUC**	**Features, n**
LR	0.662	2	0.682	6	0.685	9
RF	0.740	5	0.746	6	0.756	9
SVM	0.636	2	0.636	2	0.699	15
KNN	0.683	2	0.686	3	0.695	9
NB	0.687	3	0.707	7	0.713	16
XGB	0.734	6	0.734	6	0.764	17
LDA	0.664	2	0.691	7	0.692	8
MLP	0.653	2	0.683	6	0.692	9

Table 4.3. Test of different tolerance levels for all machine learning algorithms. The ROC-AUC score and the selected features are visualized at each tolerance level for each algorithm.

The highest ROC-AUC scores can be seen when no tolerance is implemented in the feature selection. This means that the model chooses the number of features that result in the highest ROC-AUC score. However, having no tolerance results in more features, which might overfit the model. An example can be seen in Figure 4.5, where the green dot illustrates the number of features that result in the highest performance.

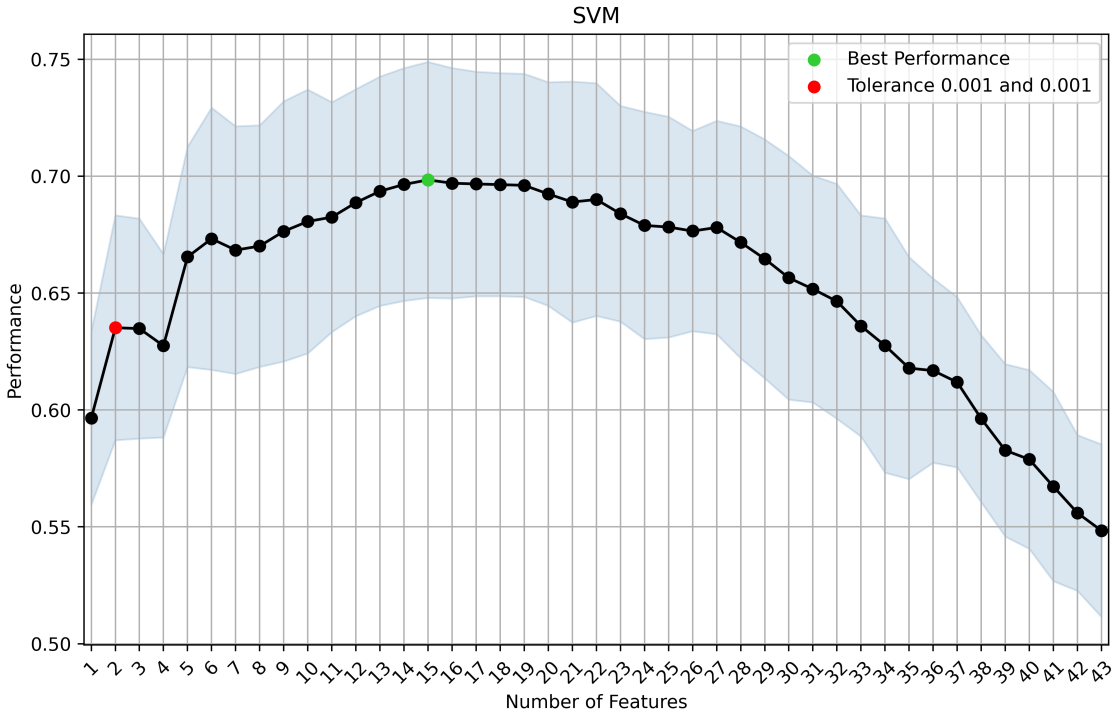


Figure 4.5. Example of feature selection for Support Vector Machine with best performance as green and tolerances of 0.01 and 0.001 as red.

However, it can be seen that the surrounding number of features results in similar performances. Therefore, choosing the number of features with the best performance might incorporate overfitting. Consequently, a tolerance is often implemented to reduce overfitting. However, as seen with the red dot in Figure 4.5, a tolerance may stop the feature selection too early due to a small decrease in performance.

From a clinical perspective, too many unnecessary features would increase the workload as more data is needed to detect nonadherence using a model. Therefore, this study implemented a tolerance of 0.001 to reduce the number of features and potential overfitting, as it has the best trade-off between the number of features and the ROC-AUC score. This means that the sequential forward feature selection stops when adding features, which does not increase the model performance by more than 0.1%. Sequential forward feature selection was implemented in all the developed machine learning models.

4.2.3 Hyperparameter optimization

The machine learning algorithms have different hyperparameters that can be tuned to increase performance. Therefore, various combinations of hyperparameters are tested to boost the performance further in detecting nonadherence. Hyperparameter optimization was done using a grid search, where all possible combinations of hyperparameters were tested. The grid search chose the combination of hyperparameters that yielded the highest cross-validation score. Hyperparameter optimization with grid search was performed on all machine learning models.

4.2.4 Model evaluation

All developed models were evaluated to quantify the individual performances and find the best-performing model. In this study, ROC-AUC was used as a scoring parameter in the cross-validation. A ROC curve shows the performance of a classifier, where the true positive rate (TPR) is plotted against the false positive rate (FPR). TPR is the probability that a nonadherent patient is correctly classified as nonadherent, whereas FPR is the probability that an adherent patient is incorrectly classified as nonadherent. Therefore, the curve estimates how well the model classifies at different thresholds, which can be seen as the ROC-AUC values.

Additionally, positive predictive value (PPV) and negative predictive value (NPV) were calculated to evaluate the performance of the models even further. PPV indicates how many are correctly classified as nonadherent out of all people classified as nonadherent. Likewise, NPV indicates how many are correctly classified as adherent out of all those classified as adherent. The calculations for PPV and NPV are as follows:

$$PPV = \frac{TruePositives}{TruePositives + FalsePositives} \tag{4.1}$$

$$NPV = \frac{TrueNegatives}{TrueNegatives + FalseNegatives} \tag{4.2}$$

To calculate PPV and NPV, different sensitivities with an interval of 0.1, starting from 0.5, were fixed on the ROC-AUC curve. The values of PPV and NPV can be used in a clinical setting to find the right balance between misdiagnosed and correctly diagnosed patients.

Furthermore, permutation feature importance was used to evaluate the impact of the features selected in the feature selection. It measures the importance of a feature by looking at the decrease in performance when the feature values are randomly shuffled. The reduction in performance was calculated using equation 4.3.

$$PI_X = Error_{base} - Error_{Xshuffled} \tag{4.3}$$

PI_X is the permutation feature importance for feature X, $Error\_base$ is the model error calculated using the original dataset, and $Error_X shuffled$ is the model error of the dataset with shuffled values for feature X. Therefore, the importance of a feature increases when the difference gets higher. [68]

Lastly, SHapley Additive exPlanations (SHAP) values were calculated to investigate potential tendencies in the data. SHAP values give transparency to the machine learning models, as they visualize the impact each observation has on the performance of the models. Tendencies in data can be revealed, as it can be seen if high or low values of the selected features impact the performance positively or negatively. [69]

Results 5

5.1 Characteristics

Table 5.2 was made to get an overview of the features and characteristics of the population, adherent, and nonadherent groups. Each feature and characteristic was shown with either ranging or numerical values. The ranging values were abbreviated to two decimals and shown with the standard deviation. The numerical values were the sum of all positive answers in each feature; for example, *living alone* was the sum of all participants who answered that they lived alone at the beginning of the trial. In addition, the numerical values were shown with the percentage share of how much the positive answers constituted within the group.

A statistical analysis was made for each feature to examine if there was a significant difference between the adherent and nonadherent groups. To do so, all features were tested for normal distribution using a Shapiro-Wilk test. If the data was normally distributed, a 2-sample t-test was made, and if not, a Mann-Whitney U test was made. Both tests were made with a significance level of <0.05. The results can be seen in the rightmost column in Table 5.2.

