Machine Learning-Based Adherence Detection of Type 2 Diabetes Patients on Once-Daily Basal Insulin Injections

Thyde, Daniel N; Mohebbi, Ali; Bengtsson, Henrik; Jensen, Morten Lind; Mørup, Morten

Published in:
Journal of Diabetes Science and Technology

Link to article, DOI:
10.1177/1932296820912411

Publication date:
2021

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):
Machine Learning Based Adherence Detection of Type 2 Diabetes Patients on Once-daily Basal Insulin Injections

Daniel N. Thyde, M.Sc.1*, Ali Mohebbi, M.Sc.1,2*, Henrik Bengtsson, MBA2, Morten Lind Jensen, M.D., Ph.D.3, Morten Mørup, Ph.D.1

Author Affiliations: 1DTU Compute, Department of Applied Mathematics and Computer Science, Kgs. Lyngby, Denmark; 2 Novo Nordisk A/S, Device R&D, Hillerød, Denmark, 3 Novo Nordisk A/S, Medical & Science, Søborg, Denmark
* These authors have contributed equally to this work.

Abbreviations: (AUROC) area under the receiver operator curve, (CGM) continuous glucose monitoring, (CHO) carbohydrates, (CNN) convolutional neural network, (CV) coefficient of variance, (DL) deep learning, (DOC) day of classification, (FC) fully connected, (HCP) healthcare professional, (LR) logistic regression, (META) mean ensemble test accuracy, (ML) machine learning, (MLP) multilayer perceptron, (MVP) Medtronic Virtual Patient, (NHU) number of hidden units, (PG) plasma glucose, (PK) pharmacokinetic, (ReLU) rectified linear unit, (ROC) receiver operator curve, (SD) standard deviation, (SI) insulin sensitivity, (SMPG) self-monitoring plasma glucose, (T2D) type 2 diabetes, (T1D) type 1 diabetes, (TOI) time of injection

Keywords: adherence detection, continuous glucose monitoring, type 2 diabetes, machine learning, deep learning, virtual patient

Corresponding Author: Morten Mørup, Danmarks Tekniske Universitet, Richard Petersens Plads, Building 321, 2800 Kgs. Lyngby, Denmark; email address mmor@dtu.dk

Figures and tables: 7 figures and 2 tables
Abstract

Background: Lack of treatment adherence can lead to life-threatening health complications for people with Type 2 Diabetes (T2D). Recent improvements and availability in Continuous Glucose Monitoring (CGM) technology have enabled various possibilities to monitor diabetes treatment. Detection of missed once-daily basal insulin injections can be used to provide feedback to patients, thus improving their diabetes management. In this study, we explore how Machine Learning (ML) based on CGM data can be used for detecting adherence to once-daily basal insulin injections.

Methods: In-silico CGM data were generated to simulate a cohort of T2D patients on once-daily insulin injection (Tresiba®). Deep Learning methods within ML based on automatic feature extraction including Convolutional Neural Networks were explored and compared to simple feature-engineered ML classification models for adherence detection. It was further investigated if fused expert-dependent and automatically learned features could improve performance, resulting in a comparison of 6 different detection models. Adherence was detected throughout each day with an increasing amount of CGM data available.

Results: The adherence detection accuracy improved as more CGM data became available on the day of classification. The 3 classification models based on expert-engineered features obtained mean accuracies of 78.6%, 78.2% and 78.3%. The classification model based purely on learned features obtained a mean accuracy of 79.7%. The 2 classification models fusing expert-engineered and learned features obtained mean accuracies of 79.7% and 79.8%. All the mentioned results were obtained 16 hours after time of injection.

Conclusion: The results suggest that adherence detection based on CGM data is feasible. Even though our study based on in-silico data indicates only slightly improved performance of more complex models, the question remains whether advanced models
would outperform the simple in a real-world setting. Thus, future studies on adherence monitoring using real CGM data are relevant.

**Introduction**

Diabetes has become a major public health problem worldwide with an estimate of 463 million people diagnosed and 232 million still undiagnosed\(^1\). Type 2 Diabetes (T2D) is more prevalent than Type 1 Diabetes (T1D) and accounts for ~ 90% of the diagnosed\(^2\). Elevated plasma glucose (PG) is caused by this chronic disease due to insufficient insulin production or decreased insulin sensitivity. Life-threatening health complications can occur in case of prolonged elevated PG. Hence, in progressed T2D, it may be required to supplement with long-acting basal insulin to maintain acceptable glucose levels. Thus, adherence to prescribed treatment and self-monitored plasma glucose (SMPG) are crucial for all diabetes patients\(^3\)–\(^6\). However, despite the minimal requirement of once-daily injections, studies report high variance in adherence levels\(^7\)–\(^9\).

