Therapy-driven Deep Glucose Forecasting

Eleonora Maria Aiello[a], Giuseppe Lisanti[b], Lalo Magni[c], Mirto Musci[a], Chiara Toffanin[a,*]

[a] *University of Pavia, Department of Electrical, Computer and Biomedical Engineering,via Ferrata 5, Pavia 27100, Italy.*
[b] *University of Bologna, Department of Computer Science and Engineering, Mura Anteo Zamboni 7, Bologna 40126, Italy.*
[c] *University of Pavia, Department of Civil and Architecture Engineering, via Ferrata 5, Pavia 27100, Italy.*

Abstract

The automatic regulation of blood glucose for Type 1 diabetes patients is the main goal of the artificial pancreas, a closed-loop system that exploits continue glucose monitoring data to define an optimal insulin therapy. One of the most successful approaches for developing the artificial pancreas is the model predictive control, which exhibits promising results on both virtual and real patients. The performance of such controller is highly dependant on the reliability of the glucose-insulin model used for prediction purpose, which is usually implemented with classic mathematical models. The main limitation of these models consists in the difficulties of modeling the physiological non linear dynamics typical of this system. The availability of big amount of *in silico* and *in vivo* data moved the attention to new data-driven methods which are able to easily overcome this problem. In this paper we propose Deep Glucose Forecasting, a deep learning approach for forecasting glucose levels, based on a novel, two-headed Long-Short Term Memory implementation. It takes in input the previous values obtained through continue glucose monitoring, the carbohydrate intake, the suggested insulin therapy and forecasts the interstitial glucose level of the patient. The proposed architecture has been trained on 100 virtual adult patients of the UVA/Padova simulator, and tested on both virtual and real patients. The proposed solution is able to generalize to new unseen data, outperforms classical population models and reaches performance comparable to classical personalized models when fine-tuning is exploited on real patients.

Keywords: diabetes, forecasting, prediction, deep learning, LSTM

1. Introduction

Type 1 Diabetes (T1D) is a chronic metabolic disease characterized by high Blood Glucose (BG) level, known as hyperglycaemia. Hyperglycaemia can cause long-term complications including damage to blood vessels, eyes, kidneys, and nerves and it is caused by the dysfunction of pancreatic β-cells responsible for the

*Corresponding author
Email addresses: eleonoramaria.aiello01@universitadipavia.it (Eleonora Maria Aiello), giuseppe.lisanti@unibo.it (Giuseppe Lisanti), lalo.magni@unipv.it (Lalo Magni), mirto.musci@unipv.it (Mirto Musci), chiara.toffanin@unipv.it (Chiara Toffanin)

production of insulin. This hormone regulates the BG concentration by allowing cells and tissues to absorb glucose from the bloodstream. T1D patients need exogenous insulin injections to keep the glucose concentration in the euglycemic range. Their goal is to minimize diabetes complications related to hyperglycemia and simultaneously avoid hypoglycemia, a condition that could be caused by excessive insulin administration. The automatic regulation of the BG concentration for people affected by T1D through exogenous insulin administrations [1, 2, 3] is the main purpose of the so-called artificial pancreas. The artificial pancreas is a closed-loop system that exploits the glucose measurements obtained via Continuous Glucose Monitor (CGM) to compute and automatically deliver the proper amount of insulin via subcutaneous insulin pump. The core of the artificial pancreas is the control algorithm that defines the optimal insulin amount to infuse. The Model Predictive Control (MPC) resulted into one the most promising approach to this problem in the last years, obtaining successful results both *in silico* and *in vivo* [4, 5, 6, 7, 8, 9, 10]. The MPC approach exploits a glucose-insulin model to forecast the BG values in order to compute the optimal insulin therapy. For this reason, the predictive performance of the model plays a key role in the overall control performance. Classical mathematical model used in these applications are not able to fully describe the nonlinear glucose-insulin dynamics. In order to overcome this limitation, the complexity of the model has to be increased and new effective identification techniques are required. Recently, a branch of the research was moved towards new identification techniques in order to have more effective models to be used for both the control algorithms and the safety systems. A complete review can be found in [11]. Data-driven approaches have been successfully applied to real-life applications [12, 13, 14, 15]. Depending on the task at hand, the aim of these approaches is to learn a model directly from the data. Thanks to the availability of a huge amount of data collected during long-period trials in free-living conditions new data-driven approaches have also been studied in the artificial pancreas research field, with promising results [16, 17]. However, their performance are limited by the use of a fixed and simple structure of the chosen model. Data-driven approaches based on deep learning architecture have received an increasing attention in the last few years mainly because of the remarkable performance obtained in several research fields [18, 19]. Among these approaches, recurrent neural networks represent a family of deep learning architectures which have been explicitly designed to model the evolution over time of a phenomenon. In particular, given an input composed of a sequence of observations from a signal, such as the BG level in our scenario, these models try to predict its future value or values.

The main goal of this work is the development of a new forecasting model able to predict the future BG of a patient subject to several possible insulin treatments in order to define his/her optimal future insulin therapy. In this perspective, we propose a deep learning architecture which is able to forecast the BG level of T1D patients. Our architecture is composed of two models, one observing the CGM measurements, insulin injections and carbohydrate intakes up to a given time t and a second model that receives as input the future insulin that will be administered to the patient and the future carbohydrates that he/she will assume.

Both models are composed of stacked Long-Short Term Memory (LSTM) [20] networks. The output of the two models is combined and given as input to a Fully Connected (FC) layer which is used to predict the future values of the Interstitial Glucose (IG), considering a fixed prediction horizon. Training is performed in a supervised fashion on a subset of identities, separated from those that will be considered as test, in order to obtain a model which is able to generalize to new unseen data. The proposed architecture obtains state-of-the-art performance on both *in silico* and *in vivo* data, considering several prediction horizons.

This paper has two main contributions with respect to the state of art: 1) as far as we know [21, 22, 23, 24, 25], this is the first work that introduces a deep learning architecture that exploits two models composed of LSTM networks for a therapy driven approach; 2) the proposed architecture is shown to be able to generalize on *in vivo* population of real patients, even if it is trained only on *in silico* data.

2. Related Works

Several works treated the model identification problem in the last years by exploiting *in silico* [26, 27, 28, 29, 30, 31, 32, 33, 34, 35] and *in vivo* [16, 17, 36] data. For a comprehensive literature review please refer to [11, 37]. The *in silico* data used for model identification [26, 27, 31, 33] were obtained through realistic closed-loop clinical protocols simulated via the UVA/Padova simulator [38, 39] in order to produce a sufficient input-output excitation for identification purpose. This simulator is equipped with a cohort of virtual patients and represents a powerful tool for the design and the test of new insulin therapies since it has been accepted by Food and Drug Administration (FDA) as a substitute to animals trials, making the control algorithms directly testable on real patients. On the other hand, *in vivo* data were collected during either short and controlled trials on hospitalized patients [27] and trials outside the hospital environment in free-living conditions [4, 5, 6, 7, 8, 40].

Classical models can be either nonlinear or linear: the nonlinear models usually allow for a better approximation of the underlying dynamic, but being based on partial differential equations, they are extremely costly from a computational point-of-view. On the other hand, the linear model approximation is less precise, but their computational load is limited.

Recently, several MPC algorithms have been developed using different type of models. Among them, the MPC proposed in [41] used a nonlinear models whose parameters were re-estimated at each control step in order to take into account the daily variability of the patient. This MPC was tested in several clinical trials [42, 43, 44, 45, 46] with promising results. The Zone-MPC [47] was based on the third-order, discrete-time, linear time-invariant model proposed in [48]. Although the simple structure of this model, the Zone-MPC obtained good results *in vivo* on both adults and adolescents [49, 50]. The MPC described in [51] (referred in the following with MPC-P) was synthesized on the basis of the average linear time-invariant metabolic model computed by averaging the model parameters of the virtual population of the UVA/Padova

simulator. The MPC-P was tested in several clinical trials [40, 52] obtaining good results thanks to capability of the simulator to represent all the diabetic population and not only a restricted subgroup of individuals. However, a model with better predictive capabilities would increase the overall control performance. In this paper we design a nonlinear model based on deep learning techniques to be included in the MPC-P and we analyze its prediction capabilities. The proposed model is then compared to the linearized average model (AVG) currently used in the MPC-P. It is important to stress out that in this kind of applications, the real data collected during clinical trials are usually not publicly released, thus the choice of MPC-P is critical for the availability of data.

2.1. Solutions based on Neural Networks

The first solution exploiting neural networks for modeling the BG metabolism of a T1D patient was proposed in [53]. In particular, the authors tried to predict the glucose level of a diabetic patient by training a Recurrent Neural Network (RNN) architecture which receives insulin levels, meals and level of exercise as inputs, alongside current and previous estimates of BG. However, the data used for both train and test is acquired from a single patient and this may result in a lack of generalization for the final model.