Feature	**Total (n = 279)**	**Adherent (n = 182)**	**Nonadherent (n = 97)**	**p-value**
Demographic data				
Age, years	57.20 ± 11.26	57.25 ± 11.13	57.11 ± 11.54	0.755
Height, cm	173.69 ± 8.97	172.65 ± 9.24	175.64 ± 8.13	0.008†*
Weight, kg	100.10 ± 21.18	99.18 ± 20.83	101.84 ± 21.84	0.400
BMI	33.19 ± 6.60	33.30 ± 6.61	32.98 ± 6.60	0.673
Living alone, n	78 (27.96%)	48 (26.37%)	30 (30.93%)	0.421
Handyman, n	92 (32.97%)	56 (30.77%)	36 (37.11%)	0.284
Primary school, n	35 (12.54%)	25 (13.74%)	10 (10.31%)	0.412
Highschool, n	20 (7.17%)	11 (6.04%)	9 (9.28%)	0.319
Medium education, n	112 (40.14%)	76 (41.76%)	36 (37.11%)	0.449
Long education, n	20 (7.17%)	14 (7.69%)	6 (6.19%)	0.644
CGM data				
CGM variance, mmol/L	7.25 ± 4.42	6.82 ± 4.18	8.03 ± 4.74	0.022*
CGM mean, mmol/L	10.04 ± 2.49	9.69 ± 2.11	10.69 ± 2.98	0.010*
CGM max, mmol/L	17.98 ± 3.20	17.63 ± 3.12	18.63 ± 3.27	0.009*
CGM min, mmol/L	4.64 ± 1.73	4.51 ± 1.50	4.89 ± 2.07	0.187
Glycemic variability, %	25.94 ± 6.14	25.85 ± 6.07	26.1 ± 6.29	0.800
Hypoglycemic events in the last seven days, n	0.86 ± 2.03	0.77 ± 1.75	1.02 ± 2.47	0.818

Presence of hypoglycemia during the last seven days, n	78 (28.05%)	53 (29.12%)	25 (25.77%)	0.554
Number of hyperglycemic events in the last seven days, n	14.38 ± 6.93	14.73 ± 6.90	13.72 ± 6.98	0.318
Presence of hyperglycemia during the last seven days, n	278 (99.6%)	181 (99.45%)	97 (10%)	0.469
Time above range (>10 mmol/L), minutes	3756.77 ± 2487.35	3457.66 ± 2366.0	4317.99 ± 2621.81	0.009*
Time in range (3.9-10 mmol/L), minutes	5077.22 ± 2489.03	5438.54 ± 2406.13	4399.28 ± 2512.57	0.001*
Time below range (<3.9 mmol/L), minutes	45.84 ± 122.52	37.58 ± 95.96	61.34 ± 160.42	0.304
Laboratory data				
Telemonitored, n	145 (51.9%)	108 (59.34%)	37 (38.14%)	0.001*
HbA1c, mmol/mol	64.04 ± 14.22	61.93 ± 12.0	67.99 ± 17.04	0.006*
Diastolic, mmHg	81.61 ± 11.08	81.35 ± 10.51	82.08 ± 12.12	0.600†
Systolic, mmHg	138.11 ± 17.35	137.67 ± 17.10	138.94 ± 17.85	0.519
Mean arterial pressure, mmHg	100.44 ± 11.51	100.12 ± 10.86	101.03 ± 12.70	0.531
Diabetes-related information				
Minimum one diabetes-related complication, n	174 (62.4%)	119 (65.38%)	55 (56.70%)	0.155
Sum of diabetes-related complications, n	1.08 ± 1.13	1.09 ± 1.1	1.04 ± 1.19	0.427
Insulin type (basal + bolus), n	117 (42.04%)	81 (45.05%)	36 (37.11%)	0.235
Sum of anti-diabetic medication, n	1.68 ± 0.93	1.69 ± 0.95	1.66 ± 0.88	0.691
Ever experienced hypoglycemia, n	191 (68.5%)	124 (68.13%)	67 (69.07%)	0.873
Minimum one additional medication, n	265 (95.08%)	173 (95.05%)	92 (94.85%)	0.951
Sum of other medications, n	2.32 ± 1.03	2.28 ± 1.04	2.38 ± 1.0	0.289
Comorbidities				
Number of comorbidities, n	2.68 ± 0.95	2.66 ± 0.94	2.72 ± 0.98	0.443
Overweight, n	226 (81%)	148 (81.32%)	78 (80.41%)	0.855
Hypertension, n	220 (78.9%)	145 (79.67%)	75 (77.32%)	0.648
Cardiovascular disease, n	86 (30.82%)	54 (29.67%)	32 (32.99%)	0.569
Hyperlipidaemia, n	216 (77.42%)	137 (75.27%)	79 (81.44%)	0.239
Physical and mental data				
>5 hours of exercise, n	89 (31.9%)	61 (33.52%)	28 (28.87%)	0.429
Overall good health status, n	202 (72.4%)	135 (74.18%)	67 (69.07%)	0.369
Emotional state, (1; negative - 6; positive)	3.23 ± 1.43	3.2 ± 1.45	3.28 ± 1.40	0.739
Sadness within last four weeks, (1; all time - 6; at no time)	5.07 ± 1.08	5.17 ± 0.98	4.88 ± 1.22	0.081

Table 5.2. Features and characteristics of all participants divided into total, adherent, and nonadherent groups. The binary features are shown with the number of positive answers and the percentage share of participants in the group. The integers and floats are shown with the mean ± standard deviation. Mann Whitney U test with $p<0.05$ is performed on all features except the ones listed with †, where a 2-sample t-test with $p<0.05$ is performed. All features that are significantly different between adherent and nonadherent groups are listed with an *.

The results show that eight features significantly differ between the adherent and nonadherent groups. Nonadherent patients were significantly taller (p = 0.008) than the adherent group. Furthermore, it can be seen that 59.34% of the telemonitored patients were in the adherent group, whereas 38.14% of the telemonitored patients were in the nonadherent group. This significant difference (p = 0.001) indicates that patients are more adherent when they are telemonitored. HbA1c is a measure of how well the regulation of blood glucose has been in the previous two to three months. A high HbA1c value indicates poor regulation of blood glucose. In this study, HbA1c was a feature that significantly differs (p = 0.006) between the adherent and nonadherent groups. Nonadherent patients had a higher mean value of 67.99 mmol/mol, whereas adherent patients had a lower mean value of 61.93 mmol/mol. This corresponded with the features *CGM variance*, *CGM mean*, *CGM max*, *Time above range*, and *Time in range* also being significantly different between the adherent and nonadherent groups. The values and standard deviation of the features *CGM variance*, *CGM mean*, *CGM max*, and *Time above range* were significantly higher in the nonadherent group, whereas the values of *Time in range* were significantly lower for the nonadherent group compared to the adherent group.

5.2 Sequential forward feature selection

Sequential forward feature selection was used in the development of machine learning algorithms to reduce overfitting and improve performance. A test was made to investigate different tolerance levels, which is described in Section 4.2.2. Based on the results, a tolerance of 0.001 was implemented in the feature selection. The features that resulted in the highest ROC-AUC score for each model, can be seen in Table 5.3.

Model	ROC-AUC	Feature names
LR	0.683	Time below range, Time in range, Height, Telemonitored, HbA1c, Sadness
RF	0.746	Time in range, Systolic, Health status, Insulin type, Telemonitored, HbA1c
KNN	0.686	CGM mean, Presence of hypoglycemia, Time in range
XGB	0.734	CGM min, CGM min, Time below range, Time in range, Mean arterial pressure, Sum of diabetes-related complications, Minimum one diabetes complication
NB	0.707	CGM mean, Time below range, Time in range, Height, Insulin type, Telemonitored, Sadness
MLP	0.684	CGM max, Number of hypoglycemic events, Minimum one diabetes complication, Telemonitored, HbA1c, Sadness
LDA	0.692	Time below range, Time in range, Sum of other medications, Hyperlipidaemia, Insulin type, Telemonitored, HbA1c
SVM	0.636	CGM mean, Height

Table 5.3. Selected features resulted in the highest ROC-AUC score for each model when a tolerance of 0.001 was implemented.

