

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

Journal of Diabetes Science and Technology  
2021, Vol. 15(1) 98–108  
© 2020 Diabetes Technology Society



Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/1932296820912411  
journals.sagepub.com/home/dst



Daniel N. Thyde, MSc<sup>1\*</sup> , Ali Mohebbi, MSc<sup>1,2\*</sup> ,  
Henrik Bengtsson, MBA<sup>2</sup>, Morten Lind Jensen, MD, PhD<sup>3</sup>,  
and Morten Mørup, PhD<sup>1</sup>

## 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 with simple feature-engineered ML classification models for adherence detection. It was further investigated whether fused expert-dependent and automatically learned features could improve performance, resulting in a comparison of six 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 three 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 two 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.

## Keywords

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

## Introduction

Diabetes has become a major public health problem worldwide with an estimate of 463 million people diagnosed and 232 million still undiagnosed.<sup>1</sup> Type 2 diabetes (T2D) is more prevalent than type 1 diabetes (T1D) and accounts for ~90% of the diagnosed.<sup>2</sup> Elevated plasma glucose (PG) is caused by this chronic disease due to insufficient insulin production or decreased insulin sensitivity ( $S_i$ ). 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

<sup>1</sup>Department of Applied Mathematics and Computer Science, DTU Compute, Kgs. Lyngby, Denmark

<sup>2</sup>Novo Nordisk A/S, Device R&D, Hillerød, Denmark

<sup>3</sup>Novo Nordisk A/S, Medical & Science, Søborg, Denmark

\*These authors have contributed equally to this work.

### Corresponding Author:

Morten Mørup, PhD, Danmarks Tekniske Universitet, Richard Petersens Plads, Building 321, 2800 Kgs., Lyngby, Denmark.  
Email: mmor@dtu.dk

patients.<sup>3-6</sup> However, despite the minimal requirement of once-daily injections, studies report high variance in adherence levels.<sup>7-9</sup>

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.<sup>6,10-16</sup> CGM provides the ability to continuously monitor interstitial glucose levels with a pre-defined interval (usually between 5 and 15 minutes). Even though the use of CGM devices is currently mainly used in T1D, there is a clear expansion and progress toward CGM use in the treatment of T2D as well.<sup>10,13</sup>

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 (ie, adherence level, intraday variation of  $S_p$ , intersubject variance in level of consumed carbohydrate [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.<sup>17,18</sup>

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.<sup>19,20</sup> In particular, DL methods have proven powerful for automatic feature extraction using large data sets to reduce the need for feature engineering by domain experts.<sup>20</sup> 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, that is, 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® states that a delayed dose should be taken within eight hours of the next injection.<sup>21</sup>

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 nonadherence to prescribed treatment and higher mortality and hospitalization.<sup>5</sup> Therefore, an early alarm system designed to detect adherence/nonadherence 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). Additionally a fused approach were considered, in which engineered and learned features were combined and explored for two different network architectures. This resulted in a comparison of six different detection models in total.

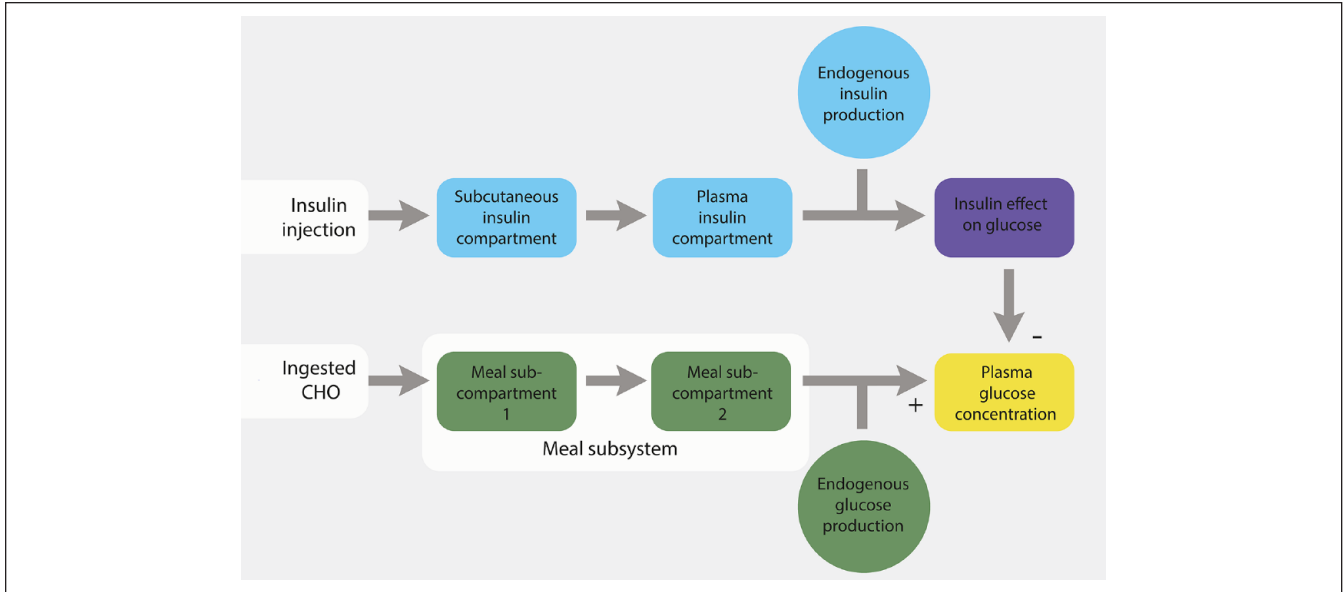
The speed and precision of adherence detection were explored after an injection of once-daily basal insulin using the simulated CGM data. Our recent preliminary study by Mohebbi et al<sup>22</sup> indicated the feasibility and potential of such an approach which is further evaluated and improved upon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

In the simulations, adherence/nonadherence was 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 nonadherence was defined as a total dose omission.