Although SMPG is the common approach to measure PG levels, recent Continuous Glucose Monitoring (CGM) technology has enabled new opportunities and transformed the approach to diabetes care\(^6,10\)–\(^16\). CGM provides the ability to continuously monitor interstitial glucose levels with a predefined interval (usually between 5 to 15 minutes). Even though the use of CGM devices is currently mainly used in T1D, there is a clear expansion and progress towards CGM use in the treatment of T2D as well\(^10,13\).

At present, the availability and accessibility of real-life CGM data from people with T2D are limited, and even more so with reliable labels with respect to adherence to prescribed treatment. Given this limitation, we simulate the required data using a modified T2D version of the well-established Medtronic Virtual Patient (MVP) model. Such simulation
models enable generation of large amounts of CGM data and provide opportunities to investigate the impact of various factors (i.e. adherence level, intra-day variation of insulin sensitivity, inter-subject variance in level of consumed CHO, etc.) on the CGM response. Most importantly for this study, the models provide flexible and easy-to-use simulations to facilitate accelerated development of decision support tools utilizing CGM data. However, existing simulation models are not perfect in terms of providing realistic data. Therefore, considering the scope of this study, opportunities to improve the existing MVP model were sought\textsuperscript{17,18}.

Recently, Deep Learning (DL) has achieved state-of-the-art performance within Machine Learning in different research areas as well as industrial and medical applications\textsuperscript{19,20}. In particular, DL methods have proven powerful for automatic feature extraction using large datasets to reduce the need for feature engineering by domain experts\textsuperscript{20}. Given the increasing availability of connected medical devices including CGM and injection data, diabetes research can benefit from DL, and incrementally so as large CGM data sets become available.

The aim of this study was to develop an early alarm system for adherence detection using supervised learning based on large amounts of in-silico CGM and injection data. A common initial insulin treatment for T2D is once-daily basal insulin injection, and even though it currently is the simplest form of insulin therapy, this daily injection could be forgotten. It is therefore a relevant question whether a missing once-daily basal insulin injection can be detected based on CGM data from the same day, i.e. retrospective adherence detection, up to 16 hours after expected time of injection (TOI). Following the present setup, this is the latest TOI for a missed dose as the label of Tresiba\textsuperscript{®} states that a delayed dose should be taken within 8 hours of the next injection\textsuperscript{21}. 
Adherence to prescribed treatment is essential to achieve glycemic control. However, people with diabetes may in their daily lives often forget elements of the complicated treatment of diabetes, which can affect their treatment. For instance, a previous study has shown a clear link between non-adherence to prescribed treatment and higher mortality and hospitalization\(^4\). Therefore, an early alarm system designed to detect adherence/non-adherence is likely to provide clinically relevant treatment recommendations. Ultimately, this could delay or prevent onset of secondary conditions to T2D and improve the overall well-being of the patient.

**Methods**

In the context of this study, engineered CGM features based on three network architectures of varying complexity were contrasted against DL generated features based on *Convolutional Neural Networks* (CNNs) in addition to a fused approach in which engineered and learned features were combined considering two network architectures. This resulted in a comparison of 6 different detection models in total.

The speed and precision of adherence detection were explored after an injection of once-daily basal insulin using the simulated CGM data. Our recent preliminary study by Mohebbi et al.\(^22\) indicated the feasibility and potential of such an approach which is further evaluated and improved upon in this study.

In the simulations, adherence/non-adherence were defined binarily as either taken injection or missed injection at the defined TOI, thus there were no late or suboptimal dose injections. In the context of this study, adherence was defined as a full optimal dose taken within the introduced time variation and non-adherence was defined as a total dose omission.

*Medtronic Virtual Patient Model*
CGM data of T2D patients with labelled adherence/non-adherence to the prescribed once-daily insulin injection were simulated. The in-silico CGM data were based on the well-established and T2D modified MVP model emulating the excursions of PG concentration\(^{17,18}\). A schematic of the utilized compartment model is presented and described in Fig. 1.

