Computers in Biology and Medicine 125 (2020) 103956

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: http://www.elsevier.com/locate/compbiomed

After-meal blood glucose level prediction using an absorption model for neural network training

Rebaz A.H. Karim [a], István Vassányi [a,*], István Kósa [b,c]

[a] Medical Informatics Research & Development Center, University of Pannonia, Veszprém, Hungary
[b] Cardiac Rehabilitation Institute of the Military Hospital, Balatonfüred, Hungary
[c] Department of Preventive Medicine, University of Szeged, Szeged, Hungary

ARTICLE INFO

Keywords:
Blood glucose level prediction
Diabetes mellitus
Meal absorption model
Lifestyle support for diabetes
Outpatient care
Artificial neural network

ABSTRACT

Background: Diabetes Mellitus outpatients would benefit from a lifestyle support tool that delivers reliable short term Blood Glucose Level (BGL) predictions.
Aim: To develop a method for BGL prediction based on the baseline BGL, the insulin dosing and a dietary log.
Methods: A new training method is proposed for a neural network in which an absorption model is applied that uses the nutrient contents of meals. The numerical characteristics of the computed absorption curve are fed to the neural network as training inputs along with the applied insulin doses and BGL evolution measured by a Continuous Glucose Monitoring System. For comparison, another version of the training in which raw carbohydrate values are used as dietary inputs has also been implemented. The method was validated in a clinical trial with 5 patients using a total of 167 meals.
Results: It was found that the proposed method performed significantly better on the 60- and 120-min prediction horizons, with a Root Mean Square Error of 1.12 mmol/l and 1.75 mmol/l, respectively, and more than 96% of the predicted values falling in the 'clinically acceptable' class according to clinical practice. These results surpass those published results to which our method is directly comparable, and also those of the carbohydrate-only version (1.81 mmol/l and 2.53 mmol/l).
Conclusion: The integration of the absorption model in the training process has successfully contributed to the success of the model. Future research will focus on a new trial with more patients to verify these promising results.

1. Introduction

Diabetes mellitus (DM) is a widespread chronic metabolic disorder in which cells of the body are unable to take up sugar from the blood in sufficient volume, resulting in abnormally high blood glucose levels (BGL). The cause of this phenomenon is the absolute or relative lack of insulin. Accordingly, type 1 or type 2 diabetes mellitus (T1DM or T2DM) can be distinguished. The two types are significantly different in etiological (causal) terms. In T1DM, due to autoimmune disease, insulin production is virtually eliminated and must be replaced externally. In the case of T2DM, there is limited insulin production and/or increased insulin resistance, consequently, cells have limited ability to absorb circulating glucose. All T1DM and some T2DM patients use external insulin, most often in the form of subcutaneous injections, typically one injection for each main meal, and all DM patients must take special care of their diet to prevent overly low BGL (hypoglycemia), which can lead to an emergency, as well as high BGL (hyperglycemia), which may cause severe complications if it is sustained for a long time. In practice, this means that the patients on external insulin must estimate their insulin needs such that it matches their daily meals—for which they can rely on some general medical guidelines, frequent fingertip BGL measurements, and their personal experience. The main objective of our work is to assist DM patients, especially those on external insulin, by providing short term (1–4 h) BGL predictions based on their dietary and insulin administration log. If the prediction is reliable and the method is integrated into a mobile lifestyle management application, the patients could be warned of hypo/hyperglycemia in time to reconsider their insulin dosage or planned meal. It should be emphasized that in contrast to artificial pancreas research, the aim now is not to give a recommendation for insulin dosing, but to provide only a prediction—either as an

* Corresponding author. Medical Informatics Research & Development Center, University of Pannonia, H-8200, Veszprém, Egyetem u. 10, Hungary.
E-mail addresses: rebaz.ahkarim@virt.uni-pannon.hu (R.A.H. Karim), vassanyi@almos.vein.hu (I. Vassányi), kosa.istvan@med.u-szeged.hu (I. Kósa).

https://doi.org/10.1016/j.compbiomed.2020.103956
Received 29 June 2020; Received in revised form 31 July 2020; Accepted 31 July 2020
Available online 18 August 2020

educational aid or as a tool to test 'what-if' scenarios.