Recently, a few solutions exploiting deep learning techniques for glucose level prediction in diabetic patients have been proposed [21, 22, 23, 24, 25]. Similarly to [53], Allam et. al. [21] proposed to use CGM signals to train a RNN for predicting future values of the glucose concentration, considering several prediction horizons. Again, the data used for both train and test were selected from the same population, which may result in a model that hardly generalize to new unseen data.

LSTM networks achieve state-of-the-art performance in modeling several time-dependent phenomena. For this reason, the authors of [22] proposed to exploit LSTM in a model which takes CGM values, insulin dosages and carbohydrate intake as inputs and tries to predict the glucose level at prediction horizons of 30 and 60 minutes. Data incoming from four patients acquired using different CGM devices has been used in both train and test. Unfortunately, the training needed to be repeated multiple times, mainly because of initialization issues which left the optimization stuck in a bad local optima. An LSTM-based architecture has also been exploited in [23]. In this case, the model is trained on the measurements provided by CGM systems, and used to predict a singular value after a pre-defined prediction horizon. The output is modeled as a univariate Gaussian distribution, so as to be able to follow the uncertainty of the prediction. The LSTM dimension is set to 128 and it is trained on the Ohio T1DM Dataset [54], considering the first 80% of the glucose level as training data for each patient, and validating on the last 20%. A more complex architecture was designed by Sun et. al. [24]. In particular, they propose to use a sequential model with one LSTM layer, one bidirectional LSTM layer and several fully connected layers to predict BG levels for different prediction horizons. The model is trained on the CGM measurements of both *in silico* and *in vivo* data coming from 20 real patients.

Convolutional RNNs have also been exploited to predict the BG level [25]. The concatenated time series of glucose level, carbohydrate and insulin is firstly preprocessed by a deep convolutional networks, so that the recurrent LSTM layers accepts these features instead of the CGM measurements directly. The model is trained on *in silico* data consisting on a small sample of 10 adult T1D subjects simulated using the UVA/Padova simulator.

3. Method

Glucose concentration depends mainly on the injected insulin and carbohydrate intake, which have opposite effects on glucose levels. Indeed, it is well-known from physiology that an increase in insulin results in a decrease in glucose concentration, while a meal intake produces a glucose rise. Of course, the future evolution of the BG is also influenced by its current value and it is consistent with its trend. All these variables can be easily measured without an invasive data collection.

Thus, the inputs of the model proposed in this work is composed by three measurable signals sampled at a given rate T_s: the injected insulin (ins) recorded by subcutaneous insulin pump, the carbohydrate amount (cho) inserted manually by the patient, and the glucose concentration (cgm) measured by the CGM sensor. The output of the model is the interstitial fluid glucose concentration (ig). Specifically, the signal cgm is the interstitial (i.e. subcutaneous) glucose concentration measured by a CGM device and affected by measurement noise, while ig is the real interstitial glucose. These measurements have different ranges of values according to the units adopted by the UVA/Padova simulator: injected insulin doses and carbohydrate amounts are about 100 times larger than the glucose measurements in this dataset. In order to eliminate the units of measurements for data and to guarantee that all features contribute equally in the training process, a data preprocessing step is introduced. In particular, each signal is independently rescaled using the minimum and maximum values and then subdivided in samples of fixed size, depending on the prediction horizon (ph) in analysis. These sub-samples constitute the training and testing data for our model, as detailed in Section 4.

Denoting the current time with t_0 and given $ph \in \mathrm{PH}$, where $\mathrm{PH} = [5, 10, \cdots, 60]$ is the set of the

considered prediction horizons, let's define the following signals:

$$\begin{aligned}
\overleftarrow{\mathbf{cgm}}(t_0, ph) &= \Big[\mathrm{cgm}(t_0 - ph), \mathrm{cgm}(t_0 - ph{+}1), \cdots, \mathrm{cgm}(t_0 - 1)\Big], \\
\overleftarrow{\mathbf{ins}}(t_0, ph) &= \Big[\mathrm{ins}(t_0 - ph), \mathrm{ins}(t_0 - ph{+}1), \cdots, \mathrm{ins}(t_0 - 1)\Big], \\
\overleftarrow{\mathbf{cho}}(t_0, ph) &= \Big[\mathrm{cho}(t_0 - ph), \mathrm{cho}(t_0 - ph{+}1), \cdots, \mathrm{cho}(t_0 - 1)\Big], \\
\overrightarrow{\mathbf{ins}}(t_0, ph) &= \Big[\mathrm{ins}(t_0), \mathrm{ins}(t_0 + 1), \cdots, \mathrm{ins}(t_0 + ph{-}1)\Big], \\
\overrightarrow{\mathbf{cho}}(t_0, ph) &= \Big[\mathrm{cho}(t_0), \mathrm{cho}(t_0 + 1), \cdots, \mathrm{cho}(t_0 + ph{-}1)\Big], \\
\hat{\mathbf{ig}}(t_0, ph) &= \Big[\hat{\mathrm{ig}}(t_0), \hat{\mathrm{ig}}(t_0 + 1), \cdots, \hat{\mathrm{ig}}(t_0 + ph{-}1)\Big]
\end{aligned} \tag{1}$$

where $\overleftarrow{\mathbf{cgm}}(t_0, ph)$, $\overleftarrow{\mathbf{ins}}(t_0, ph)$, and $\overleftarrow{\mathbf{cho}}(t_0, ph)$ are the CGM data, the delivered insulin and the ingested carbohydrates in the past ph minutes, respectively, while $\overrightarrow{\mathbf{ins}}(t_0, ph)$, $\overrightarrow{\mathbf{cho}}(t_0, ph)$ are the suggested amount of insulin and the meal information in the future ph minutes. For each ph a single model is identified as described in Section 3.1. The aim of each model is to depict the relation between ig in the future ph minutes $\Big(\hat{\mathrm{ig}}(t_0), \hat{\mathrm{ig}}(t_0 + 1), \cdots, \hat{\mathrm{ig}}(t_0 + ph{-}1)\Big)$, collected in the vector $\hat{\mathbf{ig}}$, and the above mentioned signals. In particular, each single $\hat{\mathrm{ig}}$ value can be described as:

$$\hat{\mathrm{ig}}(t_0 + k, ph) = g\left(\overleftarrow{\mathbf{cgm}}(t_0, ph), \overleftarrow{\mathbf{ins}}(t_0, ph), \overleftarrow{\mathbf{cho}}(t_0, ph), \overrightarrow{\mathbf{ins}}(t_0, ph), \overrightarrow{\mathbf{cho}}(t_0, ph)\right) \quad k = 0, 1, 2, \cdots, ph{-}1. \tag{2}$$

In the perspective of employing our solution in an MPC and in order to be able to accurately predict a glucose trend, an ensemble of these models can be trained, independently for each ph, and the predictions from these models can be combined to obtain a trend of future glucose concentration:

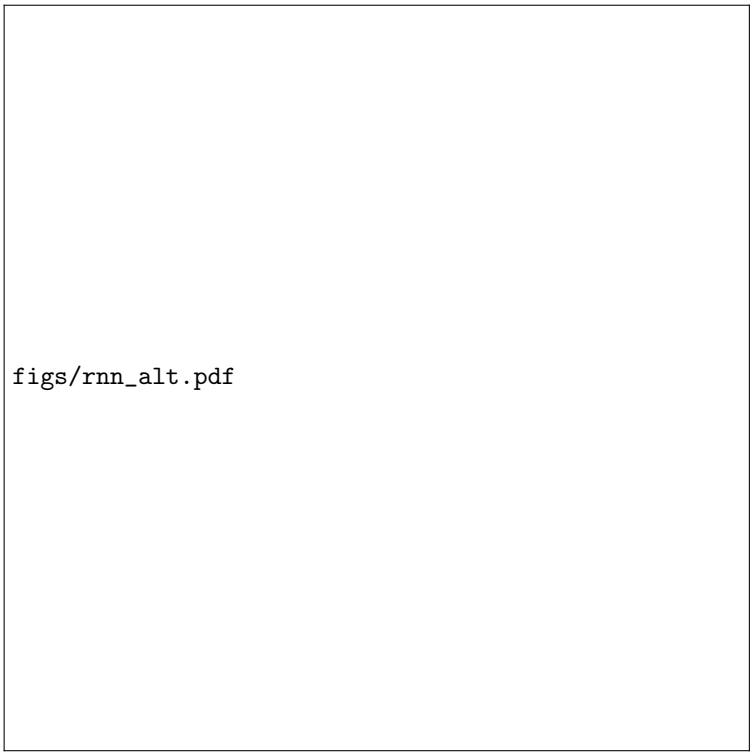
$$\begin{aligned}
\hat{\mathbf{ig}}\,(t_0) = [&\hat{\mathrm{ig}}(t_0, 5), \hat{\mathrm{ig}}(t_0 + 1, 5), \cdots, \hat{\mathrm{ig}}(t_0 + 4, 5), \\
&\hat{\mathrm{ig}}(t_0 + 5, 10), \hat{\mathrm{ig}}(t_0 + 6, 10), \cdots, \hat{\mathrm{ig}}(t_0 + 9, 10), \\
&\vdots \\
&\hat{\mathrm{ig}}(t_0 + 54, 60), \hat{\mathrm{ig}}(t_0 + 55, 60), \cdots, \hat{\mathrm{ig}}(t_0 + 59, 60)].
\end{aligned} \tag{3}$$

3.1. Proposed Architecture

We chose a simple model based on stacked Long Short-Term Memory (LSTM) cells [20, 55]. LSTMs are a special kind of RNNs, which are able to learn how to filter (e.g. forget) part of their hidden state during the inference process in order to model long-term temporal dependencies.