In short, the model is based on the MVP model\(^{17}\), a physiological model for simulating 24-hour insulin-glucose dynamics in T1D patients along with quantified parameter sets for 10 patients. Moreover, a T2D augmentation was implemented by Aradóttir et al.\(^{18}\), introducing a linear endogenous insulin production and a two-compartment model describing absorption of carbohydrates (CHO)\(^{17,18,23,24}\). Furthermore, relevant parameters were varied in accordance with typical physiological differences between T1D and T2D patients including body weight, insulin sensitivity (\(S_I\)) and fasting glucose\(^{18,22}\). The full set of coupled differential equations describing the compartment model and further details are presented by Aradóttir et al.\(^{18}\).

**Data Simulation**

Several elements were implemented to the T2D MVP model in order to achieve more realistic in-silico CGM data. Each element is listed below followed by a short description in addition to conceptual plots as presented in Fig 2.

a) **Random adherence sequence**

Random adherence was implemented by drawing a random number from a uniform distribution before the simulation of each CGM day and thresholding at the desired level of adherence (Fig. 2a). All data in this paper were simulated with an adherence level of 95% in order to imitate clinical trial data acquired under a controlled setup where a high adherence level is expected.
b) **Variance in TOI**

If a day was adherent, the TOI would be at 08:00 AM ± 15 minutes. The standard deviation (SD) of 15 minutes was added to the TOI by adding a random number from a gaussian distribution (Fig. 2b). If a day was non-adherent the dose was fully omitted, and no actions were taken at the expected TOI.

c) **Intra-day variation of S₁**

We implemented an intra-day variance in S₁ in order to achieve a more realistic CGM dynamic (Fig. 2c). The daily excursions of S₁ presented by Shultz et al.²⁵ were extracted and confined to be within the diurnal and nocturnal subject specific S₁ values presented by Kanderian et al.¹⁷. Additionally, noise was added to the S₁-value before each PG estimation. The noise was acquired from a gaussian distribution with a SD of 5% of the difference between maximum and minimum of the intra-day S₁ (Fig. 2c).

d) **Inter-subject variance in level of consumed CHO**

To introduce an inter-subject variance in level of consumed CHO a meal factor was implemented (Fig. 2d). The meal factor was chosen at random before each simulation and multiplied on all meals each day. The factor was limited such that the daily CHO intake was below the upper boundary of ingested CHO presented by Kanderian et al.¹⁷ and above the recommended minimum intake of CHO for T2D patients²⁶.
e) **Removal of simulation initiation bias**

Each simulation initiates with a constant PG at the subject specific basal glucose level until actions are taken, such as a meal or insulin injection. To overcome the pattern, we removed this initiation bias by excluding the first 10 days of each simulation scenario consisting of 100 days (Fig. 2e). Hence, a series of random choices were introduced prior to the 11th day entailing 90 days of usable in-silico CGM.

f) **A stochastic simulation scheme**

Preliminary studies showed a low intra-day variance in the simulated data when compared to clinical data. A higher degree of variance was achieved in the simulated CGM data by applying the Euler-Maruyama method\(^{27}\) introducing stochastic noise during simulations (Fig. 2f).

In order to simulate CGM signals of a T2D cohort the parameter variation and meal case structure applied by Mohebbi et al.\(^{22}\) was used: 6 fasting glucose (\(G_b\)) levels between ~108-198 mg/dL (6-11 mmol/L), decreasing \(S_i\) by 30%, 50% and 70% and increasing body weight (\(B_W\)) by 10%, 30% and 50%. These parameter variations were applied to 9 out of 10 subjects from the original MVP model resulting in \(9_{\text{subjects}} \cdot 6_{G_b} \cdot 3_{S_i} \cdot 3_{B_W} = 486\) different subject configurations\(^{18,22}\). Subject number 10 showed unexplainable unrealistic dynamics and was therefore excluded. Each configuration was simulated 20 times resulting in \(486_{\text{config}} \cdot 90_{\text{days}} \cdot 20 = 874800\) CGM days.

To further ensure realistic data simulations, CGM sensor noise was included following Facchinetti et al.\(^{28}\). The noise model was developed using actual CGM data, implemented as a sum of two autoregressive models and applied after the data were simulated.
The study concerns patients treated with Tresiba® (Insulin Degludec), a once-daily long-acting insulin with a half-life of approximately 25 hours\textsuperscript{21}. A pharmacokinetic (PK) profile of Insulin Degludec described by Heise et al. was used in the simulations\textsuperscript{29}. Prior to each simulation, the patient specific optimal dose was determined using a stepwise titration algorithm presented in a study by Zinman et al.\textsuperscript{30}. Dose adjustments were based on the mean fasting PG of 3 days above target, or the minimum fasting PG if below target. A patient was considered in target if PG was between ~70-90 mg/dL (3.9-4.9 mmol/L).