Due to the significance of the problem, the characteristics of BGL evolution have been researched extensively in healthy and DM persons in the past decades, and BGL was found to be influenced by physical activity [1], stress [2], mental state, and most of all, nutrition. In order to model the effect of nutrition, i.e. the effect of absorbed carbohydrates entering the circulation, on the BGL regulation system, several hundred mathematical models of various complexity have been proposed [3]. Generally, complex models with many parameters can simulate the human metabolism better than simple models with few parameters, but they are increasingly harder to 'personalize' for a real DM outpatient, due to the significant personal (natural) variations in the model parameters.

In an earlier study the authors applied genetic algorithms (GA) and other methods to find the personalized parameter set using a simple but quite powerful state-of-the-art model for BGL regulation. The training input to the model was a detailed nutrition and medication log of a clinical trial, complemented with frequent BGL readings from a Continuous Glucose Monitor (CGM) device. Stress and excessive physical activity were avoided by the patients during the trial. The glucose intake profiles of the consumed meals were computed with an absorption model that could also handle the effect of other nutrients like dietary fiber and the glycemic mix of the logged items. Evolutionary parameter search and diurnal parameter profiles were applied during model training [4,5] and special representation of long-acting insulins [6] to decrease the errors of the model. The results were promising, especially compared to published results of similar trained or untrained models for outpatient care.

In this work, a 'gray-box' approach is proposed by keeping only the simple absorption model and not using any BGL regulation model at all. Instead, the patients' reactions to the computed glycemic load are predicted using an Artificial Neural Network (NN) that is trained by the past meals and corresponding BGL measurements of the patient. Compared to similar work on NN based BGL prediction, the novelty of this approach lies in the use of an absorption model output instead of raw carbohydrate values, for the training. Our hypothesis is that the integration of 'some' a priori domain-specific knowledge in the training process will counter-balance the limited number and variety of meal samples available for NN training, which will in turn improve the prediction.

The differences from our research group's previous work [4,5] can be summarized as follows:

- In this work, data from a new clinical trial is used in which longer CGM records were available
- The previously proposed method used a BGL regulation model for short and long term BGL prediction. The input of the model was the direct output of an absorption model. In contrast, in this paper a method is proposed for short term BGL prediction only that uses a trained NN. The input of the NN is a feature vector that is computed from the output of the same absorption model as was used in our previous work.

The rest of this paper is structured as follows. In the State of the art section, the problem domain and the results of the relevant published studies are briefly reviewed. The Methods section introduces our approach, focusing on the new contribution. Results are described and discussed in the Results and Discussion sections. Finally, conclusions are drawn.

2. State of the art

NN's are computational tools with a structure resembling biological neural networks, often used to learn the behavior of systems that are generally too complex for accurate modeling and identification.

NN's consist of processing units connected by controlled, weighted links. The processing units are similar to biological neurons, which is why they are referred to as 'artificial neurons'. The structure of artificial neurons is shown in Fig. 1. The artificial neuron generates its current output value from its input values using the activation function (a thresholding in the simplest case) applied to the sum of the weighted values of the inputs.

NN's are usually layered, with each layer consisting of artificial neurons. In general, there is a connection between all the neurons in the adjacent layers, while there is no connection between the non-adjacent layers. The input of each neuron is the output value of the neurons of the previous layer. The input of the first layer is provided by the input of the network. The layers between the first (input) and the last (output) layers are referred to as hidden layers. The NN is trained in several iterations by comparing the current output to the desired output (taken from the training set) and back-propagating the error through the layers. The training is accomplished by the neurons adaptively changing the synaptic weights based on the error.

To yield a practically usable prediction model, the NN must be provided with a sufficient amount of training data—in our case meal/insulin log and BGL records. Key parameters of an NN are the number of hidden neural layers, the activation function used in the neurons, the number and interpretation of the inputs and outputs as well as several other algorithmic parameters of the training process [7]. The most common NN structures are as follows.

- Feed-forward NN (FNN): The first and simplest type, the structure of which is described above, used for known inputs and given outputs.
- Recurrent NN (RNN): A more complex structure having an infinite dynamic response due to hidden layers with directed feedback connections. The information passes through a loop so when the neuron provides an output, it can take into account the response to previous inputs.
- Autoregressive NN (ARNN): Such NN's model current values of a series as a function of past values and have a finite dynamic response.