Formally, a single LSTM cell with input $\mathbf{x}(t)$, output $\mathbf{h}(t)$ and an internal cell state $\mathbf{c}(t)$ is described by

(a)



(b)

Figure 1: (a) The basic structure of an LSTM cell. For each arrow pointing to a circle, an addition is performed. Dots represent vector/matrix multiplications. (b) Temporal unfolding and data flow on n stacked LSTM cells.

the following equations, also represented in graphical form in Figure 1(a):

$$\begin{aligned}
\mathbf{c}_{\text{in}}(t) &= \tanh(\mathbf{W}_{xc}\mathbf{x}(t) + \mathbf{W}_{hc}\mathbf{h}(t-1) + \mathbf{b}_c) \\
\mathbf{i}(t) &= \text{sigmoid}(\mathbf{W}_{xi}\mathbf{x}(t) + \mathbf{W}_{hi}\mathbf{h}(t-1) + \mathbf{b}_i) \\
\mathbf{f}(t) &= \text{sigmoid}(\mathbf{W}_{xf}\mathbf{x}(t) + \mathbf{W}_{hf}\mathbf{h}(t-1) + \mathbf{b}_f) \\
\mathbf{o}(t) &= \text{sigmoid}(\mathbf{W}_{xo}\mathbf{x}(t) + \mathbf{W}_{ho}\mathbf{h}(t-1) + \mathbf{b}_o) \\
\mathbf{c}(t) &= \mathbf{f}(t)\mathbf{c}(t-1) + \mathbf{i}(t)\mathbf{c}_{\text{in}}(t) \\
\mathbf{h}(t) &= \mathbf{o}(t)\tanh\mathbf{c}(t)
\end{aligned} \tag{4}$$

where each weight matrix $\mathbf{W}_x$, $\mathbf{W}_h \in \mathbb{R}^{d\times d}$ and $\mathbf{b}$, $\mathbf{x}(t)$, $\mathbf{h}(t)$, $\mathbf{c}_{in}(t)$, $\mathbf{i}(t)$, $\mathbf{f}(t)$, $\mathbf{o}(t)$, $\mathbf{c}(t) \in \mathbb{R}^d$ while d represent the LSTM *dimension*, an hyperparameter defined upfront by design and constant among all cells. Respectively, $\mathbf{i}(t), \mathbf{f}(t), \mathbf{o}(t)$ are called the *input*, *forget* and *output* gates, while $\mathbf{c}_{\text{in}}(t)$ contains a vector of new candidate values for the cell state.

During temporal unfolding, both $\mathbf{h}(t)$ and $\mathbf{c}(t)$ are passed to the temporal replica of the next cell in the fold. Models made of multiple, stacked LSTM cells can be easily conceived, by making the output of a given

cell the input of the next one in the stack. The process of training through unfolding n-stacked LSTM cells is illustrated in Figure 1(b).

We trained multiple models, one for each $ph \in \text{PH}$. Depending on ph, the whole signal is sampled as described in Eq. (1) and each sub-sample is split into two arrays $\overleftarrow{\mathbf{X}}$ and $\overrightarrow{\mathbf{X}}$, the former representing past information given to the model, the latter representing the suggested therapy and meals for the future:

$$\overleftarrow{\mathbf{X}}(t_0, ph) = \begin{bmatrix} \overleftarrow{\mathbf{cgm}}(t_0, ph) \\ \overleftarrow{\mathbf{ins}}(t_0, ph) \\ \overleftarrow{\mathbf{cho}}(t_0, ph) \end{bmatrix}, \quad \overrightarrow{\mathbf{X}}(t_0, ph) = \begin{bmatrix} \overrightarrow{\mathbf{ins}}(t_0, ph) \\ \overrightarrow{\mathbf{cho}}(t_0, ph) \end{bmatrix} \tag{5}$$

$\overleftarrow{\mathbf{X}}$ and $\overrightarrow{\mathbf{X}}$ are separately processed through two identical branches of the architecture, each being a stack of n LSTM cells. The output of both branches is then concatenated and processed through a final fully connected layer that produces the intended output. Since the main goal of this work is to predict the future BG of a patient subject to different insulin therapy in order to define the optimal treatment, the second branch containing the suggested future therapy cannot be excluded. As the model aims to forecast the IG signals, the supervised architecture assumes to have access to the IG signal during training in order to use them as ground truth.

More formally, leaving out the flowing of the internal cell states, the model is described by:

$$\begin{array}{rclrcl} \overleftarrow{\mathbf{h}}_1(t_0, ph) & = & \text{LSTM}_1(\overleftarrow{\mathbf{X}}(t_0, ph)) & \overrightarrow{\mathbf{h}}_1(t_0, ph) & = & \text{LSTM}_1(\overrightarrow{\mathbf{X}}(t_0, ph)) \\ \overleftarrow{\mathbf{h}}_2(t_0, ph) & = & \text{LSTM}_2(\overleftarrow{\mathbf{h}}_1(t_0, ph)) & \overrightarrow{\mathbf{h}}_2(t_0, ph) & = & \text{LSTM}_2(\overrightarrow{\mathbf{h}}_1(t_0, ph)) \\ & \vdots & & & \vdots & \\ \overleftarrow{\mathbf{h}}_n(t_0, ph) & = & \text{LSTM}_n(\overleftarrow{\mathbf{h}}_{n-1}(t_0, ph)) & \overrightarrow{\mathbf{h}}_n(t_0, ph) & = & \text{LSTM}_n(\overrightarrow{\mathbf{h}}_{n-1}(t_0, ph)) \end{array} \tag{6}$$

$$\hat{\mathbf{ig}}(t_0, ph) = \mathbf{W}_{FC}\ [\overleftarrow{\mathbf{h}}_n(t_0, ph)\ \ \overrightarrow{\mathbf{h}}_n(t_0, ph)] + \mathbf{b}_{FC}$$

where $\mathbf{W}_{FC} \in \mathbb{R}^{d \times ph}$, $\mathbf{b}_{FC} \in \mathbb{R}^{ph}$ and LSTM_n represents the n-th LSTM layer in the stack and it is described by Eq. (4).

The training process uses a Mean Squared Error loss function (MSE) with a default Adam optimizer (learning rate 10^{-3}, batch size of 200, 180 epochs), so that for each sample:

$$\text{MSE} := \frac{1}{ph} \sum_{t=t_0}^{t_0+ph} (\hat{\mathbf{ig}}(t, ph) - \mathbf{ig}(t, ph))^2. \tag{7}$$

The complete solution is shown in Figure 2 and from now on we will refer to it as therapy-driven Deep Glucose Forecasting (DGF).