### MVP Model

CGM data of T2D patients with labeled adherence/nonadherence 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.<sup>17,18</sup> A schematic of the utilized compartment model is presented and described in Figure 1.

In short, the model is based on the MVP model,<sup>17</sup> a physiological model for simulating 24-hour insulin-glucose dynamics in T1D patients along with quantified parameter sets for ten patients. Moreover, a T2D augmentation was implemented by Aradóttir et al,<sup>18</sup> introducing a linear endogenous insulin production and a two-compartment model describing absorption of CHOs.<sup>17,18,23,24</sup> Furthermore, relevant parameters were varied in accordance with typical physiological differences between T1D and T2D patients including body weight,  $S_p$ , and fasting glucose.<sup>18,22</sup> The full set of coupled differential equations describing the compartment model and further details are presented by Aradóttir et al.<sup>18</sup>



**Figure 1.** Schematic representation of the compartment model used to simulate CGM data. The ingested CHO affect the PG concentration through meal subcompartments 1 and 2 along with endogenous glucose production. The lowering effect of insulin on PG is the sum of both the subcutaneously injected insulin and the endogenous insulin production. The final PG is denoted by the plasma glucose concentration compartment from where CGM values are acquired.

### 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 Figure 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 (Figure 2(a)). All data in this article 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 8:00 AM  $\pm$  15 minutes. The standard deviation (SD) of 15 minutes was added to the TOI by adding a random number from a Gaussian distribution (Figure 2(b)). If a day was nonadherent the dose was fully omitted, and no actions were taken at the expected TOI.

#### c) Intraday variation of $S_I$

We implemented an intraday variance in  $S_I$  in order to achieve a more realistic CGM dynamic (Figure 2(c)). The daily excursions of  $S_I$  presented by Schulz et al<sup>25</sup> were extracted and confined to be within the diurnal and nocturnal subject-specific  $S_I$  values presented by Kanderian et al.<sup>17</sup> Additionally, noise was added to the

$S_I$  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 intraday  $S_I$  (Figure 2(c)).

#### d) Intersubject variance in level of consumed CHO

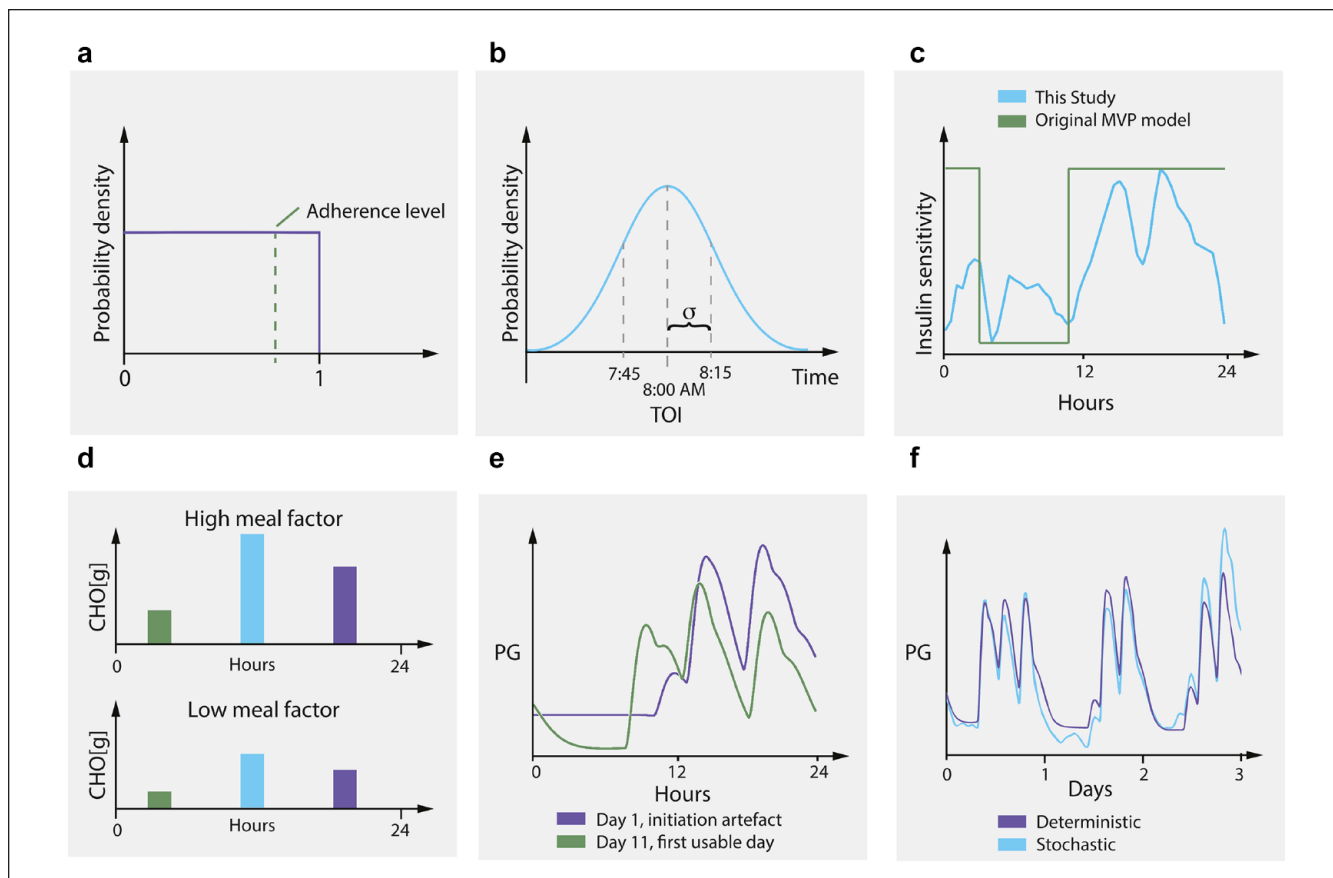
To introduce an intersubject variance in level of consumed CHO, a meal factor was implemented (Figure 2(d)). The meal factor was chosen at random before each simulation and multiplied on all meals each day. The factor was limited such that the daily CHO intake was below the upper boundary of ingested CHO presented by Kanderian et al<sup>17</sup> and above the recommended minimum intake of CHO for T2D patients.<sup>26</sup>