The quality of BGL predictions is usually evaluated in the literature by the mean absolute error (MAE) and, more often, by the root mean square error (RMSE), where the error is the difference between the predicted and the measured BGL values at all time instances for which a BGL measurement is available over a prediction time frame, see Equations (1) and (2).

$$MAE = \frac{\sum_i |(x_i - y_i)|}{n} \tag{1}$$

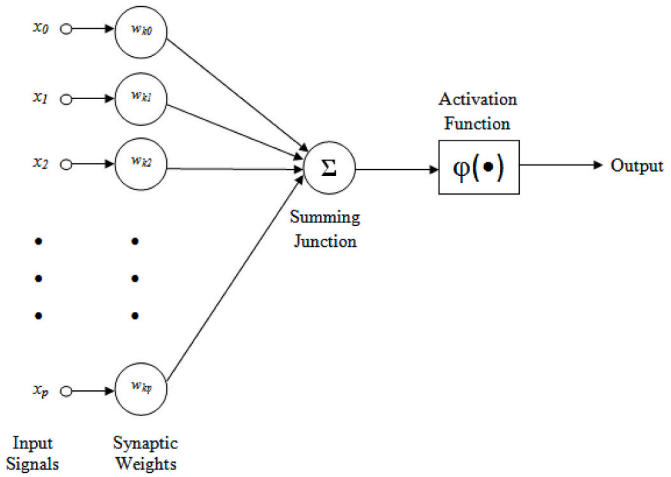

Fig. 1. Structure of the artificial neuron.

$$RMSE = \sqrt{\frac{\sum_i (x_i - y_i)^2}{n}} \quad (2)$$

where x_i is the measured glucose value at the time instant t_i, y_i is the predicted BGL at the same time instant, and n is the total number of blood glucose measurements in each dataset.

A CGM delivers a BGL estimated from the tissue serum glucose concentration every couple of minutes. The prediction time frame, also referred to as *horizon*, typically lasts for 15–240 min and the prediction is often started after a meal and insulin administration event. From the clinical point of view, the practical goal of the prediction is to estimate the patient's glycemia between two main meals of the day, so 15-min predictions have limited applicability (however they can be used to assess model performance). The unit used for RMSE is either mmol/l or mg/dl; in this paper we'll use only mmol/l for consistency. It should be noted that the error range of widely used fingertip BGL meters and calibrated CGM systems is around or above 1 mmol/l.

Besides RMSE, the clinical reliability of BGL predictions is often evaluated with Clarke's Error Grid Analysis (EGA) [8]. EGA classifies predictions into 5 classes A-B-C-D-E with respect to the clinical outcome of an insulin dosing based on the predicted BGL (see Fig. 2). The worst scenario (Class D and E) is an overly high BGL prediction when the actual BGL of the patient is in the <4 mmol/l range, because relying on such a prediction may lead to hypoglycemia, an emergency situation. Thus, the same absolute numerical error may be classified into various classes depending on the real BGL range and the sign of the error. A predicted value is termed 'clinically acceptable' if it is classified into either the A or B EGA class. CG-EGA is a variation of the EGA grid in which the 'accurate' domain is roughly equivalent to the EGA 'clinically acceptable' classification [9].

2.1. An overview of artificial neural networks used for BGL prediction

The studies published in the literature can be characterized by the application area (real outpatients vs. simulated data), the number of patients, the model inputs (CGM or fingertip BGL log, diet, insulin dosing, physical activity, symptoms, etc.), the length of the training data set and the NN structure.

One of the first works to propose NN for BGL prediction was due to Sandham et al., in 1998 who used two T1DM patients' data sets of 10 days each [10]. The training input consisted of insulin, diet (meaning CH quantity), exercise, BGL and an 'X' vector that included parameters such as stress and illness. The FNN had a hidden layer of 95 neurons and the output layer represented the predicted BGL values and used a linear activation function. The input data consisted of 122 events in 20 days, out of which 97 were used for training and the rest for evaluation. As a result, they found that most of the predictions were very close to the measured values (difference of 1.5 mmol/l or less).