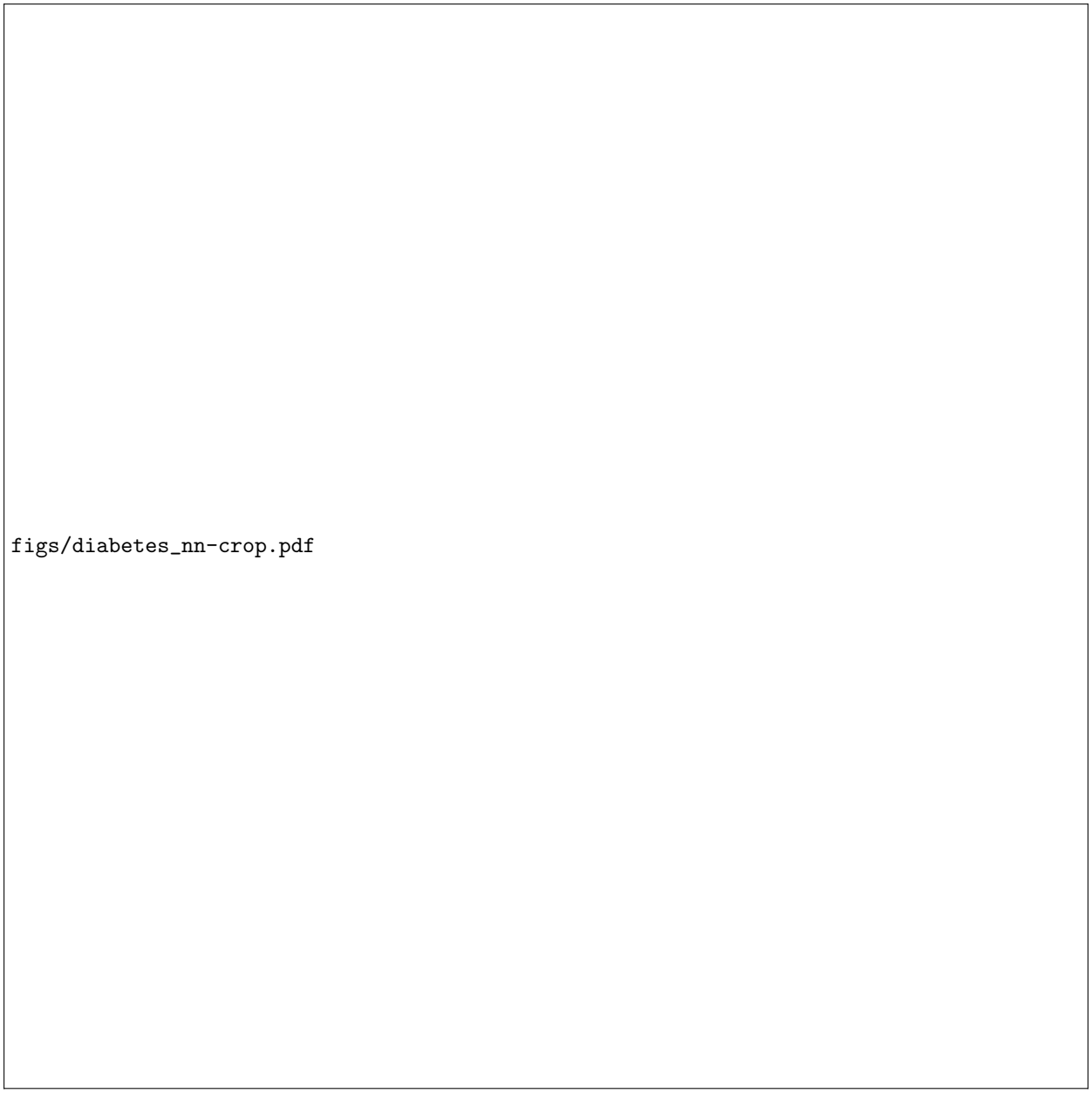


Figure 2: Deep Glucose Forecasting (DGF) architecture. The input is split into two sets: past observations and estimated future inputs. Both branches are processed by n-stacked LSTM cells with dimension d. The output of the branches is concatenated into a final Fully Connected (FC) layer.

4. Experiments

In this section we report on a series of experiments in order to assess the performance of the proposed solution. We first describe the datasets used in our experiments, then we discuss the parameters settings and the measures used to evaluate our approach. In Section 4.3 we analyze how the performance of DGF varies by considering different configurations for the dimension d of each LSTM and the number n of stacked LSTMs, in order to identify the best combinations of these hyperparameters. Finally, we evaluate the prediction capabilities of the proposed solution on *in vivo* data collected during clinical trials [4].

4.1. Dataset

The *in silico* dataset has been generated using the UVA/Padova simulator [38, 39], which is equipped with a cohort of virtual patients and accepted by Food and Drug Administration (FDA) as a substitute to animals trials. This acceptance allowed the *in silico* synthesis of control algorithms directly testable on real patients. The UVA/Padova simulator includes a large nonlinear compartmental model able to simulate the glucose-insulin dynamics of the diabetic population [39]. The inter-subject variability of this population is modelled through different sets of metabolic parameters of this model.

The simulator is equipped with three virtual populations (children, adolescents and adults), each composed of 100 subjects. In the most recent version of the simulator, the circadian variability of insulin sensitivity and meal absorption parameters have been added [56, 57, 58]. The UVA/Padova simulator allows also to simulate the so-called meal announcement, i.e. the patient can announce a meal intake to the controller in advance. It is the unique tool accepted by FDA to test an insulin therapy in order to obtain the approval to clinical studies on real patients [59]. As demonstrated in [60] the virtual subjects of this simulator are representative of the T1D population observed in clinical trials. In this paper, the population of 100 adults of the UVA/Padova simulator is used and two different scenarios are designed. The use of different scenarios is devoted to simulate the realistic intra-subject change in eating habits in terms of timing and meal size variations. Different food habits imply different insulin therapies, which in turn impact differently on glucose levels. Hence, *in- silico* data collected by running two different scenarios allow to a richer and more realistic data set. Table 1 shows Scenario 1, which is a four-day protocol simulated in closed-loop using the MPC-P to define the optimal insulin therapy. The first three days are used for model training, while the remaining day is used for validation purposes. The training scenario involves three meals per day with additional snacks in each day. Moreover, in order to define a real-life scenario, possible errors in the meal announcement are included, i.e. a limited events of unannounced meals or meals announced with a wrong estimation of the amount. Scenario 2 lasts three days and it is reported in Table 2. The meal amounts and times of this protocol are designed to reproduce a real life scenario. Hypo treatments of 15 g are administrated to the patient in case glucose concentration falls below 65 mg/dl in both scenarios.

	Time	CHO (g)	Insulin Bolus
Day 1	08:00	60	Bolus on time
	13:00	60	Bolus at 14:00
	17:00	30	No bolus
	20:00	80	Bolus on time
Day 2	08:00	50	Bolus on time
	10:00	15	No bolus
	13:00	35	Bolus on time
	19:00	80	Bolus on time
	22:00	20	Bolus on time
Day 3	06:30	40	Bolus on time
	09:30	20	No bolus
	12:30	45	Bolus at 12:00 for 50 g
	17:00	20	Bolus on time
	20:00	70	Bolus at 20:30
	23:00	20	No bolus
Day 4	08:00	35	Bolus on time
	11:30	20	Bolus on time
	13:30	60	Bolus at 13:30 for 30 g
	16:30	20	No bolus
	20:30	90	Bolus on time

Table 1: Training Scenario.

	Time	CHO (g)	Insulin Bolus
Day 1	08:00	50	Bolus on time
	13:00	50	Bolus on time
	19:00	70	Bolus on time
	23:00	20	Bolus on time
Day 2	06:30	50	Bolus on time
	09:30	15	No bolus
	13:00	60	Bolus at 12:00 for 50 g
	17:00	25	Bolus on time
	20:00	90	Bolus at 20:15 for 70 g
	23:00	15	No bolus
Day 3	08:30	50	Bolus on time
	11:30	20	Bolus on time
	14:00	60	Bolus at 13:00 for 30 g
	17:00	20	No bolus
	20:30	100	Bolus on time

Table 2: Testing Scenario.

Scenario 1 is used to perform model training, while Scenario 2 is exploited to assess the prediction capabilities of the proposed model. Specifically, Scenario 2, i.e. the testing scenario, is defined to reproduce eating patterns different from those present in Scenario 1, i.e. the training scenario. Moreover, an *in vivo* dataset is considered and it is composed of clinical data of a single T1D patient of the Padova clinical centre collected during experiments involved in the AP@Home project [61]. The considered clinical trial lasted for a month, and it has been conducted through a fully automatic closed-loop control [4]. The closed-loop system was composed of an suitably modified android smartphone (the DiAs platform [38]), communicating wirelessly with the G4 Platinum CGM system, Dexcom Inc. and the AccuCheck Spirit Combo insulin pump, Roche Diagnostic. This dataset is challenging because the clinical trial has been conducted in free-living conditions, i.e. the patient could have a normal life without any type of restriction. This particular dataset was used to identify the classic mathematical model described in [17] which represents the state-of-art reached so far via the classic identification techniques developed by the authors. Testing on a dataset not belonging to the training set and acquired following a real-life scenario allows the evaluation of the robustness of this approach to new unseen data and subjects.

4.2. Parameter Settings and Evaluation Protocol

The accuracy of the model predictions is assessed by considering various prediction horizons, which are expressed in terms of minutes. In this paper, ph from 5 minutes to 60 minutes are considered. Since the proposed model is aimed to be included in the MPC-P controller, that is characterized by a sample time of $T_s = 5$ minutes for the predictions, we consider PH $=[5, 10, \cdots, 60]$ and the predicted signals are sampled every T_s. For a given patient p and a specific $ph \in$ PH, we denote with $\widehat{\mathbf{ig}}(\cdot, ph)$ the ph-steps ahead prediction on the entire testing scenario, $\mathbf{ig}(\cdot)$ the considered reference in Scenario 2, and $\overline{\text{ig}}$ its average value. The predictions of the model are evaluated in terms of Coefficient Of Determination (COD), the index of fitting called FIT, and Root Mean Square Error (RMSE). These metrics are the standards used to evaluate performance in system identification [62, 63]. The RMSE is an absolute quantity that assesses the variance of the prediction error (the larger it is, the poorer is the prediction) while FIT and COD are two normalized metrics commonly used in system identification. Moreover, this choice allows a fair comparison of the proposed model with respect to previously published solutions [17].