*Day-to-day Variance*

Initial exploratory analysis indicated the fasting PG to be a primary feature in the detection of adherence. To ensure that the simulated CGM data had a degree of day-to-day variance proportional to that of actual PG measures, the simulated data were compared to clinical data. The coefficient of variance (CV) of the pre-breakfast PG in the simulated CGM data were compared to the CV of pre-breakfast SMPG measurements from a large clinical study containing 770 patients with T2D on once-daily Tresiba® treatment\textsuperscript{30}. For the simulated CGM data the pre-breakfast PG was defined as the lowest mean hour (lowest mean acquired from a sliding window of 12 consecutive PG values\textsuperscript{31}), between 06AM - 09AM. The pre-breakfast PG CV of the clinical study was 17% in comparison to 18% in the simulated CGM data.

*Deep Learning*

DL builds on neural networks and is characterized by having multiple hidden layers. As opposed to relying on features designed by domain experts, this enables the automatic extraction of features from large datasets by learning feature representations of increased complexity as the layers are traversed.
In this study, we investigated the potential for using CGM for adherence detection considering DL architectures for classification. Ground truth treatment adherence was available for the in-silico CGM data entailing the two class-labels of adherence and non-adherence as output. The training error used is the cross-entropy error function defined by,

\[ E = -\sum_{n=1}^{N} y_n \ln \hat{y}_n + (1 - y_n) \ln (1 - \hat{y}_n) \]  

where \( y_n \) and \( \hat{y}_n \) are the desired output and the estimated class probability for the \( n \)’th CGM signal. The estimated class probability is defined by the logistic sigmoid activation function,

\[ \hat{y}_n = \frac{1}{1 + e^{-z_n}} \]  

In case of logistic regression (LR) \( z_n \) is given by the linear model \( z_n = \sum_i w_i x_{ni} + b \) in which \( x_{ni} \) is the \( i \)'th engineered feature (described in Tab. 1) of CGM observation \( n \). We further considered a multilayered feed-forward neural network based on Multilayer Perceptrons (MLPs) and CNNs, the latter based on the raw CGM as input to the model. As non-linear activation functions we employed rectified linear units (ReLUs)\(^{32}\).

In order to find the optimal values for the model parameters, represented by weights \( w_i \) and \( b \) in the case of LR, the cross-entropy function was minimized. Here, LR is a convex problem whereas the MLP and CNN models are non-convex with issues of local minima. The minimum was found using the Adam optimizer based on stochastic gradient optimization\(^{33}\). The classification methods were implemented using the PyTorch framework version 0.4.1 in Python version 3.6.5.

Data Input
The input of the LR and MLP models were based on 11 engineered features of the CGM signal described in Tab. 1. Out of the 11 features, 7 were chosen with inspiration from the key metrics from the international consensus on the use of CGM and the remaining 4 were lowest mean hour, pre-breakfast PG, minimum and maximum measurements.

Considering the input to the feature-based architectures, the first 10 features were computed every 4 hours throughout the day of classification (DOC) (midnight-4 AM, 4 AM-8 AM, …, 8 PM-midnight), resulting in a total of 6 feature sets (6 x 10 features). In the same manner, historical data were also included to the input with the 4-hour interval feature sets acquired from 1-5 days prior to DOC. Furthermore, a set of all 11 features were calculated for a 24-hour interval when a full day of CGM was available (from midnight to midnight). The additional feature, restricted to 24-hour intervals, was the pre-breakfast PG defined as lowest mean hour between 6 AM-9 AM. For a clearer understanding Eq. 3 is included showing an example of input for adherence detection on a DOC at 8 AM:

\[
DOC_{8AM} = \text{days}_{\text{prior}} \cdot (\text{features}_{24h,\text{prior}} + 6 \cdot \text{features}_{4h,\text{prior}}) + 2 \cdot \text{features}_{4h,\text{DOC}}
\]

\[
= 5 \cdot (11 + 6 \cdot 10) + 2 \cdot 10
\]

\[
= 375 \text{ input features}
\]

Similarly, a new feature set was added to the input data for each added 4-hour interval. Furthermore, a set of 24-hour features were added at midnight on the DOC.