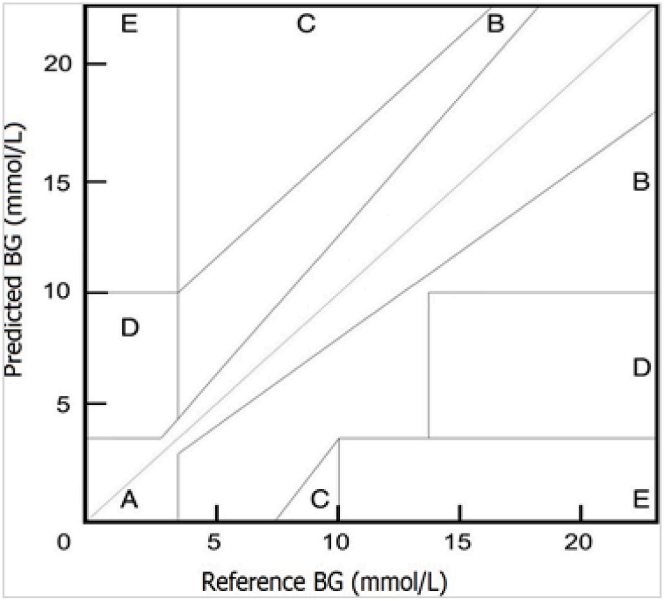

Fig. 2. Error Grid Analysis, adapted from [8].

Later, especially in the last decade, several more results were published in this field, due to the significance of the problem. Here an overview is given of only 10 recent studies that were selected as most relevant to this work. For a more comprehensive review, see [11].

Some authors use no dietary log for the prediction, only the past CGM or fingertip BGL data of the patient [12–15]. It should be noted that such an approach assumes that the patient has a very stable daily schedule with similar or controlled meals every day—an assumption that usually does not hold for real patients. The most recent of these is due to Ali et al., who used fingertip BGL data recorded from 13 T1DM patients. The prediction horizons were 15, 30, 45, and 60 min, and the resulting RMSE values 0.36, 0.4, 0.45l and 0.5 mmol/l, respectively [12]. Two years earlier, the work of Frandes et al. was very similar as they monitored 17 T1DM patients for 4–7 days in free-living conditions, with slightly less accurate results (30-min: 0.1, 60-min: 0.2, and 90-min: 1.2 mmol/l) [13]. Zarkogianni et al. had 6 patients monitored for 7–15 days and trained a special adaptive neuro-fuzzy inference system with wavelet activation functions that integrated both NN and fuzzy logic principles. They validated the model according to EGA. For the 30-min horizon, 94% of the predicted values were in the A class, which fell to 72% for the 60-min horizon [15]. As for the earlier results, Daskalaki et al. compared the performance of a NN model to that of an autoregressive model with or without external insulin input, using a simulator for validation with 30 virtual patients. The NN provided more accurate results compared to other models for the 45-min horizon with an RMSE of only 0.3 mmol/l versus 1.6 mmol/l and 1.4 mmol/l for the autoregressive models [14].

The more typical approach is to use the CH content of the meal consumed and the bolus insulin dose administered before the meal as inputs for the prediction, thus supporting a more realistic application scenario [16–22]. The latest of these results is that of Li et al., who could use a very long training sample of 1–3 months from 10 real and 10 simulated patients. The results for the real patients are impressive (30-min 1.17, 60-min 1.85 mmol/l RMSE) [21]. Mirshekarian et al. had worse results (30-min 1.19, 60-min 2.11 mmol/l RMSE), but their training sample contained only 400 records collected from 10 T1DM patients [22]. Jankovic et al. focused on the effect of physical activity on BGL evolution in their study involving 6 T1DM patients and using a hybrid NN they tried to predict post-exercise BGL based on training with pre-exercise data [19]. Mathiyazhagan & Schechter monitored only 2 patients with CGM but for a longer period (over 8 weeks each). The inputs contained exercise type and duration as well as the time of day. Instead of RMSE, they published MAE error for the 30-min (1.7 mmol/l), 60-min (3.2 mmol/l) and 120-min (5.7 mmol/l) horizons [20]. An earlier result is due to Pappada et al. with the highest reported number of real patients (27) and many kinds of inputs including emotions in a carefully designed and elaborate clinical trial. The performance of the model was validated on 10 patients not included in the model training set, so the objective was slightly different from ours i.e. personalized prediction model research. An RMSE of 2.44 mmol/l was found on the 75-min horizon [16].