These metrics are defined as follows:

$$\begin{aligned} COD_p(ph) &= 100 * \left(1 - \frac{||\widehat{\mathbf{ig}}(j, ph) - \mathbf{ig}(j)||^2}{||\,\mathbf{ig}(j) - \overline{\text{ig}}||^2}\right) \\ FIT_p(ph) &= 100 * \left(1 - \frac{||\widehat{\mathbf{ig}}(j, ph) - \mathbf{ig}(j)||}{||\,\mathbf{ig}(j) - \overline{\text{ig}}||}\right) \\ RMSE_p(ph) &= \frac{1}{N_{\text{sample}}} ||\,\mathbf{ig}(j) - \widehat{\mathbf{ig}}(j, ph)|| \end{aligned} \tag{8}$$

where $j = ph, ph + T_s, \cdots, N_{\text{sample}} \cdot T_s$ is used to index the considered samples, and N_{sample} is the number of samples of the signal when sampled every T_s minutes for Scenario 2. COD and FIT are equal to 100% for perfect predictions and can reach negative values in case of bad performances. The average value of each metric ($\overline{COD}$,$\overline{FIT}$,$\overline{RMSE}$) for all PH is used as main outcome to evaluate the overall performance as follows:

$$\begin{aligned} \overline{COD} &= \frac{1}{N_{ph}} \sum_{i=1}^{N_{ph}} \left(\frac{1}{N_p} \sum_{p=1}^{N_p} COD_p(ph(i)) \right) \\ \overline{FIT} &= \frac{1}{N_{ph}} \sum_{i=1}^{N_{ph}} \left(\frac{1}{N_p} \sum_{p=1}^{N_p} FIT_p(ph(i)) \right) \\ \overline{RMSE} &= \frac{1}{N_{ph}} \sum_{i=1}^{N_{ph}} \left(\frac{1}{N_p} \sum_{p=1}^{N_p} RMSE_p(ph(i)) \right) \end{aligned} \tag{9}$$

where N_{ph} is the total number of $ph \in PH$ (i.e. $N_{ph} = 12$) and N_p is the total number of patients involved in each testing experiment.

4.3. Ablation Study

In order to have a comprehensive idea of the model behaviour, an ablation study is performed to assess the model's structure and the function of its different components. Specifically, the goal of this study is to evaluate the contribution of the model components and to observe how these affected the predictive capabilities.

In order to maximize the generalization capability of the proposed algorithm, we train the model on different patients with eating habits drastically different with respect to those observed in the testing scenario. To do so, the population of 100 adults is split in two parts ($N_p = 50$): the model is trained on the first N_p patients in Scenario 1, and the tests are conducted on the second half of the patients in Scenario 2. The same is performed but considering the other half of patients in each scenario. The final results are obtained by averaging the estimates from the two different train and test groups. This data separation has been chosen in order to test the capability of the model to represent subjects not belonging to the training set but also to check the model robustness against a variation in meal sizes and correlation of meal sizes between a day. The same experiments have also been performed by testing the two trained models described above on the real patient (i.e. *in vivo*) and taking the average of the two results. This experiment has been performed in order to assess the generalization capability of the model in a real-world scenario.

Firstly, the study focuses on the choice of both the size of the hidden units in each layer ($d \in \{16, 32, 64\}$) and the number of LSTM layers ($n \in \{1, 2, 3\}$). The choice of powers of 2 in the number of hidden units follows a standardized practice. We have also studied the performance of the scenarios characterized by either $n >= 3$ and $d >= 128$. In both scenarios the performance dropped sharply and the training time increase significantly, so the results are not reported here. From Table 3 it is possible to observe that increasing the number of hidden units for each LSTM entails a slight improvement of the performance indices while the number of stacked LSTM does not significantly affect the final performance. Generally speaking models with more parameters are able to improve prediction performance only up to a point, that is when the amount and variability of available data is sufficient to train the model. The result presented in this section suggest that given the available data, the best configuration is the one with a single LSTM with $d = 64$. However, the performance of the single-LSTM implementation drops sharply (by more than 5%) on the real patient. As the only significant difference is in the number of parameters, we can deduce that this behavior is due to over fitting on the training set. This behavior on the real patients was not observed on models with multiple LSTMs. For these reasons all subsequent experiments have been performed with the configuration $n = 2, d = 64$.

Secondly, we analyze how the different features considered as input for the network influence the final performance. For this reason, we removed past insulin ($\overleftarrow{\mathbf{ins}}$) and carbohydrates ($\overleftarrow{\mathbf{cho}}$) from the input stream. Denoting with $\overleftarrow{\mathbf{X}}^*$ the modified input array, the vector representing past information given to the model is

		$\overline{COD}$ (%)			$\overline{FIT}$ (%)			$\overline{RMSE}$		
	d \ n	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
SILICO	**16**	79.48	79.27	79.12	55.93	55.59	55.40	12.57	12.65	12.72
	32	81.31	81.14	79.60	57.94	57.12	56.27	11.93	11.79	12.41
	64	82.10	81.93	81.32	58.84	58.66	57.98	11.68	11.72	11.93
VIVO	**16**	67.67	71.92	71.61	45.03	48.17	47.92	29.09	27.43	27.56
	32	67.20	72.57	72.05	44.96	48.93	48.65	29.13	27.03	27.18
	64	62.06	71.70	72.50	41.41	48.44	49.17	31.01	27.29	26.91

Table 3: Performance metrics on each combination of $d - n$.

defined as follows:

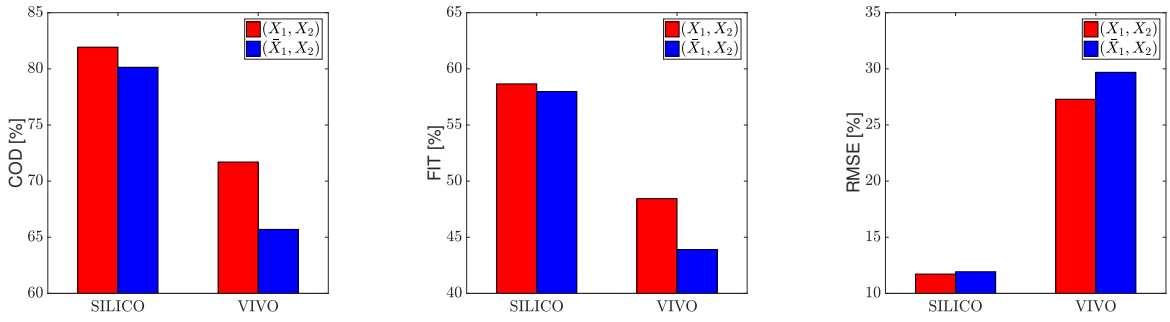
$$\overleftarrow{\mathbf{X}}^*(t_0, ph) = \left[\overleftarrow{\mathbf{cgm}}(t_0, ph)\right]. \tag{10}$$

Figure 3 shows the comparison of the prediction performance of the models with $\left[\overleftarrow{\mathbf{X}}^*(t_0, ph)\ \ \overrightarrow{\mathbf{X}}(t_0, ph)\right]$ and $\left[\overleftarrow{\mathbf{X}}(t_0, ph)\ \ \overrightarrow{\mathbf{X}}(t_0, ph)\right]$ as inputs, respectively. It denotes that the introduction of the information regarding the past insulin therapy and ingested meals guarantees an improvement in the prediction performance of the model. In particular, considering the results obtained by using *in silico* data, there is a slight improvement in performance. Indeed, the virtual patients belonging to the training and testing groups are subsets of the same population.

Since the LSTM is able to learn the behaviour of the population, the additional information about past history does not provide a significant improvement. On the other hand, Figure 3 shows a significant gap in performance computed on real life data. Indeed, if the LSTM is trained on *in silico* data, the lack of past therapy information lowers the performance on *in vivo* testing dataset. Hence, the past evidence $\overleftarrow{\mathbf{ins}}$ and $\overleftarrow{\mathbf{cho}}$ help mitigating the differences in the data distribution. The model obtained considering these additional information is able to generalize to new unseen data and improve the overall glucose control.

5. Discussion

The proposed solution is a population average model identified on the 100 adults of the UVA/Padova simulator. An average model could ideally limit the performance since it describes the average dynamics of the population, so we aim to test both its prediction and generalization capabilities on a dataset different from the one used in training. Firstly, we have re-trained the model considering the entire adult virtual population as a training dataset in order to maximize the available information provided to the training procedure. Then, the proposed algorithm is tested on a 1-month dataset containing all the data collected



(a) (b) (c)

Figure 3: Comparison of the prediction performance on both *in silico* and *in vivo* testing data in term of COD (a), FIT (b) and RMSE (c) on Scenario 2 and real-life data.

Predictor	**COD (%)**	**FIT (%)**	**RMSE**
AVG	15.23	11.48	46.82
DGF	69.85	48.44	27.29
DMP [17]	80.79	61.44	–
DGF Fine Tuned	73.23	49.69	26.63
DGF Fine-tuned + Exp. Filtering	84.05	60.14	21.09

Table 4: Prediction matrices.