The input to the raw CGM models was the PG measurements having 288 evenly distributed samples per day (every 5 minutes). As for the feature models, CGM readings from 5 prior days were included in addition to the available data from the DOC. The input of the models based on both raw CGM and features take the entire feature input as described above in
addition to the raw CGM signal. It was chosen to include the previous 5 days as Tresiba® is expected to still have a presence in the plasma given the PK dynamics.

As mentioned before, 874,800 days of CGM were simulated with an adherence level of 95%. Hence 43,740 and 831,060 non-adherent and adherent days respectively. It was desired to have class-balance in the dataset to have a baseline accuracy of 50%. Therefore, 43,740 of the adherent days were selected at random resulting in a dataset of $2 \cdot 43,740 = 87,480$ CGM observations with a forced class-balance. Fig. 3 presents examples of an adherent and non-adherent CGM observation including DOC and the 5 prior days of CGM data used for detection.

**Architectures and Hyperparameters**

Fig. 4 show the architectures investigated in this study. A progression in model complexities was examined as shown in Fig. 4 going from simple feature-based models (A0-A2) to more complex architectures based on raw CGM data (A3) in addition to a combination of these (A4 and A5). This was done in order to examine whether advanced models based on automatic feature-extraction would outperform simple expert-dependent feature models as well as if added expert insight would increase the performance in the fused models. The green circles indicate CGM feature input whereas orange circles represent raw CGM signal.

A0 is a LR model represented by a single hidden unit in a single hidden layer. A1 and A2 are MLP architectures with 1 and 2 hidden fully connected (FC) layers, respectively. A3 is the architecture developed by Mohebbi et al. based on CNN layers (automatic feature-extraction) followed by a FC layer accountable for the classification. Architecture A4 combines the automatic feature extraction part (CNN layers) from A3 with the expert-
dependent features in a FC layer and performing classification based on all the features. A5 is similar to A4, but with an added FC layer before the output. In general, adding hidden FC layers introduces further nonlinearity into the model. Despite A0 being a simple LR model it was trained using the exact same framework as the other models for consistency.

*Leave-one-subject-out Cross-validation and Ensembling*

During training of the models, the simulated data from subjects 1-6 were used in a leave-one-subject-out cross-validation setup as indicated in Fig. 5. For each fold the model yielding the best validation accuracy was chosen based on 5 re-initializations and training of model parameters. CGM data for subjects 7-9 were used as the final unknown and unused test set. In this context, *ensembling* was obtained by averaging the probabilities obtained from different models before making a classification. Specifically, we ensemble the output probabilities acquired from the 6 models obtained during the cross-validation step. Furthermore, 4 consecutive runs (repeating the random selection of adherent days from the dataset followed by the process illustrated in Fig. 5) were performed for each tuned classification model in order to check for consistency/robustness of the developed models. Hence, the performance metric is qualitatively depicted by the mean ensemble test accuracy (META) ± SD across the 4 repetitions. For a single repetition, accuracy is defined by Eq. 4,

\[
\text{Accuracy} = \frac{TP + TN}{N}
\]  

(4)

where \( TP \) is the true positives, \( TN \) is the true negatives and \( N \) is the total number of observations.

*Statistical Considerations*
Due to the exploratory scope and nature of this study comparisons are descriptive. Although formal statistical testing could be employed, the present study is based on simulated data for which arbitrary power can be achieved by additional simulations. Furthermore, the validity of the simulated data and how the results generalize to real CGM data can be questioned. The present study indicates the potential for adherence detection using both simple and advanced DL approaches which forms an important framework for future investigations using formal assessments based on real CGM data.

**Results**

The study shows that for the best performing DL architecture, adherence and non-adherence could be detected 16 hours after expected TOI with a META of $79.8 \pm 0.5\%$. The result was achieved by the most complex architecture A5 and the closely related architectures A3 and A4 had similar performance. The simplest architecture A0 performed a META of $78.6 \pm 0.6\%$ at 16 hours after expected TOI. Fig. 6 presents a plot of the META of the architectures with increasing data available after the expected TOI.