#### 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 (Figure 2(e)). 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 intraday variance in the simulated data when compared with clinical data. A higher degree of variance was achieved in the simulated CGM data by applying the Euler-Maruyama method<sup>27</sup> introducing stochastic noise during simulations (Figure 2(f)).



**Figure 2.** Conceptual plots illustrating six additions to the simulation setup. (a) Random adherence sequence, (b) variance in time of injection, (c) intraday variation of insulin sensitivity, (d) intersubject variance in level of consumed carbohydrate, (e) removal of simulation initiation bias, and (f) introduction of a stochastic simulation scheme.

In order to simulate CGM signals of a T2D cohort, the parameter variation and meal case structure applied by Mohebbi et al<sup>22</sup> was used: six fasting glucose ( $G_b$ ) levels between  $\sim 108$  and  $198$  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 nine out of ten 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.<sup>18,22</sup> 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.<sup>28</sup> 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  $\sim 25$  hours.<sup>21</sup> A pharmacokinetic (PK) profile of Insulin Degludec described by Heise et al was used in the simulations.<sup>29</sup> Prior to each simulation, the patient-specific optimal dose was

determined using a stepwise titration algorithm presented in a study by Zinman et al.<sup>30</sup> Dose adjustments were based on the mean fasting PG of three days above target, or the minimum fasting PG if below target. A patient was considered in target if PG was between  $\sim 70$  and  $90$  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 with clinical data. The coefficient of variance (CV) of the prebreakfast PG in the simulated CGM data were compared with the CV of prebreakfast SMPG measurements from a large clinical study containing 770 patients with T2D on once-daily Tresiba® treatment.<sup>30</sup> For the simulated CGM data the prebreakfast PG was defined as the lowest mean hour (lowest mean acquired from a sliding window of 12 consecutive PG values),<sup>31</sup> between 6:00 AM and 9:00 AM. The prebreakfast PG CV of the clinical study was 17% in comparison with 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 data sets 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 *nonadherence* as output. The training error used is the *cross-entropy error function* defined by:

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

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*:

$$\hat{y}_n = \hat{y}_n(\hat{z}_n) = \frac{1}{1 + e^{-z_n}} \quad (2)$$

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 Table 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 nonlinear activation functions we employed *rectified linear units*.<sup>32</sup>

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 nonconvex with issues of local minima. The minimum was found using the *Adam* optimizer based on stochastic gradient optimization.<sup>33</sup> 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 was based on 11 engineered features of the CGM signal described in Table 1. Out of the 11 features, 7 were chosen with inspiration from the key metrics from the international consensus on the use of CGM<sup>34</sup> and the remaining 4 were lowest mean hour, pre-breakfast PG, and minimum and maximum measurements.

Considering the input to the feature-based architectures, the first ten features were computed every four hours throughout the day of classification (DOC) (midnight to 4:00 AM, 4:00 AM to 8:00 AM, . . . , 8:00 PM to midnight),

**Table 1.** List of Engineered CGM Features Used to Detect Adherence and Nonadherence.

Features
Minimum PG measure of the interval
Maximum PG measure of the interval
Mean of entire interval
SD of the interval
Percent of interval with PG above 90 mg/dL (5 mmol/L)
Percent of interval with PG above 108 mg/dL (6 mmol/L)
Percent of interval with PG above 126 mg/dL (7 mmol/L)
Percent of interval with PG above 144 mg/dL (8 mmol/L)
Area under the PG measures in the interval
Lowest mean hour of the interval
Lowest mean hour between 6:00 AM and 9:00 AM*

Abbreviations: CGM, continuous glucose monitoring; PG, plasma glucose; SD, standard deviation.