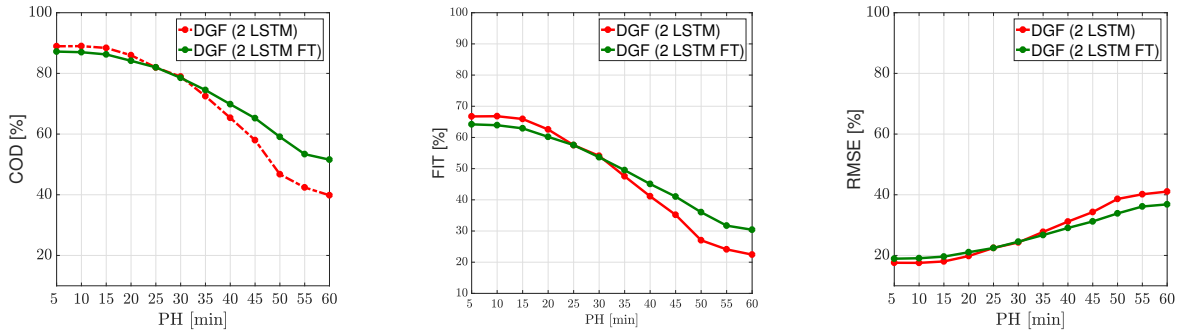
during the clinical trial [4] for a single patient. This dataset is challenging because it includes eating patterns not present in the training dataset, but also it includes all the problems experienced during the clinical trial.

The performance obtained by the DFG model and the linearized average model (AVG) of the UVA/Padova simulator are reported in Table 4. By considering the first two rows of Table 4, the DFG model shows superior prediction performance against the AVG. Moreover, DGF approach proves to be able to generalize over different datasets by achieving interesting improvements with respect to AVG, despite being an average model.

5.1. Fine tuning

The drawback of an average model is that it cannot fully describe the variety of individual glucose response of the entire population. The definition of an individualized insulin therapy by exploiting a patient-tailored model can substantially improve the effectiveness of the glucose control. Hence, in order to improve the DGF model performance, the LSTM trained on *in silico* data is fine-tuned on data of the specific patient.

In this context, Fine-Tuning (FT) slightly modifies the weights of the LSTM in order better fit the individual behaviour of the real patient. For the purpose of this paper, we have decided to extend the



(a) (b) (c)

Figure 4: A comparison of the prediction performance of 2-layer stacked LSM compared to its fine-tuned version in term of COD (a), FIT (b) and RMSE (c) on the clinical dataset.

partial retraining entailed by any FT to the entire architecture, by using a suitably small learning rate (10^{-5}) for 10 additional epochs on the FT dataset. A filtered version of the collected CGM data (ig_R) is used as reference for the signal *ig*. The signal ig_R is obtained by using a retrofitting algorithm, which is able to reconstruct an accurate continuous-time BG profile by exploiting BG samples from the fingerstick and CGM data from the sensor [64]. In order to provide a good amount of information to describe the dynamic of a specific patient, the FT dataset contains two days picked up randomly among the available ones.

The FT technique improves the accuracy by 6% on the 2-layer stacked LSTM, as reported in the row 4 of Table 4. Figure 4 reports the performance metrics as a function of *ph* for the 2-layer stacked LSTM with *in vivo* testing dataset. The higher the *ph*, the lower the performance. This is motivated by the fact that the further you want to predict the more difficult it becomes. The improvements of the fine-tuned models are evident for large *ph* values where the performance increases with respect to the original model without FT. These improvements are less obvious if we compare their performance against the performance of the individualized model (Daily Model Predictor, DMP) presented in [17] and reported in Table 4. In [17] the individualized model is identified from real-data on the base of the a-priori knowledge acquired through the analysis of the patient real-data, while here real-data are used to adjust the model pre-trained on *in silico* data.

5.2. Output filtering

The main limitation of the proposed approach is that the network is trained on noisy input measurements but a noiseless signal is required as output. Since the measurement noise that affects the CGM data is not negligible, the network would try to reduce the noise on the output in order to improve the overall quality of the prediction. However, this effect can be limited and in order to further improve the prediction smoothness an exponential filtering is applied to the predictions. The exponential filter decreases gradually the weights

on the past observations and, considering the decay of the effects of meals and insulin on the glucose, it represents the natural choice for this kind of applications. The exponential filter also allows to forget erroneous values predicted in the previous steps.

The predicted $\hat{\text{ig}}(\cdot, ph)$ is filtered by the following exponential filter:

$$\hat{\text{ig}}_{\text{exp}}(t_0 + k, ph) = \alpha \cdot \hat{\text{ig}}(t_0 + k, ph) + (1 - \alpha) \cdot \hat{\text{ig}}_{\text{exp}}(t_0 + k - 1, ph) \qquad k = 0, 1, \cdots, ph-1 \tag{11}$$

where $\alpha \in (0, 1)$ is the smoothing factor and is defined as follows:

$$\alpha = \frac{2}{w_{\text{exp}} + 1} \tag{12}$$

and w_{exp} is the length of the window used by the filter. It is set to 5, i.e. the minimum window observed by the model and it is kept fixed for all models. Figure 5 shows the noisy CGM data, the output of the retrofitting algorithm ig_R, which represents the ground truth, and the output of the exponential filter $\widehat{\text{ig}}_{\text{exp}}$. It may be noted in Figure 5 that there is a scaling problem in the signal $\widehat{\text{ig}}_{\text{exp}}$. This effect can be explained by considering that $\widehat{\text{ig}}_{\text{exp}}$ is rescaled with the minimum and maximum training values of **ig** while these two values for the test data cannot be know *a priori*. However, it is important to highlight that the glucose metabolism is highly affected by the food and lifestyle habits of the specific patient. This implies that the range of the injected insulin, carbohydrate amounts and glucose levels are individual characteristics of the patient. In order to cope with this problem, a larger dataset of the patient is required to capture the individual variability without compromising the testing dataset. The last row of Table 4 shows the results obtained by applying the exponential filter to the output of the 2-layer stacked LSTM model. It is evident that the filtering technique is able to highly improve the performance by partially removing the noise that affects the CGM data.

6. Conclusions

A novel solution which follows a therapy-driven approach based on deep learning has been introduced in order to predict a trend of future glucose concentration in T1D patients. The solution entails multiple models trained on the *in silico* adult patients of the UVA/Padova simulator. Each model is used to predict a glucose profile for a fixed prediction horizon and the individual predictions are then aggregated to obtain a profile of future glucose levels.

In order to assess the generalization ability, the models have been tested on real data collected during a 1-month clinical trial in free-living conditions of a single patient. The achieved results show that the proposed approach can significantly improve predictive performance despite being an average model. In order to individualize the trained models, fine-tuning is applied to each model separately considering a small portion of the data pertaining to a specific patient. Satisfactory results have been obtained in terms of prediction

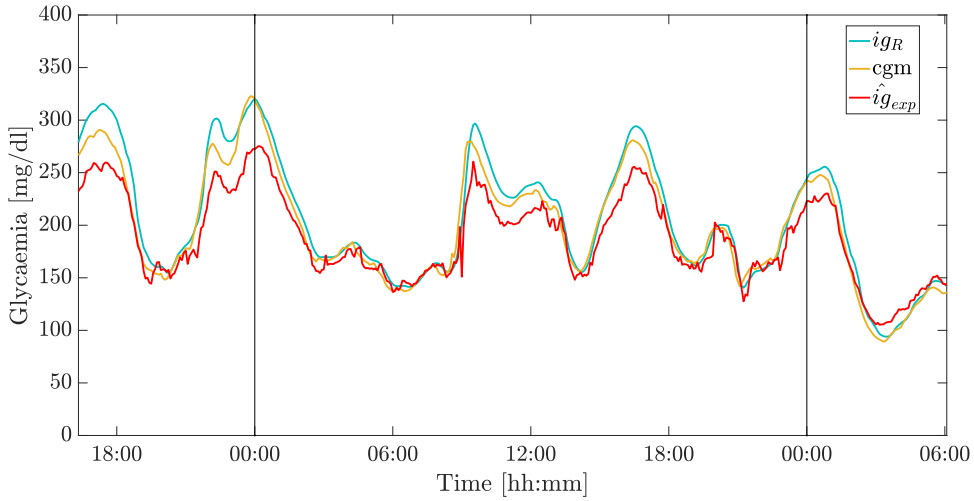


Figure 5: Glucose profile of a 24h study-case: the yellow line is the noisy CGM data, the light blue line is the output of the retrofitting algorithm ig_R presented as reference, and the red line is the output of the exponential filter $\widehat{\mathrm{ig}}_{\mathrm{exp}}$.

capabilities. The use of a single real patient for the final validation of the model is the main limitation of this study. However, the strength of the results obtained in this work is enforced by the large changes of the patient habits and the time-varying nature of the system under study. We plan to investigate more in details how the data affect the model training and to improve the individualized models by exploiting a larger amount of patient data. However, in order to validate the model on a significantly large number of real patients, new clinical trials are required.