The closest approach to ours is perhaps that of Zecchin et al., who first proposed a predictor based 'jump' NN, trained with CH and CGM data form 10 patients, resulting in an RMSE of 0.9 mmol/l on the 30-min horizon [18]. In 2016 the authors tested their jump NN for 20 T1DM patients, with 4 different model versions with respect to the input: (1) CGM only, (2) CGM and insulin dosing, (3) CGM and CH, (4) CGM, insulin and CH. The MAE error was 0.79 mmol/l in scenario (1), 0.8 mmol/l in scenario (2), 0.75 mmol/l in scenario (3) and 0.78 mmol/l in

scenario (4), all for the 30-min horizon, showing that meal intake improved prediction accuracy more than insulin information [17].

All the above trials used T1DM patients as BGL prediction for T2DM is much less researched, though T2DM patients form the majority of the DM population. An exception, shown in the first row of Table 1, is the ongoing work of Kim et al., in which they used an FNN for 16 T2DM patients [23]. The RMSE results were 2.1 mmol/l (30-min) and 2.3 mmol/l (45-min).

As a summary, Table 1 shows a compact reference for the main parameters of the above prediction models and the results achieved.

The literature survey can be concluded by stating that, concerning the nutrition information as training input for a NN, none of the published prediction models used other than raw CH content of the meal to our best knowledge. This is the point that this work tries to improve.

2.2. Some recent results on BGL prediction achieved by other methods

Though this paper focuses on NN-based methods, some of the latest results on BGL prediction using any other technique are mentioned here, in order to provide a basis for comparison.

Using random forests with grammatical evolution engines, Ignacio et al. generated models of BGL, and selected the models to assemble with bagging techniques. The results for 5 patients were evaluated over a 15-day period according to the EGA classes, 60-min A/B: 97.45%, 120-min: 95.63% [24].

The Kernel Ridge Regression technique was used by Marcus et al. for 11 T1DM patients with 7–50 days of CGM data, with the result of 45-min 1.13 mmol/l RMSE [25].

Liu et al. used a compartmental glucose/insulin model based on a deconvolution method on the CGM signal. Besides CGM, the inputs included the quantity of insulin, CH, and optionally, the type of the absorption speed (slow/medium/fast)—for the latter the authors assumed that breakfast and snack were fast absorption meals, while lunch and dinner were assumed to be medium absorption meals. The results on 10 T1DM and 10 simulated datasets in a two-week clinical trial were 30-min 0.98, 60-min 1.68, 90-min 2.12 and 120-min 2.25 RMSE, the 60-min EGA A/B was 97.1% [26].

3. Methods

The concrete goal of this work is to predict the short term BGL evolution for insulin-dependent DM patients using the following input:

- Baseline (starting) BGL
- Insulin dosing administered by the patient
- Detailed dietary log

Since the essence of the contribution is the more precise modeling of the nutritional input, the absorption model is introduced first.

3.1. The glucose absorption model

There are many methods for modeling nutrient absorption proposed in the literature [27], the most well-known of which is the one used in the Diabetes Advisory System (DIAS) [28]. DIAS uses a one-compartment (stomach) absorption model, without considering the effect of the glycemic index of the various carbohydrates contained in the meal, nor the fiber and other nutrient content. The insulin-glucose dynamics model due to Dalla Man also starts from the gastro-intestinal tract, but it does not use parameters related to the absorption rate of various carbohydrates [29].

In contrast to the above model, the two-compartment model due to Arleth et al., has a separate compartment for the intestine and it can model the timing of the absorption processes, such as the breakdown of

Table 1
Recent studies on NN based BGL prediction. CH means the total CH content of the meal consumed. All CGM data is recorded every 5 min. All RMSE and MAE results are in mmol/l. NDA: no data available, h: hours, d: days, m: months, sim.: simulated, acc.: accuracy.