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

resulting in a total of six feature sets ( $6 \cdot 10$  features). In the same manner, historical data were also included to the input with the four-hour interval feature sets acquired from one to five 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 6:00 AM and 9:00 AM. For a clearer understanding, Equation (3) is included showing an example of input for adherence detection on a DOC at 8:00 AM:

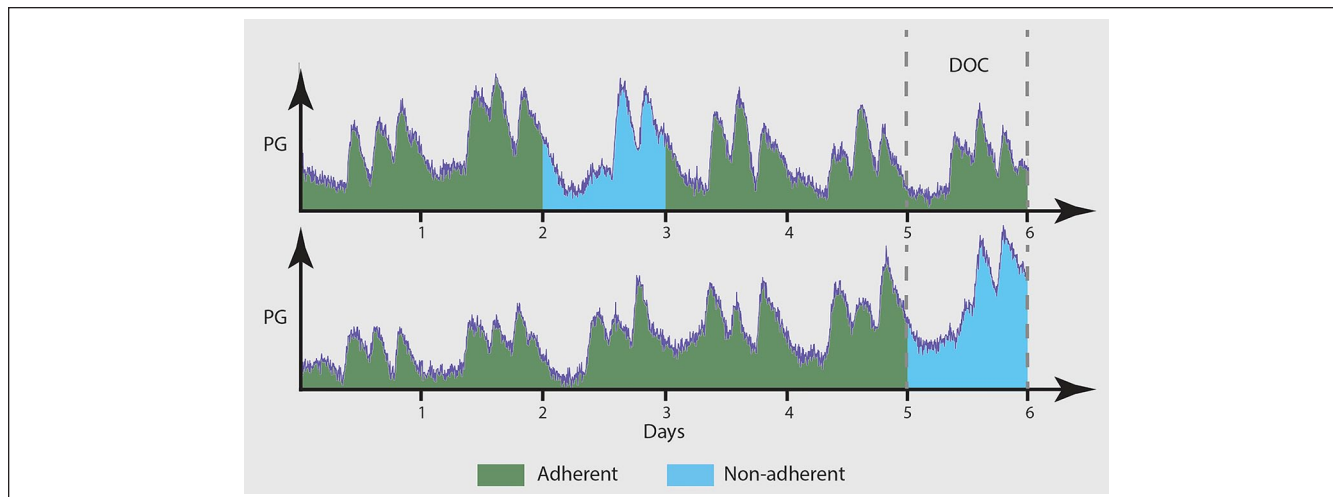
$$\begin{aligned} \text{DOC}_{8\text{AM}} &= \text{days}_{\text{prior}} \cdot \left( \begin{array}{l} \text{features}_{24\text{h,prior}} \\ +6 \cdot \text{features}_{4\text{h,prior}} \end{array} \right) \\ &+ 2 \cdot \text{features}_{4\text{h,DOC}} \\ &= 5 \cdot (11 + 6 \cdot 10) + 2 \cdot 10 \\ &= 375 \text{ input features} \end{aligned} \quad (3)$$

Similarly, a new feature set was added to the input data for each added four-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 five minutes). As for the feature models, CGM readings from five 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 takes the entire feature input as described above in addition to the raw CGM signal. It was chosen to include the previous five 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





**Figure 3.** An illustrative example of an adherent and nonadherent day including the five days prior to the day of classification.

831 060 nonadherent and adherent days, respectively. It was desired to have class-balance in the data set to have a baseline accuracy of 50%. Therefore, 43 740 of the adherent days were selected at random resulting in a data set of  $2 \cdot 43\,740 = 87\,480$  CGM observations with a forced class-balance. Figure 3 presents examples of an adherent and nonadherent CGM observation including DOC and the five prior days of CGM data used for detection.

### Architectures and Hyperparameters

Figure 4 shows the architectures investigated in this study. A progression in model complexities was examined as shown in Figure 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 circles on the left of each architecture indicate CGM feature input or raw CGM input.

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<sup>22</sup> 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 performs 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