The feature *Time in range* was chosen in six of the eight models, whereas the feature *HbA1c* was selected in four of the eight models. Table 5.2 shows these were also significantly different (p = 0.001, p = 0.006) between the adherent and nonadherent groups. This indicates that these features strongly predict nonadherence in people with T2D. *Telemonitored* was chosen in five of the eight models, implying that telemonitoring may affect adherence. However, it can be seen that most of the features were based on CGM data, implying that CGM data is essential in identifying nonadherence.

5.3 Hyperparameter optimization

Different hyperparameter combinations were tested for each model to boost the performance further. The ROC-AUC scores and the tested hyperparameters can be seen in Table 5.4.

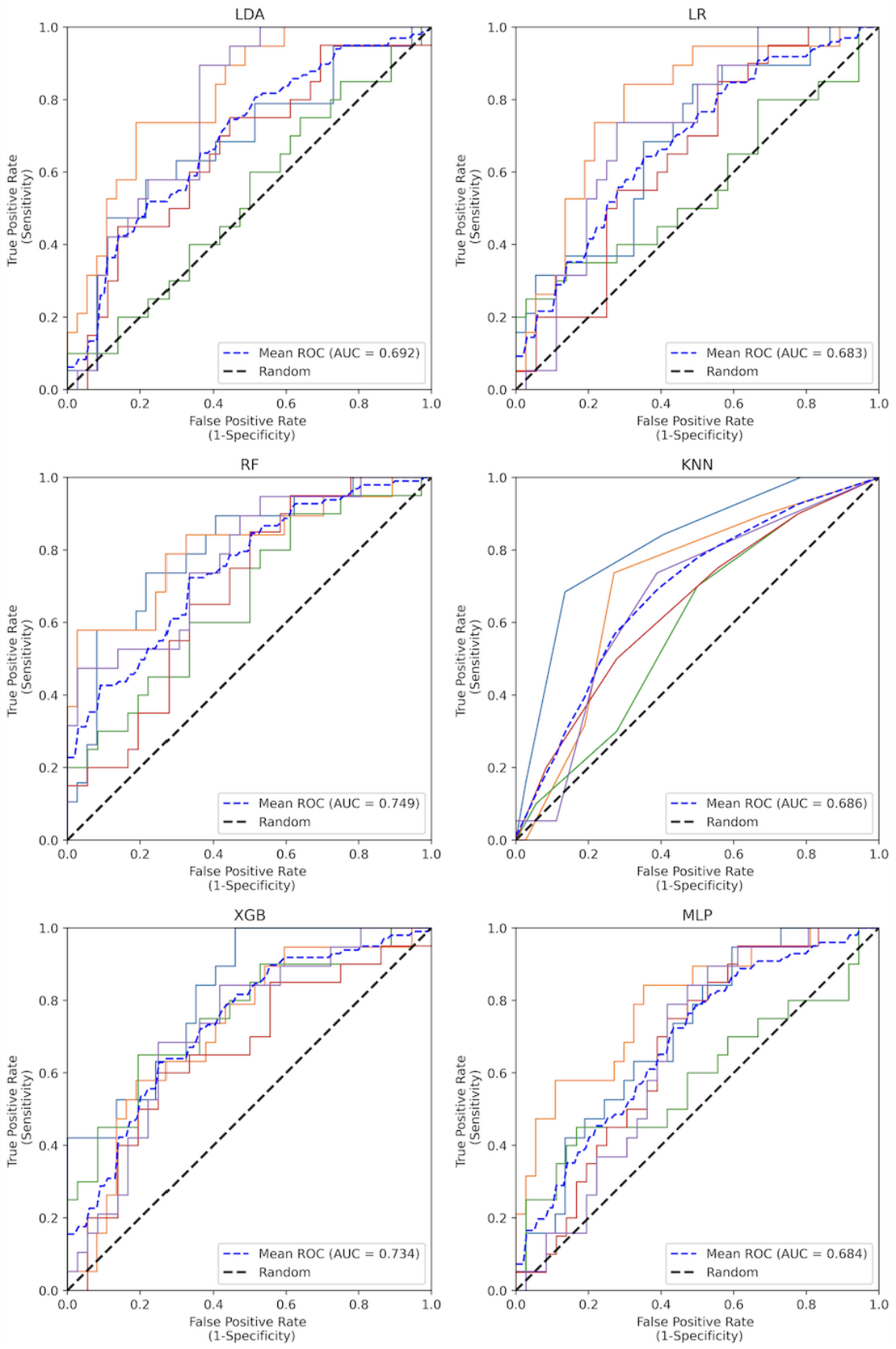
Model	ROC-AUC	Parameters	Attempted values
LR	0.683	solver	newton-cg, liblinear, newton-cholesky, saga, sag, **lbfgs**
		penalty	none, **l2**, l1, elasticnet
		C	0, 0.01, 0.1, 0.2, 0.4, 0.6, 0.8, **1.0**, 1.2, 1.4, 10, 100
SVM	0.636	kernel	**linear**, rbf, sigmoid, poly
		gamma	auto, **scale**
		C	0.1, 0.2, **1.0**, 1.5, 2.0, 5.0
RF	0.749	max_depth	5, 6, 7, 8, 9, **10**, 15, 30, 50
		min_samples_leaf	1-16 **(1)**
		n_estimators	50, **100**, 200, 300, 1000
LDA	0.692	solver	**svd**, lsqr, eigen
KNN	0.686	n_neighbours	1-30 **(5)**
		weights	**uniform**, distance
		algorithm	**auto**, ball_tree, kd_tree, brute
		leaf size	20, **30**, 50
		metric	Manhattan, euclidean, **minkowski**, mahalanobis
XGB	0.734	max_depth	**2**, 3, 4, 5, 6
		learning_rate	0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, **1.0**, 1.1, 1.2
		gamma	**0**, 1.0, 1.1, 1.25, 1.5, 2
		scale_pos_weight	0, 0.25, 0.5, **1.0**, 3.0
MLP	0.684	hidden_layer_sizes	(25,), (50,), (100), (200,), **(300,)**, (400,), (500,)
		activation	identity, logistic, tanh, **relu**
		learning_rate	**constant**, invscaling, adaptive
		learning_rate_init	0.0001, **0.001**, 0.005, 0.01, 0.05
NB	0.707	None	None

Table 5.4. Hyperparameter optimization for all models, where different settings of hyperparameters were tested in the algorithms using grid search. The ROC-AUC score and the best combination of hyperparameters are listed in the table. The selected hyperparameters are underlined and marked in bold.

As seen in Table 5.4, there is a variation of 0.113 in the ROC-AUC scores across all models. However, it can be seen that the three-based models provide higher ROC-AUC scores than the other models. The overall best performance can be seen using RF with a max depth of 10, 100 trees, and a minimum of 1 sample in the leaf node.

5.4 Model evaluation

A ROC-AUC curve was made for each model to evaluate the individual performances and show each fold's results in the cross-validations. Figure 5.1 shows an overview of all ROC-AUC curves.