First Author (Year)	NN Type	Inputs	No. patients or data sets	Length of a single data set	Validation Approach (Train./Valid.)	Results (RMSE/EGA/CG-EGA) for each horizon
Kim (2019)	FNN	CGM	16 real T2DM	NDA	NDA	30-min: 2.1, 45-min: 2.3
Li (2019)	Convolutional RNN	CGM, insulin, CH	10 sim., 10 real	Sim.: 360 d Real: 6 m	50/50%	RMSE Sim.: 30-min: 0.5, 60-min: 1.05; Real: 30-min: 1.17, 60-min: 1.85
Ali (2018)	FNN	CGM only	12 real	14 d	70/30%	15-min: 0.36, 30-min: 0.41, 45-min: 0.45, 60-min: 0.5
Mirshekarian (2017)	RNN	CGM, insulin, CH quantity	10 real	400 measurements	50/50%	30-min: 1.19, 60-min: 2.11
Jankovic (2016)	ARNN vs. RNN	CGM, insulin, CH, physical activity	6 real	CGM: 48 h before and 35 h after exercise	Pre-exercise for training, post-exercise for evaluation	15-min: 0.47, 30-min: 0.98, 45-min: 1.35
Frandes (2016)	ARNN	CGM only	17 real	4–7 d	NDA	30-min: 0.13, 60-min: 0.24, 90-min: 1.23
Zecchin (2014, 2016)	Jump NN	CGM, CH	20 real	2–3 d	10 for training, the other 10 for validation	30-min: 0.92
Zarkogianni (2014, 2015)	adaptive neuro-fuzzy inference system	CGM, BGL change, physical activity	10 real	6 d	10-fold cross-validation	30-min: 0.74, 60-min: 1.26, 120-min: 2.08, CG-EGA acc. in hypo-glycemic range: 60-min: 73.3%, 120-min: 33.7%
Mathiyazhagan (2014)	Adaptive network-based fuzzy inference system	CGM, insulin, CH	2 real	56 d	Both patients: 6 pieces of 2-h CGM records for training.	MAE: 30-min; 1.72, 60-min: 3.16, 120-min: 5.71
Daskalaki (2012)	ARNN	CGM, insulin	30 sim.	8 d	50/50%	30-min:0.2, 45-min: 0.3; CG-EGA acc.: 89% (93% in hypoglycemic range)
Pappada (2011)	FNN	CGM, insulin, CH, emotions, symptoms	27 real	115 CGM h for training, 39 h for validation (calculated)	17 for training, the other 10 for validation	75-min: 2.43; EGA 'A/B' 92.3%, 'A': 62.3%
Sandham (1998)	RNN	fingertip BGL, insulin, CH, illness, stress, pregnancy	2 real	10 d (122 total BGL records)	97 events/25 events	MAE: 1.5

starch to monosaccharide, in finer detail [30]. This model has been chosen for our work. The structure of the model is shown in Fig. 3.

The Arleth model takes the consumed quantities of lipids, proteins, dietary fibers, monosaccharides and starch as inputs. An important feature of the model is the support of a Glycemic Index (GI) parameter that can be attached to a meal item or ingredient because it facilitates the modeling of mixed meals. Ranging from 0 (for water) to 140 (for glucose itself), the GI shows the BGL raising effect of a certain food [31]. Modern dietary databases are expected to contain GI information for each ingredient containing carbohydrates, so a meal can be modeled as a glycemic mix.

The model consists of five parts, for the stomach compartment, the intestinal compartment, the breakdown of starch, the intestinal glucose absorption and the gastric emptying, respectively, using a total of 23 equations. Out of these, the stomach compartment is the most important for our topic which uses 6 simple material balance equations describing the progress of the food through the stomach and the intestine as follows.