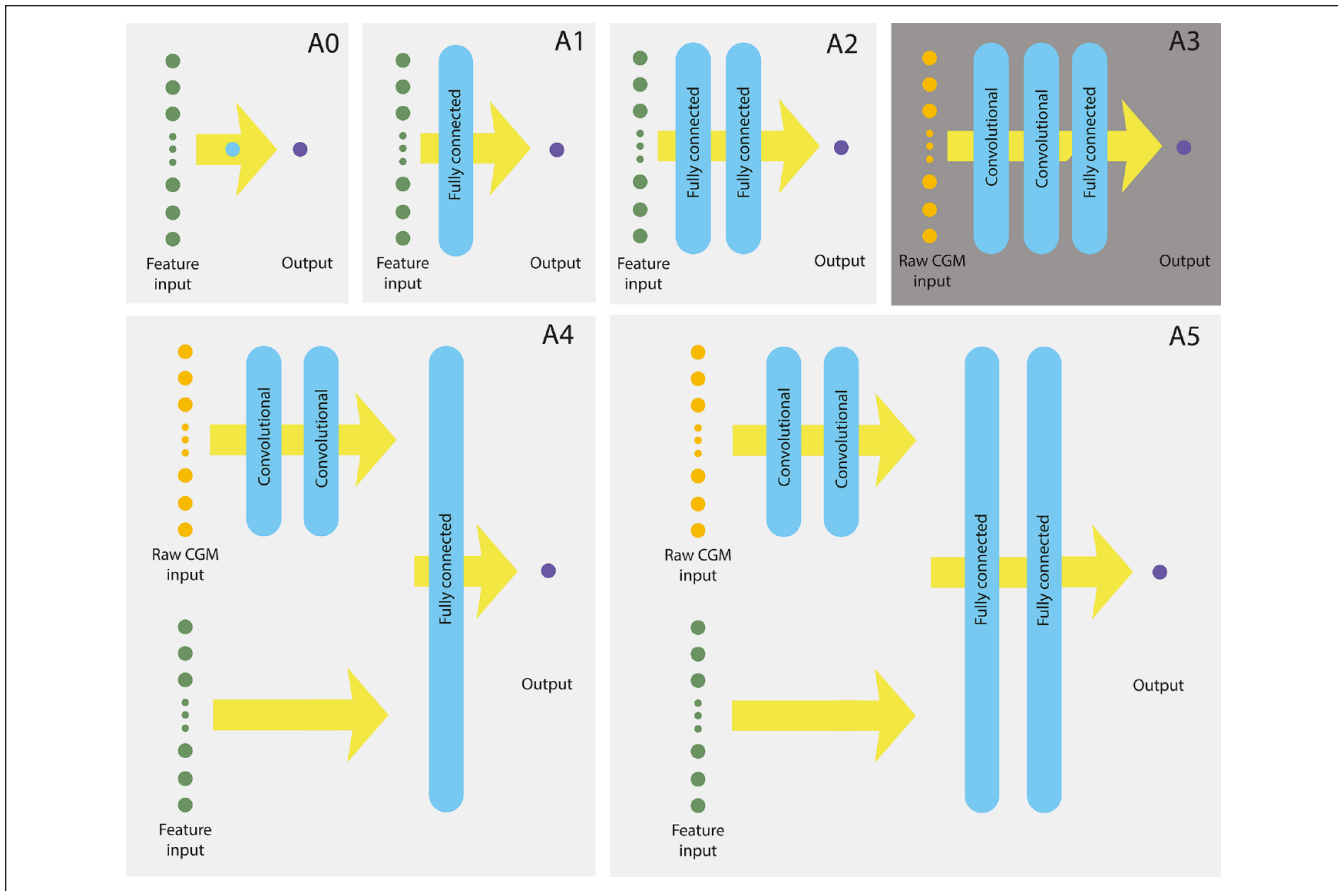
During training of the models, the simulated data from subjects 1 to 6 were used in a leave-one-subject-out cross-validation setup as indicated in Figure 5. For each fold the model yielding the best validation accuracy was chosen based on five re-initializations and training of model parameters. CGM data from subjects 7 to 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 six models obtained during the cross-validation step. Furthermore, four consecutive runs (repeating the random selection of adherent days from the data set followed by the process illustrated in Figure 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)  $\pm$  SD across the four repetitions. For a single repetition, accuracy is defined by Equation (4):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{N}, \quad (4)$$

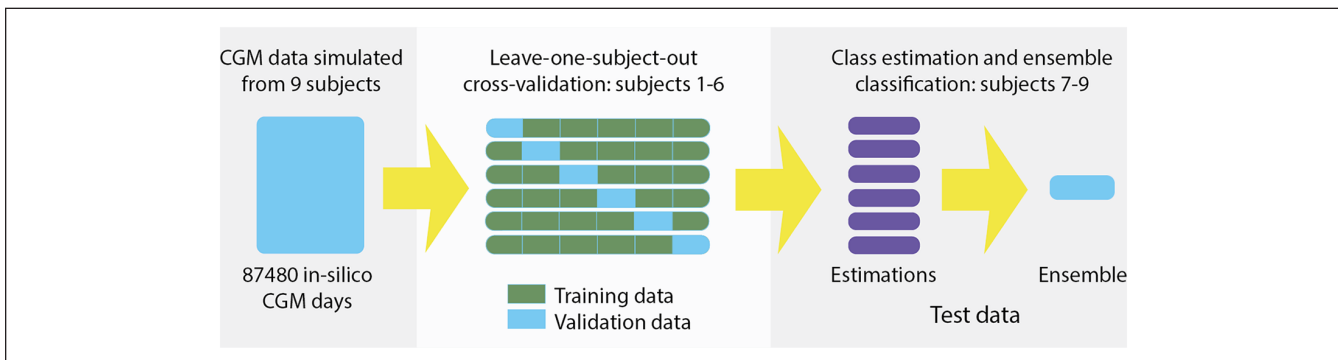
where TP is the true positives, TN is the true negatives, and N is the total number of observations.