LDA
Mean ROC (AUC = 0.692)
Random
LR
Mean ROC (AUC = 0.683)
Random
RF
Mean ROC (AUC = 0.749)
Random
KNN
Mean ROC (AUC = 0.686)
Random
XGB
Mean ROC (AUC = 0.734)
Random
MLP
Mean ROC (AUC = 0.684)
Random
True Positive Rate (Sensitivity)
False Positive Rate (1-Specificity)

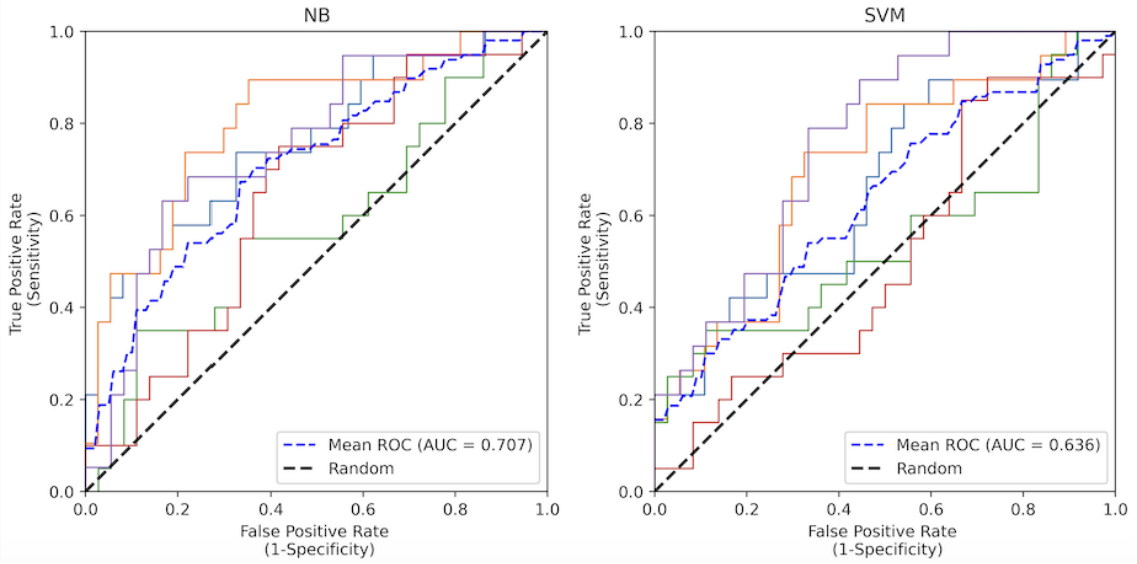


Figure 5.1. ROC-AUC curves for all models shown with each fold in the cross-validation. The blue dashed line represents the mean of the 5-fold cross-validation and the black dashed line represents a random classification.

Based on the different ROC-AUC curves, it can be seen that XGB and RF are more stable than the rest of the models. This means that the ROC-AUC values for each fold in the 5-fold cross-validation are more alike, which results in a more stable answer at each iteration. A comparison of the different mean ROC-AUC curves can be seen in Figure 5.2.

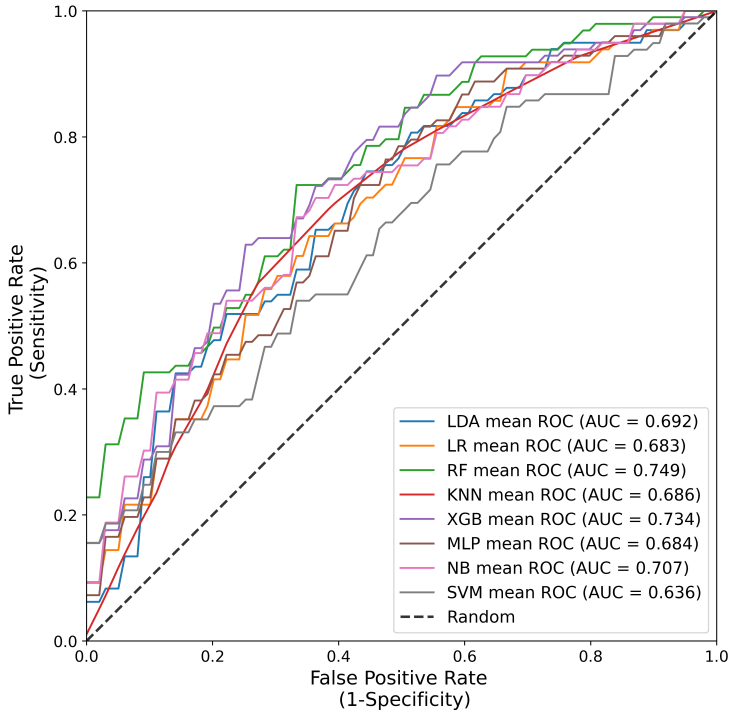


Figure 5.2. Mean ROC-AUC curves for all models based on the 5-fold cross-validation. The black dashed line represents a random classification.

To quantify the different performances and evaluate the clinical relevance, the sensitivity was fixed at different thresholds for each model. Here, the PPV, NPV, and specificity were calculated at each threshold. The results can be seen in Table 5.5 with fixed sensitivities at 0.5, 0.6, 0.7, 0.8, and 0.9.

	RF	XGB	NB	LDA	KNN	LR	MLP	SVM
Fixed sensitivity = 0.5								
Specificity	0.83	0.82	0.77	0.75	0.69	0.73	0.71	0.68
PPV	0.68	0.62	0.57	0.55	0.55	0.52	0.52	0.48
NPV	0.78	0.78	0.76	0.74	0.79	0.73	0.74	0.74
Fixed sensitivity = 0.6								
Specificity	0.71	0.76	0.69	0.66	0.63	0.65	0.63	0.58
PPV	0.56	0.59	0.54	0.53	0.53	0.51	0.48	0.45
NPV	0.80	0.80	0.77	0.80	0.80	0.76	0.76	0.73
Fixed sensitivity = 0.7								
Specificity	0.65	0.61	0.60	0.58	0.58	0.59	0.57	0.48
PPV	0.54	0.50	0.51	0.51	0.49	0.51	0.48	0.45
NPV	0.83	0.81	0.79	0.81	0.81	0.80	0.80	0.78
Fixed sensitivity = 0.8								
Specificity	0.57	0.54	0.46	0.43	0.32	0.50	0.49	0.42
PPV	0.51	0.50	0.46	0.46	0.42	0.48	0.47	0.44
NPV	0.86	0.86	0.82	0.84	0.83	0.84	0.83	0.80
Fixed sensitivity = 0.9								
Specificity	0.39	0.39	0.33	0.35	0.13	0.29	0.33	0.24
PPV	0.45	0.45	0.43	0.45	0.37	0.43	0.43	0.40
NPV	0.91	0.91	0.89	0.91	nan	0.94	0.86	0.84

Table 5.5. Calculated NPV, PPV, and specificity at different fixed sensitivities for each model. If it was not possible to fix the sensitivity at the exact threshold, the closest point above the threshold was chosen.

Evaluation of the best-performing model

Based on the ROC-AUC values, it can be seen that RF was the best-performing model. Through the feature selection, RF selected *Time in range*, *HbA1c*, *Telemonitored*, *Systolic*, *Health status*, and *Insulin type* as the features that were best at classifying nonadherent patients. The importance of these features can be seen in Figure 5.3, where *Time in range* was the feature with the highest importance.

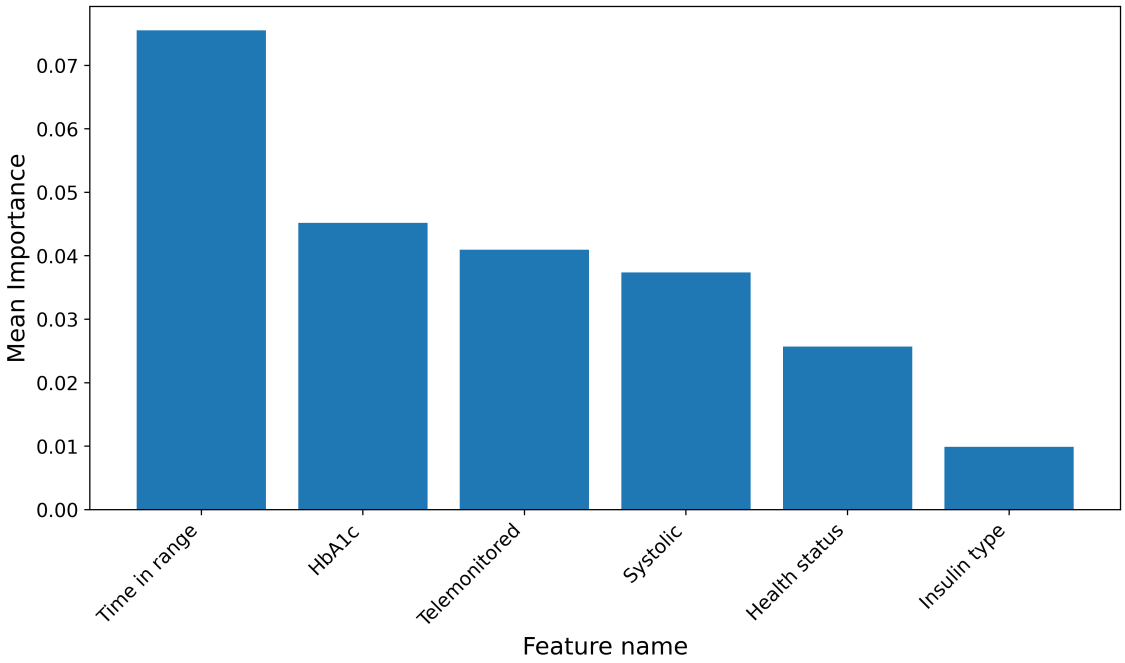


Figure 5.3. Permutation feature importance of the selected features for RF.