Considering Fig. 6 the performances based on accuracy are very similar 16 hours after TOI. However, the architectures utilizing the automatic feature extraction (i.e. A3, A4 and A5) are slightly superior in contrast to the simpler architectures which only have the expert-dependent features as input. Furthermore, it is apparent that the difference between models is reduced as the available CGM data increases. At the TOI and the following 4 hours there seem to be limited or even no information on adherence detected by the models. The limited information on adherence at 4 hours after expected TOI (i.e. at noon) can be an artefact caused by the confined TOI always being followed by post-prandial PG excursions following breakfast and lunch, thus concealing the signal caused by an injection or omission.
Fig. 7 depicts the 24 (4 runs of each of the 6 models) receiver operator curves (ROCs) for 8 hours (Fig. 7a) and 16 hours (Fig. 7b) after TOI. These ROCs support the results presented in Fig. 6. It is easy to spot the different model performances between 8 and 16 hours after TOI. In general, the trade-off between sensitivity and specificity are similar for all the models at 16 hours after TOI. However, at 8 hours after TOI a tendency of a more favorable sensitivity-specificity trade-off is observed for the automatic feature extraction and hybrid models as opposed to the models relying only on the engineered features (i.e., A0, A1, A2). As we detect adherence, it should be emphasized that a true positive rate is more important than the false positive rate. This is due to the risk of overdose if a recommendation of a delayed injection was sent even though the patient did in fact take the dose at the expected TOI. The final results for 8 and 16 hours after TOI are presented in Tab. 2 including META and the average area under the ROC (AUROC) values as performance metrics for each model.

These results support the fact that the performance of each model is similar while the classification models based on automatic feature extraction (A3-A5) provided marginally better performance within 4-8 hours after TOI.

**Discussion**

In this study, thorough modifications and additions to the MVP model were applied followed by simulation of in-silico CGM data of a diverse T2D cohort. Several subject configurations were obtained and the simulated CGM data were set to be realistic through a comparison to clinical data acquired from a similar population.
There is an indication that the developed classification models can detect daily injection adherence with an accuracy of approx. 80% within DOC (i.e., 16 hours after injection). Furthermore, there seems to be a benefit when using DL architectures both with and without engineered and expert-dependent features. Within 4-8 hours after TOI using the models based on automatic feature extraction seems superior. On the other hand, the simple feature-engineered LR model performed almost as well as the more complex DL architectures at 16 hours after TOI. Thus, even simple features of the CGM data are promising for early adherence detection if the entire CGM day is available.

It should be emphasized that there is a clear benefit of using simple models in contrast to more complex models. One benefit is the lower requirements to the necessary number of observations for training due to fewer parameters being learned. Another benefit of the simple LR model is the superior explainability, i.e., it is easier to explain how decisions regarding the adherence are made. However, a more comprehensive investigation on the architectures and hyperparameter optimization is needed to elucidate the full potential of DL adherence detection including present and other models (e.g. Recurrent Neural Networks). Hence, we expect that better tailored DL frameworks can further improve upon the presented performance.

The primary limitation of the study is the simulated data and the inherent modelling choices made. Choices pertain to the limiting of the model to only include intra-day $S_t$ variance, and not inter-day variation of the $S_t$ dynamic. Also, the choice to confine TOI is only realistic in settings with strict treatment time schedules such as the presently considered clinical setting. Patients living on a once-daily treatment scheme can vary much more in TOI producing significantly different results. However, even though real life may have larger variability in TOI, most people are habitual in their medication patterns, and the simulated data represent a reasonably habitual cohort of T2D patients.
A second limitation is due to the confinement of the adherence level to 95% which may indicate that the results could be transferrable to only a very adherent group of T2D patients. In reality, patient adherence levels vary a lot indicating the need of further investigations of different adherence levels.

A third limitation is that actual CGM data often have various device and subject induced gaps which were not considered in this study. On the other hand, incremental improvements of future devices and consistent patient usage are anticipated to lower the impact of this limitation.

The results based on simulated data are inconclusive in terms of indicating either a benefit of the more complex DL architectures or a benefit of the simpler models. Therefore, applying the models to real CGM data to assess whether the simple models would perform similarly to the complex models in a real-life context is needed. In particular, this should include exploration of the advantage of the advanced models until 8 hours after TOI due to the potential clinical implications. Detecting a missed injection already after 8 hours provides an opportunity to recommend the patient to take the day’s injection of Tresiba®, benefitting from the flexibility of injection time of Tresiba® in accordance with the label\(^2\).