Acknowledgements

This work was supported by the Italian project PRIN15: "Forget Diabetes: Adaptive Physiological Artificial Pancreas" and by the Italian project "Fondi per il Finanziamento delle Attività Base di Ricerca". The authors would like to thank Prof. Cobelli for making available the entire population of the UVA/Padova simulator. The authors would also like to thank Roberto Gallotta for his preliminary contribution. The Titan Xp GPU used for this research was donated by the NVIDIA Corporation.

References

[1] C. Cobelli, E. Renard, B. P. Kovatchev, Artificial pancreas: Past and present and future, Diabetes 60 (11) (2011) 2672–2682.

[2] F. Cameron, B. W. Bequette, D. M. Wilson, B. A. Buckingham, H. Lee, G. Niemeyer, A closed-loop artificial pancreas based on risk management, Journal of diabetes science and technology 5 (2) (2011) 368–379.

[3] H. Thabit, R. Hovorka, Coming of age: the artificial pancreas for type 1 diabetes, Diabetologia 59 (9) (2016) 1795–1805.

[4] E. Renard, A. Farret, J. Kropff, D. Bruttomesso, M. Messori, J. Place, R. Visentin, R. Calore, C. Toffanin, F. Di Palma, et al., Day-and-night closed-loop glucose control in patients with type 1 diabetes under free-living conditions: results of a single-arm 1-month experience compared with a previously reported feasibility study of evening and night at home, Diabetes Care 39 (7) (2016) 1151–1160.

[5] H. Thabit, M. Tauschmann, J. M. Allen, L. Leelarathna, S. Hartnell, M. E. Wilinska, C. L. Acerini, S. Dellweg, C. Benesch, L. Heinemann, et al., Home use of an artificial beta cell in type 1 diabetes, New England Journal of Medicine 373 (22) (2015) 2129–2140.

[6] J. Kropff, S. Del Favero, J. Place, C. Toffanin, R. Visentin, M. Monaro, M. Messori, F. Di Palma, G. Lanzola, A. Farret, et al., 2 month evening and night closed-loop glucose control in patients with type 1 diabetes under free-living conditions: a randomised crossover trial, The lancet Diabetes & endocrinology 3 (12) (2015) 939–947.

[7] S. M. Anderson, D. Raghinaru, J. E. Pinsker, F. Boscari, E. Renard, B. A. Buckingham, R. Nimri, F. J. Doyle, S. A. Brown, P. Keith-Hynes, et al., Multinational home use of closed-loop control is safe and effective, Diabetes Care 39 (7) (2016) 1143–1150.

[8] R. M. Bergenstal, S. Garg, S. A. Weinzimer, B. A. Buckingham, B. W. Bode, W. V. Tamborlane, F. R. Kaufman, Safety of a hybrid closed-loop insulin delivery system in patients with type 1 diabetes, Jama 316 (13) (2016) 1407–1408.

[9] E. Dassau, H. Zisser, R. A. Harvey, M. W. Percival, B. Grosman, W. Bevier, E. Atlas, S. Miller, R. Nimri, L. Jovanovič, F. J. Doyle III, Clinical evaluation of a personalized artificial pancreas, Diabetes care (2012) DC_120948.

[10] J. E. Pinsker, A. J. Laguna Sanz, J. B. Lee, M. M. Church, C. Andre, L. E. Lindsey, F. J. Doyle III, E. Dassau, Evaluation of an artificial pancreas with enhanced model predictive control and a glucose prediction trust index with unannounced exercise, Diabetes technology & therapeutics 20 (7) (2018) 455–464.

[11] K. Zarkogianni, E. Litsa, K. Mitsis, P.-Y. Wu, C. D. Kaddi, C.-W. Cheng, M. D. Wang, K. S. Nikita, A review of emerging technologies for the management of diabetes mellitus, IEEE Transactions on Biomedical Engineering 62 (12) (2015) 2735–2749.

[12] A. Baghban, A. Jalali, M. Shafiee, M. H. Ahmadi, K.-w. Chau, Developing an anfis-based swarm concept model for estimating the relative viscosity of nanofluids, Engineering Applications of Computational Fluid Mechanics 13 (1) (2019) 26–39.

[13] S. Samadianfard, A. Majnooni-Heris, S. N. Qasem, O. Kisi, S. Shamshirband, K.-w. Chau, Daily global solar radiation modeling using data-driven techniques and empirical equations in a semi-arid climate, Engineering Applications of Computational Fluid Mechanics 13 (1) (2019) 142–157.

[14] C. Wu, K. Chau, Rainfall–runoff modeling using artificial neural network coupled with singular spectrum analysis, Journal of Hydrology 399 (3-4) (2011) 394–409.

[15] R. Moazenzadeh, B. Mohammadi, S. Shamshirband, K.-w. Chau, Coupling a firefly algorithm with support vector regression to predict evaporation in northern iran, Engineering Applications of Computational Fluid Mechanics 12 (1) (2018) 584–597.

[16] C. Toffanin, S. Del Favero, E. Aiello, M. Messori, C. Cobelli, L. Magni, Glucose-insulin model identified in free-living conditions for hypoglycaemia prevention, Journal of Process Control 64 (2018) 27–36.

[17] C. Toffanin, E. Aiello, S. Del Favero, C. Cobelli, L. Magni, Multiple models for artificial pancreas predictions identified from free-living condition data: A proof of concept study, Journal of Process Control 77 (2019) 29–37.

[18] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: F. Pereira, C. J. C. Burges, L. Bottou, K. Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 25, 2012, pp. 1097–1105.

[19] O. Ronneberger, P.Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI), Vol. 9351 of LNCS, 2015, pp. 234–241.

[20] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.

[21] F. Allam, Z. Nossai, H. Gomma, I. Ibrahim, M. Abdelsalam, A recurrent neural network approach for predicting glucose concentration in type-1 diabetic patients, in: Engineering Applications of Neural Networks, 2011, pp. 254–259.

[22] C. Meijner, S. Persson, Blood glucose prediction for type 1 diabetes using machine learning (2017).

[23] J. Martinsson, A. Schliep, B. Eliasson, C. Meijner, S. Persson, O. Mogren, Automatic blood glucose prediction with confidence using recurrent neural networks, in: Proceedings of the 3rd International Workshop on Knowledge Discovery in Healthcare Data co-located with the 27th International Joint Conference on Artificial Intelligence and the 23rd European Conference on Artificial Intelligence (IJCAI-ECAI 2018), Stockholm, Schweden, July 13, 2018., 2018, pp. 64–68.

[24] Q. Sun, M. V. Jankovic, L. Bally, S. G. Mougiakakou, Predicting blood glucose with an LSTM and bi-lstm based deep neural network, CoRR abs/1809.03817.

[25] K. Li, J. Daniels, P. H. Viñas, C. Liu, P. Georgiou, Convolutional recurrent neural networks for blood glucose prediction, CoRR abs/1807.03043.

[26] M. Messori, C. Toffanin, S. Del Favero, G. De Nicolao, C. Cobelli, L. Magni, ndividualization for artificial pancreas, Computer methods and programs in biomedicine 171 (2019) 133–140.

[27] S. Del Favero, G. Pillonetto, C. Cobelli, G. De Nicolao, A novel nonparametric approach for the identification of the glucose-insulin system in type 1 diabetic patients, 20th IFAC World Congress (2011) 8340–8346.

[28] M. Percival, Y. Wang, B. Grosman, E. Dassau, H. Zisser, L. Jovanovič, F. J. Doyle, Development of a multi-parametric model predictive control algorithm for insulin delivery in type 1 diabetes mellitus using clinical parameters, Journal of process control 21 (3) (2011) 391–404.

[29] H. Kirchsteiger, S. Pölzer, R. Johansson, E. Renard, L. del Re, Direct continuous time system identification of miso transfer function models applied to type 1 diabetes, in: Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on, IEEE, 2011, pp. 5176–5181.

[30] A. K. Duun-Henriksen, S. Schmidt, R. M. Røge, J. B. Møller, K. Nørgaard, J. B. Jørgensen, H. Madsen, Model identification using stochastic differential equation grey-box models in diabetes, Journal of diabetes science and technology 7 (2) (2013) 431–440.

[31] A. J. Laguna, P. Rossetti, F. J. Ampudia-Blasco, J. Vehí, J. Bondia, Identification of intra-patient variability in the postprandial response of patients with type 1 diabetes, Biomedical Signal Processing and Control 12 (2014) 39–46.

[32] K. Turksoy, L. Quinn, E. Littlejohn, A. Cinar, Multivariable adaptive identification and control for artificial pancreas systems, IEEE Transactions on Biomedical Engineering 61 (3) (2014) 883–891.