SHAP values were calculated on the selected features, which can be seen in Figure 5.4. The SHAP values indicate that nonadherent patients have a low time in range, low general health, high baseline HbA1c values, and high systolic blood pressure. Furthermore, the SHAP values showed that nonadherent patients were more likely not to be telemonitored and only use basal insulin.

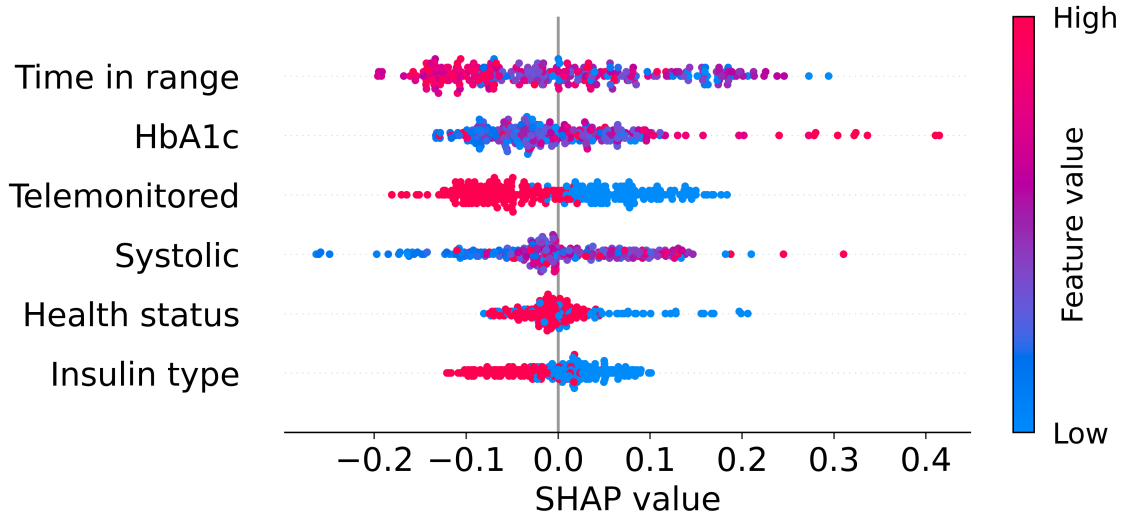


Figure 5.4. SHAP values of the selected features, where each dot represents an observation. Red dots indicate high feature values, whereas blue dots indicate low feature values.

Time management 6

At the beginning of the semester, an initial activity and time plan was created to have an overview of the project period. This plan was made in a Gantt chart as seen in Figure 6.1. Additionally, the plan was used as a part of the approval for the master thesis.

				Individual columns represent weeks.																				
		MÅNED		Februar					Marts					April					Maj				
AKTIVITET	% DONE	START DATO	SLUT DATO	5	6	7	8	9	9	10	11	12	13	14	15	16	17	18	18	19	20	21	22
Problemdomæne																							
Litteratursøgning	10%	1/2	16/2	X	X	X																	
Definer problem	0%	19/2	21/2				X																
Udfyld PRISMA	0%	22/2	23/2				X																
Lav overblik over litteratur	0%	22/2	1/3				X	X															
Metoder og udvikling																							
Modtage data	0%	Start feb	Slut marts		X					(X)	(X)	(X)	(X)										
Vælge og definere features	0%	1/3	15/3					X	X	X													
Vælg algoritmer	0%	1/3	8/3						X	X													
Udvikling af modeller	0%	4/3	4/4							X	X	X	X	X									
Analyse og resultater																							
Undersøge karakteristika	0%	1/4	15/4											X	X	X							
Statistisk analyse	0%	1/4	15/4											X	X	X							
Lav figurer	0%	16/4	26/4													X	X						
Formalia																							
Skrive artikel	0%	15/4	30/5													X	X	X	X	X	X	X	
Skrive arbejdsblade	0%	12/2	8/3			X	X	X	X	X													
Rette	0%	-	-								X	X							X	X	X	X	
Aflevering	0%	-	31/5																				X
Semestergruppemøder	0%	28/2	17/4					X								X							
Statusseminar + planlægning	0%	27/3	2/4										X	X									

Figure 6.1. Initial version of the Gantt chart made at the start of February.

The Gantt chart was separated into different project activities, regarding the problem domain, methods and development, analysis and results, and formality. The time plan was made using backcasting to set deadlines for when specific parts of the project should be finished. Using backcasting means that we started with the final deadlines and worked backward towards the beginning, ensuring enough time for the final activities. The time planning was an iterative process due to changes and unforeseen events. The final version of the Gantt chart can be seen in Figure 6.2.

				Individual columns represent weeks.																				
		MÅNED		Februar					Marts					April					Maj				
AKTIVITET	**% DONE**	**START DATO**	**SLUT DATO**	**5**	**6**	**7**	**8**	**9**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**18**	**19**	**20**	**21**	**22**
Problemdomæne																							
Litteratursøgning	100%	1/2	16/2	x	x	x	x																
Definer problem	100%	19/2	21/2				x	x	x			x											
Udfyld PRISMA	100%	22/2	23/2				x																
Lav overblik over litteratur	100%	22/2	1/3				x	x	x			x											
Metoder og udvikling																							
Modtage data	100%	Start feb	Slut marts		x	x																	
Vælge og definere features	100%	1/3	15/3						x	x	x	x											
Vælg og undersøge algoritmer	50%	1/3	8/3							x	x	x	x										
Udvikling af modeller	40%	4/3	4/4									x	x	x	x	x	x	x	x				
Analyse og resultater																							
Undersøge karakteristika	0%	1/4	15/4											x	x								
Statistisk analyse	0%	1/4	15/4											x	x								
Lav figurer og tabeller	0%	16/4	26/4				x									x	x	x	x	x			
Formalia																							
Skrive artikel	0%	15/4	30/5												x	x	x	x	x	x	x	x	
Skrive arbejdsblade	0%	12/2	8/3					x	x	x	x	x	x	x	x	x	x	x	x				
Rette	0%	-	-											x	x				x	x	x	x	x
Aflevering	0%	-	31/5																				x
Semestergruppemøder	0%	28/2	17/4					x								x							
Statusseminar + planlægning	0%	27/3	2/4										x	x									

Figure 6.2. Final version of the Gantt chart made at the beginning of May.

Through the first six weeks, we found that we had made an inaccurate definition of the problem domain due to the clinical relevance. Therefore, we started by looking back at the literature and defining a new problem. This resulted in a shift in the entire process where we used more time on individual tasks than anticipated. Furthermore, due to low performances in the machine learning models, various additional tests were made to find the potential error. This resulted in three additional weeks where we worked on the worksheets and the development of the models. Seen retrospectively, we might have been too fast in choosing a direction which meant that we needed to redo a few tasks. This emphasized the importance of thorough investigation and definition according to clinical relevance, thus we got a greater understanding of the problem domain. However, a greater understanding of the problem domain also caused a more efficient development phase late in the project.

As the Gantt chart was used to give an overview of the whole project period, a weekly time plan was also made to plan week-specific tasks. This plan was created in Microsoft Teams' Planner, which is an add-on tool to organize tasks. Here we made an assignment collection for the week and a collection for minor tasks, that needed to be done when there was time. These minor tasks could be small changes in the report or correction of figures or tables. An example of the work that needed to be made in week 15 can be seen in Figure 6.3.