### Statistical Considerations

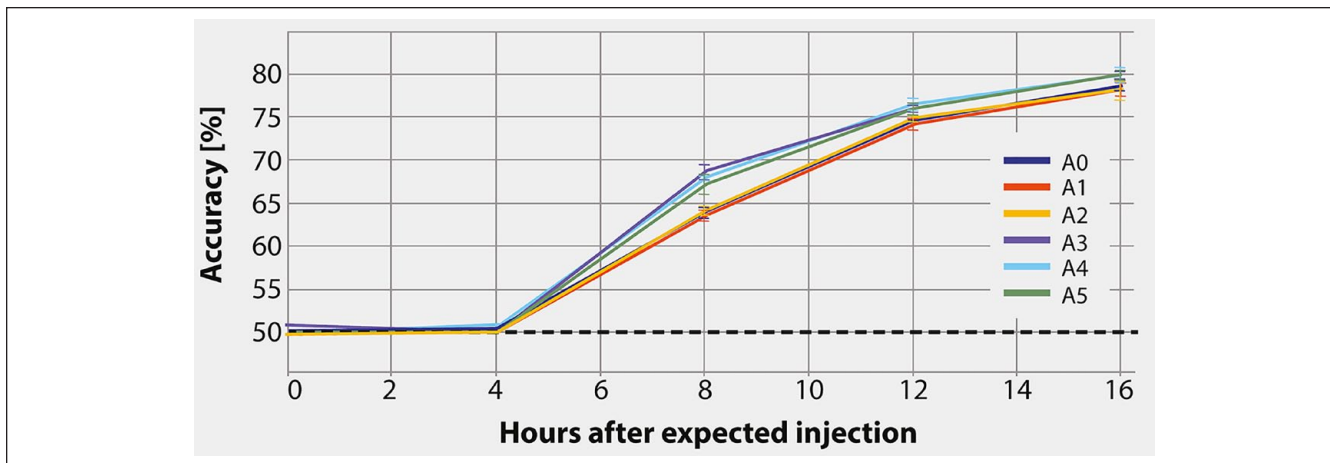
Due to the exploratory scope and nature of this study, comparisons are descriptive. Although formal statistical testing could be employed, the present study is based on simulated data for which arbitrary power can be achieved by additional simulations. Furthermore, the validity of the simulated data and how the results generalize to real CGM data can be questioned. The



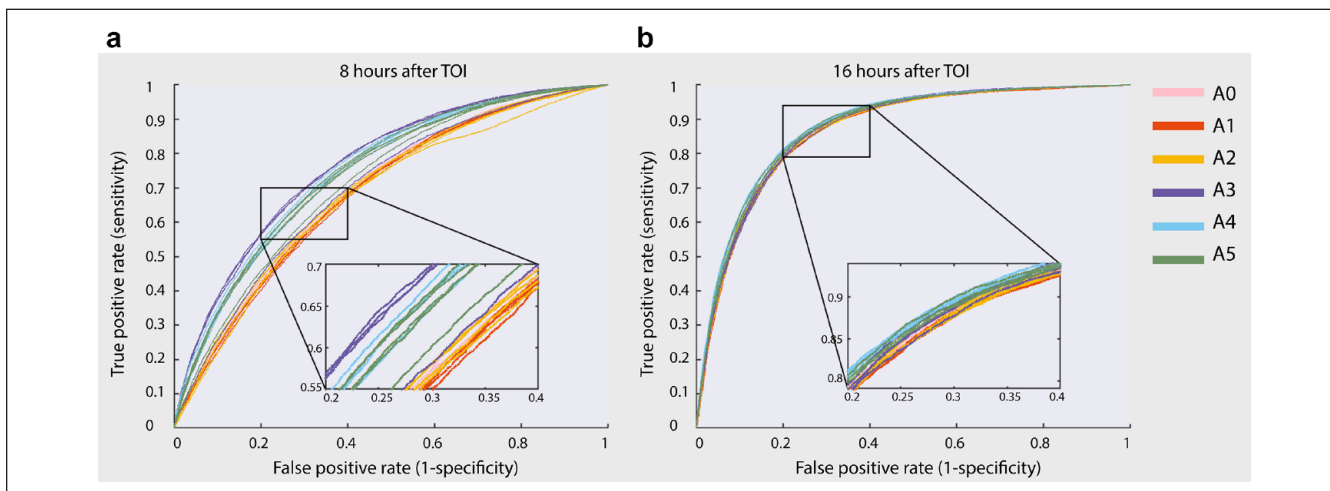
**Figure 4.** High-level schematic of the six architectures investigated in this study. The circles on the left of each architecture indicate CGM feature input or raw CGM input. The single circle to the right of each architecture illustrate the output neuron. A0 is a logistic regression model; A1 and A2 are multilayer perceptrons with 1 and 2 hidden layers, respectively; A3 is the convolutional neural network (CNN) from Mohebbi et al<sup>22</sup>; and lastly A4 and A5 are a combination of automatic and expert-dependent feature extraction with, respectively, 1 and 2 hidden layers before the output. The CNN-related hyperparameters are as follows: eight filters in each CNN layer followed by max-pooling layer of dimension [1, 2], local receptive field of length 18 (corresponding to 1.5 hours), and a stride length of 2. For all the architectures, the mini-batch size was chosen to be 250 including a total of 10 epochs for each training session.



**Figure 5.** Schematic presentation of how the simulated data were used to train, validate, and test the models. All simulations were divided into nine parts corresponding to the nine subjects. Simulations from subjects 1 to 6 were used in leave-one-subject-out cross-validation scheme used for tuning of learning rates for all architectures and number of hidden units (NHU) in the fully connected (FC) layers (no regularization employed) of each architecture excluding A3 as NHU was optimized by Mohebbi et al.<sup>22</sup> The parameters were optimized in a sequential manner with: 1) learning rates in 4 steps between  $10^{-2}$  and  $10^{-5}$  with the NHU fixed to 10, followed by 2) optimization of NHU in FC layers in 5 steps between 5 and 100. Optimal learning rate was found to be  $10^{-3}$  for all architectures and the NHU to be 5 in the FC layers for all architectures excluding A3 where NHU were 10. Simulations from subjects 7 to 9 were used as test data, and for each day, six class probabilities were calculated using the models originated from the cross-validation steps and averaged, before making an ensemble classification.