$$\mathrm{sProteins}(t_{i+1}) = \mathrm{sProteins}(t_i) + \Delta\mathrm{mProteins}(t_i) - \Delta\mathrm{eProteins}(t_i) \quad (3)$$

$$\mathrm{sLipids}(t_{i+1}) = \mathrm{sLipids}(t_i) + \Delta\mathrm{mLipids}(t_i) - \Delta\mathrm{eLipids}(t_i) \quad (4)$$

$$\mathrm{Fibres}(t_{i+1}) = \mathrm{sFibres}(t_i) + \Delta\mathrm{mFibres}(t_i) - \Delta\mathrm{eFibres}(t_i) \quad (5)$$

$$\mathrm{Monosac}(t_{i+1}) = \mathrm{sMonosac}(t_i) + \Delta\mathrm{mMonosac}(t_i) * \mathrm{CHOAvail} - \Delta\mathrm{eMonosac}(t_i) + \sum_{\mathrm{GI}} \Delta\mathrm{sStarch}_{\mathrm{GI}}(t_i) \quad (6)$$

$$\mathrm{sStarch}_{\mathrm{GI}}(t_{i+1}) = \mathrm{sStarch}_{\mathrm{GI}}(t_i) + \Delta\mathrm{mMonosac}_{\mathrm{GI}}(t_i) * \mathrm{CHOAvail} - \Delta\mathrm{eStarch}_{\mathrm{GI}}(t_i) - \Delta\mathrm{sStarch}_{\mathrm{GI}}(t_i) \quad (7)$$

The above equations use the present material amount (s prefix), the food consumed (m prefix) and the amount conveyed from the stomach into the intestine (e prefix). The CHOAvail constant represents the uptake rate of monosaccharide and starch from the food in the stomach and is set to 0.76. Besides this, the model has 4 more parameters, two in the starch breakdown part, and two in the gastric emptying part. All parameter values are determined for healthy persons. For more details on the model and the parameters, please see Ref. [32]. Though the Arleth model was developed for healthy persons, and there are known differences in the absorption system of DM patients compared to the healthy state, the model is still used in this work with the same parameter set for all patients, as a starting point, due to its simplicity and favorable input set.

That the dynamics of glucose uptake/absorption in the blood is indeed significantly dependent on GI and the presence of low-CH ingredients like fiber as well, is shown in the example below. Here the Arleth model was used to compute the theoretical glucose load curves of two real meals of a patient, similar in total CH content but different in composition (Table 2), and the CGM response was measured for the same patient in the clinical trial (Fig. 2).

As Fig. 4 shows, though the 6618 meal contains ca. 54% more CH, its model-computed absorption peak is lower and the shape of the curve is wider than that of 6615, due to its composition. The CGM-measured BGL responses verify this phenomenon.

3.2. Training data and model training

The FNN structure was chosen for its simplicity. To prepare the inputs for the training, the Lavinia application and the MenuGene dietary expert database [33] were used to find the lipid, protein, dietary fiber, monosaccharide and starch quantity of every logged meal, and then the Arleth model was run to compute its glucose absorption curve. In the next step, the shape of the computed absorption curve was 'quantified' with the following 3 numerical parameters (Fig. 5).

- p1: time elapsed to the peak of the curve [minute]
- p2: time elapsed to 50% of the peak of the curve [minute]
- p3: rate of absorption at the maximum of the curve [g/minute]

The rest two parameters of the training input vector were selected as follows.

Table 2
Components of two meals and their nutritional values.

Meal ID	Food	CH (gr)	Lipids (gr)	Protein (gr)	GI index	Fiber (gr)
6615	Ham	0.16	2.84	9.04	0	0
	Light margarine	0	2.5	0	0	0.06
	Mineral water	0	0	0	0	0
	Bread roll	30.78	0.38	5.08	75	1.78
6618	Mineral water	0	0	0	0	0
	Ham	0.4	7.1	22.6	0	0
	Green pepper	2.7	0.27	1.08	0	2.03
	Tomato	4.6	0.23	1.15	0	1.99
	Bread	39.86	1.27	6.54	66	11.6
	Cold cuts of turkey	0.05	3.75	10.3	0	0