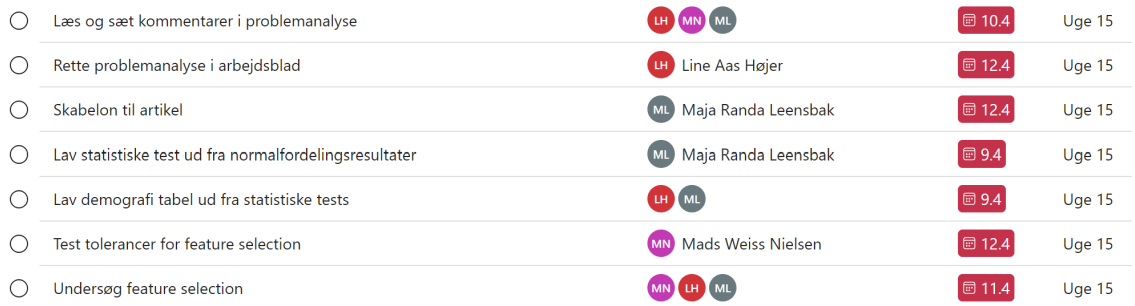


Figure 6.3. Example of week 15 in Planner with different specified tasks.

In Planner, it is possible to assign a task to a person. Thereby, everyone knew which task they were going to work on within the respective week. Furthermore, it was possible to select a deadline for when the tasks needed to be finished. When a task was finished, it was checked off from the overview, so only missing tasks were left. Besides the Gantt chart and the weekly time plan, a shared Microsoft Teams calendar was established. This calendar was used for more personalized things such as doctor appointments, work besides the study, etc.

These three planning elements were used to structure the semester to make sure that the project was finished on the deadline. Throughout this project, we used the Gantt diagram and Teams Planner thoroughly, whereas we sometimes forgot about the Teams calendar. The Planner was a great tool as tasks were written down instead of being planned orally. This ensured that no tasks were forgotten and that there was always an overview of missing tasks. The shared calendar was often forgotten as it was easier to explain upcoming plans orally and the calendar was time-consuming. However, as the project continued small plans were forgotten and it was not always possible to conduct a long-term plan. This could be optimized by using the calendar thoroughly.

Bibliography

[1] Cleveland Clinic. Diabetes: What It Is, Causes, Symptoms, Treatment and Types; 2024. [Online; accessed 19. Feb. 2024]. Available from: https://my.clevelandclinic.org/health/diseases/7104-diabetes.

[2] Mayo Clinic. Diabetes - Symptoms and causes; 2023. [Online; accessed 19. Feb. 2024]. Available from: https://www.mayoclinic.org/diseases-conditions/diabetes/symptoms-causes/syc-20371444.

[3] American Diabetes Association. Living With Diabetes | ADA; 2024. [Online; accessed 16. Apr. 2024]. Available from: https://diabetes.org/living-with-diabetes.

[4] World Health Organization. Diabetes. World Health Organization: WHO. 2023 Apr. Available from: https://www.who.int/news-room/fact-sheets/detail/diabetes.

[5] National Institute of Diabetes and Digestive and Kidney Diseases. What Is Diabetes?; 2023. [Online; accessed 19. Feb. 2024]. Available from: https://www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes.

[6] Mayo Clinic. Diabetes - Diagnosis and treatment - Mayo Clinic; 2023. [Online; accessed 20. Feb. 2024]. Available from: https://www.mayoclinic.org/diseases-conditions/diabetes/diagnosis-treatment/drc-20371451.

[7] National Institute of Diabetes and Digestive and Kidney Diseases. Insulin, Medicines, and Other Diabetes Treatments; 2023. [Online; accessed 20. Feb. 2024]. Available from: https://www.niddk.nih.gov/health-information/diabetes/overview/insulin-medicines-treatments.

[8] National Institute of Diabetes and Digestive and Kidney Diseases. Insulin, Medicines, and Other Diabetes Treatments; 2016. [Online; accessed 8. Mar. 2024]. Available from: https://www.niddk.nih.gov/health-information/diabetes/overview/managing-diabetes.

[9] Organization WH. Adherence to long-term therapies : evidence for action. World Health Organization; 2003. Available from: https://iris.who.int/handle/10665/42682.

[10] Saundankar V, Peng X, Fu H, Ascher-Svanum H, Rodriguez A, Ali A, et al. Predictors of Change in Adherence Status from 1 Year to the Next Among Patients with Type 2 Diabetes Mellitus on Oral Antidiabetes Drugs. Journal of Managed Care and Specialty Pharmacy. 2016 May;22(5):467–482. Available from: http://dx.doi.org/10.18553/jmcp.2016.22.5.467.

[11] Mishra A, Pradhan SK, Sahoo BK, Das A, Singh AK, Parida SP. Assessment of Medication Adherence and Associated Factors Among Patients With Diabetes Attending a Non-communicable Disease Clinic in a Community Health Centre in Eastern India. Cureus. 2023 Aug. Available from: http://dx.doi.org/10.7759/cureus.43779.

[12] Nazir SUR, Hassali MA, Saleem F, Bashir S, Aljadhey H. Association Between Diabetes-related Knowledge and Medication Adherence: Results From Cross-sectional Analysis. Altern Ther Health Med. 2016 Nov;22(6):8-13. Available from: https://pubmed.ncbi.nlm.nih.gov/27866175.

[13] Krishnan V S, V R. A cross sectional study on drug adherence among type 2 diabetes mellitus patients attending a tertiary care hospital in Chennai. Indian Journal of Public Health Research and Development. 2020 Mar;11(3):964-9. Available from: https://medicopublication.com/index.php/ijphrd/article/view/1497/1380.

[14] Lee DSU, Lee H. Adherence and persistence rates of major antidiabetic medications: a review. Diabetology and Metabolic Syndrome. 2022 Jan;14(1). Available from: http://dx.doi.org/10.1186/s13098-022-00785-1.

[15] Lee CS, Tan JHM, Sankari U, Koh YLE, Tan NC. Assessing oral medication adherence among patients with type 2 diabetes mellitus treated with polytherapy in a developed Asian community: a cross-sectional study. BMJ Open. 2017 Sep;7(9). Available from: http://dx.doi.org/10.1136/bmjopen-2017-016317.

[16] Iqbal Q, Bashir S, Iqbal J, Iftikhar S, Godman B. Assessment of medication adherence among type 2 diabetic patients in Quetta city, Pakistan. Postgraduate Medicine. 2017 May;129(6):637–643. Available from: http://dx.doi.org/10.1080/00325481.2017.1328251.

[17] Eby EL, Bajpai S, Faries DE, Haynes VS, Lage MJ. The Association Between Adherence to Insulin Therapy and Health Care Costs for Adults with Type 2 Diabetes: Evidence from a U.S. Retrospective Claims Database. Journal of Managed Care and Specialty Pharmacy. 2020 Sep;26(9).

[18] Shiyanbola OO, Unni E, Huang YM, Lanier C. Using the extended self-regulatory model to characterise diabetes medication adherence: a cross-sectional study. BMJ Open. 2018 Nov;8(11). Available from: http://dx.doi.org/10.1136/bmjopen-2018-022803.

[19] Jimmy B, Jose J, Al-Hinai ZA, Wadair IK, Al-Amri GH. Adherence to Medications among Type 2 Diabetes Mellitus Patients in Three Districts of Al Dakhliyah Governorate, Oman: A cross-sectional pilot study. Sultan Qaboos University Medical Journal. 2014 May;14(2):e231. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3997541.

[20] McClintock HF, Edmonds SE, Bogner HR. Adherence patterns to oral hypoglycemic agents among primary care patients with type 2 diabetes. Primary Care Diabetes. 2023 Apr;17(2):180–184. Available from: http://dx.doi.org/10.1016/j.pcd.2023.01.014.