**Figure 6.** Performance of the six architectures at expected time of injection in addition to 4, 8, 12, and 16 hours after. Each data point represents the META  $\pm$  SD of the four runs, whereas the 50% accuracy (dashed black line) is the baseline performance. META, mean ensemble test accuracy; SD, standard deviation.



**Figure 7.** Presentation of 24 (4 runs of each of the 6 models) receiver operator curves for (a) 8 hours and (b) 16 hours after time of injection.

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 nonadherence 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. Figure 6 presents a plot of the META of the architectures with increasing data available after the expected TOI.

Considering Figure 6 the performances based on accuracy are very similar 16 hours after TOI. However, the architectures utilizing the automatic feature extraction (ie, 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 (ie, at noon) can be an artifact caused by the confined TOI always being followed by postprandial PG excursions following breakfast and lunch, thus concealing the signal caused by an injection or omission.

Figure 7 depicts the 24 (4 runs of each of the 6 models) receiver operator curves (ROCs) for 8 hours (Figure 7(a))



**Table 2.** Mean  $\pm$  SD of the AUROC and META for Each of the Detection Models at 8 and 16 hours After Expected TOI (Based on Four Runs).

Model	8 hours after TOI		16 hours after TOI	
	META	AUROC	META	AUROC
	Mean $\pm$ SD	Mean $\pm$ SD	Mean $\pm$ SD	Mean $\pm$ SD
A0	63.9% $\pm$ 0.7%	0.691 $\pm$ 0.003	78.6% $\pm$ 0.6%	0.869 $\pm$ 0.001
A1	63.9% $\pm$ 0.5%	0.687 $\pm$ 0.004	78.2% $\pm$ 0.8%	0.864 $\pm$ 0.001
A2	64.0% $\pm$ 0.5%	0.688 $\pm$ 0.013	78.3% $\pm$ 1.1%	0.867 $\pm$ 0.002
A3	68.6% $\pm$ 0.9%	0.765 $\pm$ 0.005	79.7% $\pm$ 0.4%	0.870 $\pm$ 0.002
A4	68.1% $\pm$ 0.3%	0.745 $\pm$ 0.004	79.7% $\pm$ 0.8%	0.878 $\pm$ 0.001
A5	67.6% $\pm$ 1.2%	0.737 $\pm$ 0.015	79.8% $\pm$ 0.5%	0.873 $\pm$ 0.002

Abbreviations: AUROC, area under the receiver operator curve; META, mean ensemble test accuracy; SD, standard deviation; TOI, time of injection.

and 16 hours (Figure 7(b)) after TOI. These ROCs support the results presented in Figure 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 eight 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 (ie, A0, A1, and A2). As we detect an injected dose of insulin, 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 Table 2 including META and the average area under the receiver operator curve 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 four to eight 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 with 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  $\sim$ 80% within DOC (ie, 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 four to eight 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; that is, 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 (eg, 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 intraday  $S_1$  variance, and not interday variation of the  $S_1$  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 eight hours after TOI due to the potential clinical implications. Detecting a missed injection already after eight 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.<sup>21</sup>

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 health care 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  $79.8\% \pm 0.5\%$  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<sup>11,34</sup> simple feature-engineered models at 16 hours after TOI with accuracy of  $78.6\% \pm 0.6\%$ . However, eight 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,<sup>21</sup> the injection time can be flexible, as long as the injection is taken at least eight hours before next expected injection. This window of adherence detection is within the time window where a clinically relevant treatment recommendation can be provided.

## Acknowledgments

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.

## Declaration of Conflicting Interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iDs

Daniel N. Thyde  <https://orcid.org/0000-0002-0391-0646>

Ali Mohebbi  <https://orcid.org/0000-0001-7705-9332>

## References

1. Atlas & IDF Diabetes. *IDF Diabetes Atlas*. 9th ed. 2019. <https://www.diabetesatlas.org/en/>. Accessed November 29, 2019.
2. Atlas & IDF Diabetes. *International Diabetes Federation: Type 2 Diabetes*. 2019. <https://www.idf.org/aboutdiabetes/type-2-diabetes.html>. Accessed October 8, 2019.
3. Davies MJ, D'Alessio DA, Fradkin J, et al. Management of hyperglycemia in type 2 diabetes, 2018. A consensus report by the American diabetes association (ADA) and the European association for the study of diabetes (EASD). *Diabetes Care*. 2018;41(12):2669-2701.
4. Lee WC, Balu S, Cobden D, Joshi AV, Pashos CL. Medication adherence and the associated health-economic impact among patients with type 2 diabetes mellitus converting to insulin pen therapy: an analysis of third-party managed care claims data. *Clin Ther*. 2006;28(10):1712-1725.
5. Sarbacker GB, Urteaga EM. Adherence to insulin therapy. *Diabetes Spectr*. 2016;29(3):166-170.
6. Klonoff DC, Blonde L, Cembrowski G, et al. Consensus report: the current role of self-monitoring of blood glucose in non-insulin-treated type 2 diabetes. *J Diabetes Sci Technol*. 2011;5(6):1529-1548.
7. Yavuz DG, Ozcan S, Deyneli O. Adherence to insulin treatment in insulin-naïve type 2 diabetic patients initiated on different insulin regimens. *Patient Prefer Adherence*. 2015;9:1225-1231.
8. Peyrot M, Barnett AH, Meneghini LF, Schumm-Draeger PM. Insulin adherence behaviours and barriers in the multinational global attitudes of patients and physicians in insulin therapy study. *Diabet Med*. 2012;29(5):682-689.
9. Cramer JA. A systematic review of adherence with medications for diabetes. *Diabetes Care*. 2004;27(5):1218-1224.
10. Klonoff DC, Ahn D, Drincic A. Continuous glucose monitoring: a review of the technology and clinical use. *Diabetes Res Clin Pract*. 2017;133:178-192.
11. Battelino T, Danne T, Bergenstal RM, et al. Clinical targets for continuous glucose monitoring data interpretation: recommendations from the international consensus on time in range. *Diabetes Care*. 2019;42(8):1593-1603.
12. Bergenstal RM, Ahmann AJ, Bailey T, et al. Recommendations for standardizing glucose reporting and analysis to optimize clinical decision making in diabetes: the ambulatory glucose profile (AGP). *Diabetes Technol Ther*. 2013;15(3):198-211.
13. Umpierrez GE, Klonoff DC. Diabetes technology update: use of insulin pumps and continuous glucose monitoring in the hospital. *Diabetes Care*. 2018;41(8):1579-1589.