On the condition that the results in this study can be reproduced using real CGM data, an early adherence detection system could be implemented as a part of a decision and treatment support tool. It could be in the form of presenting a notification on a mobile app, to check whether a dose was injected or not. Furthermore, data from an adherence detection system could be shared to a cloud providing dynamic information and valuable insights about patient behavior to healthcare professionals (HCPs). This enables informed decisions by HCPs toward tailored diabetes treatment for each individual patient.

**Conclusion**
The T2D modified version of the MVP model was successfully used to simulate a large amount of realistic CGM data. The data were used to develop methods for treatment adherence detection. The automatically extracted features based on DL methods with added expert-dependent features performed best with accuracy of 79.8 ± 0.5% 16 hours after TOI. Although the fused CNN model with learned and expert-dependent features was the best performing model, it should be emphasized that almost equal performance could be achieved by the CGM consensus inspired simple feature engineered models at 16 hours after TOI with accuracy of 78.6 ± 0.6%. However, 8 hours after TOI the models based on automatic feature extraction indicated a clear advantage and should be further explored due to the potential clinical implications. According to the Tresiba® label, the injection time can be flexible, as long as the injection is taken at least 8 hours before next expected injection. This window of adherence detection is within the time window where a clinically relevant treatment recommendation can be provided.

**Funding Source:** none

**Acknowledgements:** The authors would like to thank Ann Kirstine Jørgensen for her graphical support in the included figures. We also appreciate Tinna B. Aradóttir and Zeinab Mahmoudi for their inputs with regards to the MVP model.

**Conflict-of-interest Disclosure:** none

**References**

4. Lee, W. C., Balu, S., Cobden, D., Joshi, A. V. & Pashos, C. L. Medication adherence and the associated health-economic impact among patients with type 2


**Tables:**

**Table 1:** List of engineered CGM-features used to detect adherence and non-adherence.

<table>
<thead>
<tr>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum PG measure of the interval</td>
</tr>
<tr>
<td>Maximum PG measure of the interval</td>
</tr>
<tr>
<td>Mean of entire interval</td>
</tr>
<tr>
<td>SD of the interval</td>
</tr>
<tr>
<td>Percent of interval with PG above 90 mg/dL (5 mmol/L)</td>
</tr>
<tr>
<td>Percent of interval with PG above 108 mg/dL (6 mmol/L)</td>
</tr>
<tr>
<td>Percent of interval with PG above 126 mg/dL (7 mmol/L)</td>
</tr>
<tr>
<td>Percent of interval with PG above 144 mg/dL (8 mmol/L)</td>
</tr>
<tr>
<td>Area under the PG measures in the interval</td>
</tr>
<tr>
<td>Lowest mean hour of the interval</td>
</tr>
<tr>
<td>Lowest mean hour between 6 AM and 9 AM*</td>
</tr>
</tbody>
</table>

*This feature is restricted in time and only calculated for 24-hour intervals.

**Table 2:** Mean ± SD of the AUROC and META for each of the detection models at 8 and 16 hours after expected TOI (based on 4 runs).

<table>
<thead>
<tr>
<th>Model</th>
<th>8 hours after TOI</th>
<th>16 hours after TOI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>META Mean ± SD</td>
<td>AUROC Mean ± SD</td>
</tr>
<tr>
<td>A0</td>
<td>63.9 ± 0.7 %</td>
<td>0.691 ± 0.003</td>
</tr>
<tr>
<td>A1</td>
<td>63.9 ± 0.5 %</td>
<td>0.687 ± 0.004</td>
</tr>
<tr>
<td>A2</td>
<td>64.0 ± 0.5 %</td>
<td>0.688 ± 0.013</td>
</tr>
<tr>
<td>A3</td>
<td>68.6 ± 0.9 %</td>
<td>0.765 ± 0.005</td>
</tr>
<tr>
<td>A4</td>
<td>68.1 ± 0.3 %</td>
<td>0.745 ± 0.004</td>
</tr>
<tr>
<td>A5</td>
<td>67.3 ± 1.2 %</td>
<td>0.737 ± 0.015</td>
</tr>
</tbody>
</table>
Figures:
Figure 1:

Figure 2:
Figure 3:

Figure 4:
Figure 5:

CGM data simulated from 9 subjects

87480 in-silico CGM days

Leave-one-subject-out cross-validation: subject 1-6

Class estimation and ensemble classification: subject 7-9

Training data
Validation data

Figure 6:

Accuracy [%]

Hours after expected injection