[21] Chen YL, Nguyen PA, Chien CH, Hsu MH, Liou DM, Yang HC. Machine learning-based prediction of medication refill adherence among first-time insulin users with type 2 diabetes. Diabetes Research and Clinical Practice. 2024 Jan;207. Available from: http://dx.doi.org/10.1016/j.diabres.2023.111033.

[22] Li M, Lu X, Yang H, Yuan R, Yang Y, Tong R, et al. Development and assessment of novel machine learning models to predict medication non-adherence risks in type 2

diabetics. Frontiers in Public Health. 2022 Nov;10. Available from: http://dx.doi.org/10.3389/fpubh.2022.1000622.

[23] Wu XW, Yang HB, Yuan R, Long EW, Tong RS. Predictive models of medication non-adherence risks of patients with T2D based on multiple machine learning algorithms. BMJ Open Diabetes Research and Care. 2020 Mar;8(1). Available from: http://dx.doi.org/10.1136/bmjdrc-2019-001055.

[24] Sun P, Lian J. Treatment adherence in newly diagnosed type 2 diabetes: patient characteristics and long-term impact of adherence on inpatient care utilization. Postgraduate Medicine. 2016 Feb;128(4):338–345. Available from: http://dx.doi.org/10.1080/00325481.2016.1151326.

[25] Kristy I, Tunceli K, Zhao C, Davies MJ, Brodovicz KG, Alexander CM, et al. Factors associated with adherence to oral antihyperglycemic monotherapy in patients with type 2 diabetes. Patient Preference and Adherence. 2015 Jan. Available from: http://dx.doi.org/10.2147/PPA.S71346.

[26] Campbell DJT, Campbell DB, Ogundeji Y, Au F, Beall R, Ronksley PE, et al. First-line pharmacotherapy for incident type 2 diabetes: Prescription patterns, adherence and associated costs. Diabetic Medicine. 2021 Jun;38(9). Available from: http://dx.doi.org/10.1111/dme.14622.

[27] Huang J, Ding S, Xiong S, Liu Z. Medication Adherence and Associated Factors in Patients With Type 2 Diabetes: A Structural Equation Model. Frontiers in Public Health. 2021 Nov;9. Available from: http://dx.doi.org/10.3389/fpubh.2021.730845.

[28] Ranjbaran S, Shojaeizadeh D, Dehdari T, Yaseri M, Shakibazadeh E. Determinants of medication adherence among Iranian patients with type 2 diabetes: An application of health action process approach. Heliyon. 2020 Jul;6(7). Available from: http://dx.doi.org/10.1016/j.heliyon.2020.e04442.

[29] Kretchy IA, Koduah A, Ohene-Agyei T, Boima V, Appiah B. The Association between Diabetes-Related Distress and Medication Adherence in Adult Patients with Type 2 Diabetes Mellitus: A Cross-Sectional Study. Journal of Diabetes Research. 2020 Mar;2020:1–10. Available from: http://dx.doi.org/10.1155/2020/4760624.

[30] Jackson IL, Adibe MO, Okonta MJ, Ukwe CV. Medication Adherence in Type 2 Diabetes Patients in Nigeria. Diabetes Technology and Therapeutics. 2015 Jun;17(6):398–404. Available from: http://dx.doi.org/10.1089/dia.2014.0279.

[31] Masaba BB, Mmusi-Phetoe RM. Determinants of Non-Adherence to Treatment Among Patients with Type 2 Diabetes in Kenya: A Systematic Review. Journal of Multidisciplinary Healthcare. 2021 Jan;Volume 13:2069–2076. Available from: http://dx.doi.org/10.2147/JMDH.S270137.

[32] Kumar PS, Ranjan SS, Pratyush M, Snehashini D. Assessment of medication adherence and its predictors in Type 2 diabetes mellitus patients in tertiary care teaching hospital: A cross-sectional observational study. Journal of Cardiovascular Disease Research. 2023 Dec;14(12):2023. Available from:

https://www.researchgate.net/publication/376349519_Assessment_of_medication_adherence_and_its_predictors_in_Type_2_diabetes_mellitus_patients_in_tertiary_care_teaching_hospital_A_cross-sectional_observational_study.

[33] Nelson LA, Wallston KA, Kripalani S, LeStourgeon LM, Williamson SE, Mayberry LS. Assessing barriers to diabetes medication adherence using the Information-Motivation-Behavioral skills model. Diabetes Research and Clinical Practice. 2018 Aug;142:374–384. Available from: http://dx.doi.org/10.1016/j.diabres.2018.05.046.

[34] Azri N, Norsa'adah B, Hassan NB, Naing NN. Insulin Adherence and Associated Factors in Patients with Type 2 Diabetes Mellitus Treated in Klang Primary Health Care Centres. Malays J Med Sci. 2021 Dec;28(6):76-87.

[35] Hashimoto K, Urata K, Yoshida A, Horiuchi R, Yamaaki N, Yagi K, et al. The relationship between patients' perception of type 2 diabetes and medication adherence: a cross-sectional study in Japan. Journal of Pharmaceutical Health Care and Sciences. 2019 Jan;5(1). Available from: http://dx.doi.org/10.1186/s40780-019-0132-8.

[36] Abdullah NF, Khuan L, Theng CA, Sowtali SN, Juni MH. Effect of patient characteristics on medication adherence among patients with type 2 diabetes mellitus: a cross-sectional survey. Contemporary Nurse. 2019 Jan;55(1):27–37. Available from: http://dx.doi.org/10.1080/10376178.2019.1583067.

[37] Nørlev JTD, Kronborg T, Jensen MH, Vestergaard P, Hejlesen O, Hangaard S. A Three-Step Data-Driven Methodology to Assess Adherence to Basal Insulin Therapy in Patients With Insulin-Treated Type 2 Diabetes. Journal of Diabetes Science and Technology. 2023 Dec. Available from: http://dx.doi.org/10.1177/19322968231222007.

[38] Sánchez-Hernández MS, Rodríguez-Caldero MC, Martín-Pérez MP, Mira-Solves JJ, Vitaller-Burillo J, Carratalá-Munuera MC. Impact of adherence to Mediterranean diet and/or drug treatment on glycaemic control in type 2 diabetes mellitus patients: DM2-CUMCYL study. Primary Care Diabetes. 2020 Dec;14(6):685–691. Available from: http://dx.doi.org/10.1016/j.pcd.2020.06.008.

[39] Sison G. The Morisky Medication Adherence Scale: An Overview; 2018. [Online; accessed 15. Apr. 2024]. Available from: https://www.pillsy.com/articles/the-morisky-medication-adherence-scale-definition-alternatives-and-overview.

[40] Meige S. Differences between MPR and PDC | OpenHealth Knowledge Base; 2024. [Online; accessed 15. Apr. 2024]. Available from: https://success.openhealth.fr/en/articles/3722666-differences-between-mpr-and-pdc.

[41] Thyde DN, Mohebbi A, Bengtsson H, Jensen ML, Mørup M. Machine Learning-Based Adherence Detection of Type 2 Diabetes Patients on Once-Daily Basal Insulin Injections. Journal of Diabetes Science and Technology. 2020 Apr;15(1):98–108. Available from: http://dx.doi.org/10.1177/1932296820912411.

[42] Syafhan NF, Donnelly R, Harper R, Harding J, Mulligan C, Hogg A, et al. Adherence to metformin in adults with type 2 diabetes: a combined method approach. Journal of

Pharmaceutical Policy and Practice. 2022 Oct;15(1). Available from: http://dx.doi.org/10.1186/s40545-022-00457-5.

[43] Eze UIH, Akhumi TF, Iheanacho CO, Saka SA. Drug therapy and medication adherence in type 2 diabetes in a care facility: A cross sectional survey. Exploratory Research in Clinical and Social Pharmacy. 2022 Dec;8. Available from: http://dx.doi.org/10.1016/j.rcsop.2022.100200.