14. Sheng T, Offringa R, Kerr D, et al. Diabetes healthcare professionals use multiple continuous glucose monitoring data indicators to assess glucose management. *J Diabetes Sci Technol.* 2018;14:271-276.
15. Bergenstal RM. Continuous glucose monitoring: transforming diabetes management step by step. *Lancet.* 2018;391(10128):1334-1336.
16. Scheiner G. CGM retrospective data analysis. *Diabetes Technol Ther.* 2016;18(suppl 2):S214-S222.
17. Kanderian SS, Weinzimer S, Voskanyan G, Steil GM. Identification of intraday metabolic profiles during closed-loop glucose control in individuals with type 1 diabetes. *J Diabetes Sci Technol.* 2009;3(5):1047-1057.
18. Aradóttir TB, Boiroux D, Bengtsson H, Kildegaard J, Orden BV, Jørgensen JB. Model for simulating fasting glucose in type 2 diabetes and the effect of adherence to treatment. *IFAC-PapersOnLine.* 2017;50(1):15086-15091.
19. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nat Med.* 2019;25(1):24-29.
20. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521(7553):436-444.
21. Food and Drug Administration. Tresiba® label. 2015. [https://www.accessdata.fda.gov/drugsatfda\\_docs/label/2015/203314lbl.pdf](https://www.accessdata.fda.gov/drugsatfda_docs/label/2015/203314lbl.pdf). Accessed November 29, 2019.
22. Mohebbi A, Aradóttir TB, Johansen AR, Bengtsson H, Fraccaro M, Mørup M. A deep learning approach to adherence detection for type 2 diabetics. Paper presented at: Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS); July 11-15, 2017:2896-2899; Seogwipo, South Korea: Institute of Electrical and Electronics Engineers Inc. doi:10.1109/EMBC.2017.8037462
23. Hovorka R, Canonico V, Chassin LJ, et al. Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes. *Physiol Meas.* 2004;25(4):905-920.
24. Ruan Y, Thabit H, Wilinska ME, Hovorka R. Modelling endogenous insulin concentration in type 2 diabetes during closed-loop insulin delivery. *Biomed Eng Online.* 2015;14(1):19.
25. Schulz B, Ratzmann KP, Albrecht G, Bibergeil H. Diurnal rhythm of insulin sensitivity in subjects with normal and impaired glucose tolerance. *Exp Clin Endocrinol.* 1983;81(3):263-272.
26. American Diabetes Association, Bantle JP, Wylie-Rosett J, et al. Nutrition recommendations and interventions for diabetes: a position statement of the American Diabetes Association. *Diabetes Care.* 2008;31(suppl 1):S61-S78.
27. Higham DJ. An algorithmic introduction to numerical simulation of stochastic differential equations. *Siam Rev.* 2001;43(3):525-546.
28. Facchinetti A, Del Favero S, Sparacino G, Castle JR, Ward WK, Cobelli C. Modeling the glucose sensor error. *IEEE Trans Biomed Eng.* 2014;61(3):620-629.
29. Heise T, Hermanski L, Nosek L, Feldman A, Rasmussen S, Haahr H. Insulin degludec: four times lower pharmacodynamic variability than insulin glargine under steady-state conditions in type 1 diabetes. *Diabetes Obes Metab.* 2012;14(9):859-864.
30. Zinman B, Philis-Tsimikas A, Cariou B, et al. Insulin degludec versus insulin glargine in insulin-naive patients with type 2 diabetes: a 1-year, randomized, treat-to-target trial (BEGIN Once Long). *Diabetes Care.* 2012;35(12):2464-2471.
31. Aradóttir TB, Bengtsson H, Jensen ML, et al. Feasibility of a new approach to initiate insulin in type 2 diabetes [published online ahead of print January 15, 2020]. *J Diabetes Sci Technol.* doi:10.1177/1932296819900240
32. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Paper presented at: Proceedings of the 25th International Conference on Neural Information Processing Systems; December 2012: 1097-11059; Lake Tahoe, NV.
33. Kingma DP, Ba J. Adam: a method for stochastic optimization. Paper presented at: ICLR May 7-9, 2015; San Diego, CA.
34. Danne T, Nimri R, Battelino T, et al. International consensus on use of continuous glucose monitoring. *Diabetes Care.* 2017;40(12):1631-1640.