[33] A. J. Laguna, P. Rossetti, F. J. Ampudia-Blasco, J. Vehí, J. Bondia, Experimental blood glucose interval identification of patients with type 1 diabetes, Journal of Process Control 24 (1) (2014) 171–181.

[34] A. Bock, G. François, D. Gillet, A therapy parameter-based model for predicting blood glucose concentrations in patients with type 1 diabetes, Computer methods and programs in biomedicine 118 (2) (2015) 107–123.

[35] A. Bhattacharjee, A. Sutradhar, Data driven nonparametric identification and model based control of glucose-insulin process in type 1 diabetics, Journal of Process Control 41 (2016) 14–25.

[36] C. Toffanin, S. Del Favero, E. Aiello, M. Messori, C. Cobelli, L. Magni, MPC model individualization in free-living conditions: A proof-of-concept case study, IFAC-PapersOnLine 50 (1) (2017) 1181–1186.

[37] S. Oviedo, J. Vehí, R. Calm, J. Armengol, A review of personalized blood glucose prediction strategies for T1DM patients, International journal for numerical methods in biomedical engineering 33 (6) (2017) e2833.

[38] B. P. Kovatchev, M. D. Breton, C. D. Man, C. Cobelli, In silico preclinical trials: A proof of concept in closed-loop control of type 1 diabetes, Journal of Diabetes Science and Technology 3 (1) (2009) 44–55.

[39] C. Dalla Man, F. Micheletto, D. Lv, M. Breton, et al., The UVA/PADOVA type 1 diabetes simulator: New features, Journal of Diabetes Science and Technology 8 (1) (2014) 26–34.

[40] M. Messori, J. Kropff, S. Del Favero, J. Place, R. Visentin, R. Calore, C. Toffanin, F. Di Palma, G. Lanzola, A. Farret, et al., Individually adaptive artificial pancreas in subjects with type 1 diabetes: a one-month proof-of-concept trial in free-living conditions, Diabetes Technology & Therapeutics 19 (10) (2017) 560–571.

[41] R. Hovorka, V. Canonico, L. J. Chassin, U. Haueter, M. Massi-Benedetti, M. O. Federici, T. R. Pieber, H. C. Schaller, L. Schaupp, T. Vering, et al., Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes, Physiological measurement 25 (4) (2004) 905.

[42] H. R. Murphy, D. Elleri, J. M. Allen, J. Harris, D. Simmons, G. Rayman, R. Temple, D. B. Dunger, A. Haidar, M. Nodale, et al., Closed-loop insulin delivery during pregnancy complicated by type 1 diabetes, Diabetes care 34 (2) (2011) 406–411.

[43] Y. M. Luijf, J. H. DeVries, K. Zwinderman, L. Leelarathna, M. Nodale, K. Caldwell, K. Kumareswaran, D. Elleri, J. M. Allen, M. E. Wilinska, et al., Day and night closed-loop control in adults with type 1 diabetes: a comparison of two closed-loop algorithms driving continuous subcutaneous insulin infusion versus patient self-management, Diabetes care 36 (12) (2013) 3882–3887.

[44] L. Leelarathna, S. Dellweg, J. K. Mader, J. M. Allen, C. Benesch, W. Doll, M. Ellmerer, S. Hartnell, L. Heinemann, H. Kojzar, et al., Day and night home closed-loop insulin delivery in adults with type 1 diabetes: three-center randomized crossover study, Diabetes Care 37 (7) (2014) 1931–1937.

[45] J. Plank, J. Blaha, J. Cordingley, M. E. Wilinska, L. J. Chassin, C. Morgan, S. Squire, M. Haluzik, J. Kremen, S. Svacina, et al., Multicentric, randomized, controlled trial to evaluate blood glucose control by the model predictive control algorithm versus routine glucose management protocols in intensive care unit patients, Diabetes care 29 (2) (2006) 271–276.

[46] L. Bally, H. Thabit, H. Kojzar, J. K. Mader, J. Qerimi-Hyseni, S. Hartnell, M. Tauschmann, J. M. Allen, M. E. Wilinska, T. R. Pieber, et al., Day-and-night glycaemic control with closed-loop insulin delivery versus conventional insulin pump therapy in free-living adults with well controlled type 1 diabetes: an open-label, randomised, crossover study, The Lancet Diabetes & Endocrinology 5 (4) (2017) 261–270.

[47] R. Gondhalekar, E. Dassau, F. J. Doyle III, Periodic zone-mpc with asymmetric costs for outpatient-ready safety of an artificial pancreas to treat type 1 diabetes, Automatica 71 (2016) 237–246.

[48] K. van Heusden, E. Dassau, H. C. Zisser, D. E. Seborg, F. J. Doyle III, Control-relevant models for glucose control using a priori patient characteristics, IEEE transactions on biomedical engineering 59 (7) (2012) 1839–1849.

[49] L. M. Huyett, T. T. Ly, G. P. Forlenza, S. Reuschel-DiVirgilio, L. H. Messer, R. P. Wadwa, R. Gondhalekar, F. J. Doyle III, J. E. Pinsker, D. M. Maahs, et al., Outpatient closed-loop control with unannounced moderate exercise in adolescents using zone model predictive control, Diabetes technology & therapeutics 19 (6) (2017) 331–339.

[50] G. P. Forlenza, S. Deshpande, T. T. Ly, D. P. Howsmon, F. Cameron, N. Baysal, E. Mauritzen, T. Marcal, L. Towers, B. W. Bequette, et al., Application of zone model predictive control artificial pancreas during extended use of infusion set and sensor: a randomized crossover-controlled home-use trial, Diabetes Care 40 (8) (2017) 1096–1102.

[51] C. Toffanin, M. Messori, F. Di Palma, G. De Nicolao, C. Cobelli, L. Magni, Artificial pancreas: Model predictive control design from clinical experience, Journal of Diabetes Science and Technology 7 (6) (2013) 1470–1483.

[52] S. Del Favero, J. Place, J. Kropff, M. Messori, P. Keith-Hynes, R. Visentin, M. Monaro, S. Galasso, F. Boscari, C. Toffanin, et al., Multicenter outpatient dinner/overnight reduction of hypoglycemia and increased time of glucose in target with a wearable artificial pancreas using modular model predictive control in adults with type 1 diabetes, Diabetes, Obesity and Metabolism 17 (5) (2015) 468–476.

[53] V. Tresp, T. Briegel, J. Moody, Neural network models for the blood glucose metabolism of a diabetic, IEEE Transactions on Neural Networks 10 (1999) 1204–1213.

[54] C. Marling, R. Bunescu, The ohiot1dm dataset for blood glucose level prediction, in: The 3rd International Workshop on Knowledge Discovery in Healthcare Data, Stockholm, Sweden, 2018.

[55] F. A. Gers, et al., Learning to forget: Continual prediction with LSTM, 1999.

[56] R. Visentin, C. Dalla Man, B. Kovatchev, C. Cobelli, The university of virginia/padova type 1 diabetes simulator matches the glucose traces of a clinical trial, Diabetes technology & therapeutics 16 (7) (2014) 428–434.

[57] R. Visentin, C. Dalla Man, Y. C. Kudva, A. Basu, C. Cobelli, Circadian variability of insulin sensitivity: Physiological input for in silico artificial pancreas, Diabetes Technology & Therapeutics 17 (1) (2015) 1–7.

[58] R. Visentin, C. Dalla Man, C. Cobelli, One-day bayesian cloning of type 1 diabetes subjects: toward a single-day uva/padova type 1 diabetes simulator, IEEE Transactions on Biomedical Engineering 63 (11) (2016) 2416–2424.

[59] R. Visentin, M. Schiavon, R. Basu, A. Basu, C. Dalla Man, C. Cobelli, Physiological models for artificial pancreas development, in: The Artificial Pancreas, Elsevier, 2019, pp. 123–152.

[60] R. Visentin, C. Dalla Man, B. Kovatchev, C. Cobelli, The university of virginia/padova type 1 diabetes simulator matches the glucose traces of a clinical trial, Diabetes technology & therapeutics 16 (7) (2014) 428–434.

[61] C. C. Palerm, H. Zisser, L. Jovanovič, F. J. D. III, A run-to-run control strategy to adjust basal insulin infusion rates in type 1 diabetes, Journal of Process Control 18 (3–4) (2008) 258–265.

[62] D. A. Finan, C. C. Palerm, F. J. Doyle III, D. E. Seborg, H. Zisser, W. C. Bevier, L. Jovanovič, Effect of input excitation on the quality of empirical dynamic models for type 1 diabetes, AIChE journal 55 (5) (2009) 1135–1146.

[63] M. Cescon, R. Johansson, Glycemic trend prediction using empirical model identification, in: Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference, IEEE, 2009, pp. 3501–3506.

[64] S. Del Favero, A. Facchinetti, G. Sparacino, C. Cobelli, Improving accuracy and precision of glucose sensor profiles: retrospective fitting by constrained deconvolution, IEEE Transactions on Biomedical Engineering 61 (4) (2014) 1044–1053.