[44] Wulandari N, Maifitrianti M, Hasanah F, Atika S, Dini Putri R. Medication adherence assessment among patients with type 2 diabetes mellitus treated polytherapy in indonesian community health center: A cross sectional-study. Journal of Pharmacy And Bioallied Sciences. 2020;12(6). Available from: http://dx.doi.org/10.4103/jpbs.JPBS_257_19.

[45] Horsburgh S, Barson D, Zeng J, Sharples K, Parkin L. Adherence to metformin monotherapy in people with type 2 diabetes mellitus in New Zealand. Diabetes Research and Clinical Practice. 2019 Dec;158. Available from: http://dx.doi.org/10.1016/j.diabres.2019.107902.

[46] Lee DSU, Lee H. Clinical Characteristics Associated with Adherence and Persistence in Patients with Type 2 Diabetes Mellitus Treated with Dulaglutide. Journal of Diabetes Research. 2023 Jun;2023:1–12. Available from: http://dx.doi.org/10.1155/2023/7917641.

[47] Vlacho B, Simarro FL, Mata-Cases M, Miravet S, Escribano-Serrano J, Asensio D, et al. Adherence to antidiabetic treatment among patients managed in primary care centres in Spain: the INTENSE study. Primary Care Diabetes. 2022 Dec;16(6):760–767. Available from: http://dx.doi.org/10.1016/j.pcd.2022.10.004.

[48] Parada H, Horton LA, Cherrington A, Ibarra L, Ayala GX. Correlates of Medication Nonadherence Among Latinos With Type 2 Diabetes. The Diabetes Educator. 2012 Apr;38(4):552–561. Available from: http://dx.doi.org/10.1177/0145721712445215.

[49] Cheng SH, Chen CC, Chin-Hsiao Tseng M. Does Medication Adherence Lead to Lower Healthcare Expenses for Patients With Diabetes? AJMC. 2013 Aug;19(8):662-70. Available from: https://rb.gy/to90ii.

[50] Nazir SUR, Hassali MA, Saleem F, Bashir S, Aljadhey H. Disease related knowledge, medication adherence and glycaemic control among patients with type 2 diabetes mellitus in Pakistan. Primary Care Diabetes. 2016 Apr;10(2):136–141. Available from: http://dx.doi.org/10.1016/j.pcd.2015.09.004.

[51] Gatwood JD, Chisholm-Burns M, Davis R, Thomas F, Potukuchi P, Hung A, et al. Differences in health outcomes associated with initial adherence to oral antidiabetes medications among veterans with uncomplicated Type 2 diabetes: a 5-year survival analysis. Diabetic Medicine. 2018 Jul;35(11):1571–1579. Available from: http://dx.doi.org/10.1111/dme.13775.

[52] Buysman EK, Liu F, Hammer M, Langer J. Impact of Medication Adherence and Persistence on Clinical and Economic Outcomes in Patients with Type 2 Diabetes Treated with Liraglutide: A Retrospective Cohort Study. Advances in Therapy. 2015 Apr;32(4):341–355. Available from: http://dx.doi.org/10.1007/s12325-015-0199-z.

[53] Wong MCS, Wu CHM, Wang HHX, Li HW, Hui EMT, Lam AT, et al. Association between the 8-item Morisky medication adherence scale (MMAS-8) score and glycaemic control among Chinese diabetes patients. The Journal of Clinical Pharmacology. 2014 Nov;55(3):279–287. Available from: http://dx.doi.org/10.1002/jcph.408.

[54] Vervloet M, Spreeuwenberg P, Bouvy ML, Heerdink ER, de Bakker DH, van Dijk L. Lazy Sunday afternoons: the negative impact of interruptions in patients' daily routine on adherence to oral antidiabetic medication. A multilevel analysis of electronic monitoring data. European Journal of Clinical Pharmacology. 2013 Apr;69(8):1599–1606. Available from: http://dx.doi.org/10.1007/s00228-013-1511-y.

[55] Yong SY, Goh GM, Loh HH. Insulin adherence and the associated factors among patients with type 2 diabetes mellitus at the Hospital Queen Elizabeth II, Sabah. Journal of Public Health. 2020 Nov;30(5):1319–1327. Available from: http://dx.doi.org/10.1007/s10389-020-01409-6.

[56] Wang FC, Chang W, Nie SL, Shen BX, He CY, Zhao WC, et al. Predicting medication nonadherence risk in the Chinese type 2 diabetes mellitus population – establishment of a new risk nomogram model: a retrospective study. Journal of International Medical Research. 2021 Sep;49(9). Available from: http://dx.doi.org/10.1177/03000605211042502.

[57] Nazir SUR, Hassali MA, Saleem F, Bashir S, Aljadhey H. Does adherence to the therapeutic regimen associate with health related quality of life: Findings from an observational study of type 2 diabetes mellitus patients in Pakistan. Pakistan Journal of Pharmaceutical Sciences. 2017 Nov;30(6):2159-66. Available from: https://go.gale.com/ps/i.do?id=GALE%7CA520323093&sid=googleScholar&v=2.1%E2%81%A2=r&linkaccess=abs&issn=1011601X&p=AONE&sw=w&userGroupName=anon%7E3841bd79&aty=open-web-entry.

[58] Sapkota S, Brien JAE, Greenfield JR, Aslani P. A Systematic Review of Interventions Addressing Adherence to Anti-Diabetic Medications in Patients with Type 2 Diabetes–Components of Interventions. PLoS One. 2015 Jun;10(6).

[59] Egede LE, Gebregziabher M, Lynch CP, Echols C. Longitudinal Effects of Medication Nonadherence on Glycemic Control. Annals of Pharmacotherapy. 2014 Feb;48:562-70.

[60] Kang Y, Hur Y. Medication Adherence and Its Associated Factors in Laotians With Type 2 Diabetes Mellitus. Clinical Nursing Research. 2019 May;29(5):331–338. Available from: http://dx.doi.org/10.1177/1054773819849349.

[61] Zongo A, Guénette L, Moisan J, Grégoire JP. Predictive Validity of Self-Reported Measures of Adherence to Noninsulin Antidiabetes Medication against Control of Glycated Hemoglobin Levels. Canadian Journal of Diabetes. 2016 Feb;40(1):58–65. Available from: http://dx.doi.org/10.1016/j.jcjd.2015.06.008.

[62] Hangaard S, Kronborg T, Hejlesen O, Aradóttir TB, Kaas A, Bengtsson H, et al. The Diabetes teleMonitoring of patients in insulin Therapy (DiaMonT) trial: study protocol for a randomized controlled trial. Trials. 2022 Dec;23(985).

[63] Delua J. Supervised vs. Unsupervised Learning: What's the Difference? - IBM Blog. IBM Blog. 2021 Mar. Available from: https://www.ibm.com/blog/supervised-vs-unsupervised-learning.

[64] Chollet F. Chapter 1. What is deep learning?; 2019. [Online; accessed 19. Apr. 2024]. Available from: https://livebook.manning.com/book/deep-learning-with-python/chapter-1/17.

[65] Coursera. Deep Learning vs. Machine Learning: A Beginner's Guide; 2024. [Online; accessed 19. Apr. 2024]. Available from: https://www.coursera.org/articles/ai-vs-deep-learning-vs-machine-learning-beginners-guide.

[66] Collins GS, Dhiman P, Ma J, Schlussel MM, Archer L, Van Calster B, et al. Evaluation of clinical prediction models (part 1): from development to external validation. BMJ. 2024 Jan;384:e074819.

[67] Brownlee J. How to Choose a Feature Selection Method For Machine Learning - MachineLearningMastery.com. MachineLearningMastery. 2020 Aug. Available from: https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data.

[68] Zohar Y. Permutation Importance (PI) : Explain Machine Learning Predictions; 2024. [Online; accessed 1. May 2024]. Available from: https://www.aporia.com/learn/feature-importance/explain-ml-models-with-permutation-importance.

[69] Trevisan V. Using SHAP Values to Explain How Your Machine Learning Model Works. Medium. 2022 Jul. Available from: https://towardsdatascience.com/using-shap-values-to-explain-how-your-machine-learning-model-works-732b3f40e137.