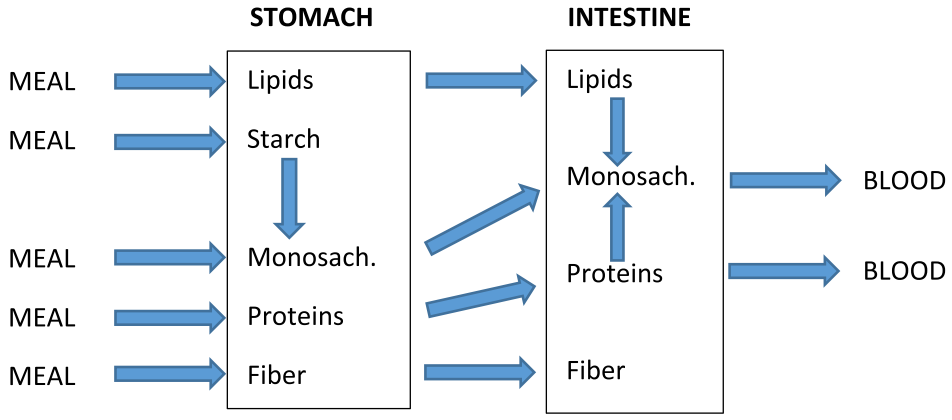

Fig. 3. Structure of the two-compartment glucose absorption model. Arrows show the transport and absorption (transformation) of the nutrients in the two compartments.

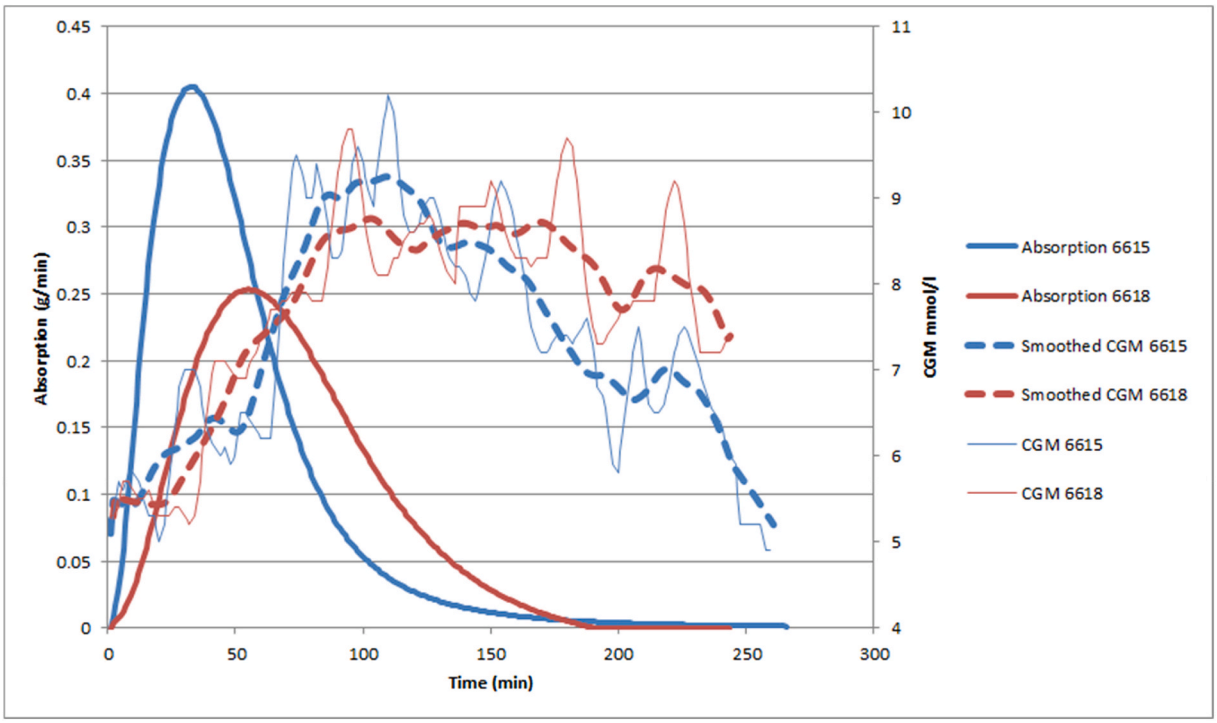

Fig. 4. Glucose absorption in the blood computed by the Arleth model of two alternative meals (solid thick blue and red curves) and the measured CGM values (thin blue and red curves) for a patient of the clinical trial. For clarity, dashed lines show the smoothed version of the CGM. The smoothed curves were calculated over a sliding window of length 30 across neighboring data points.

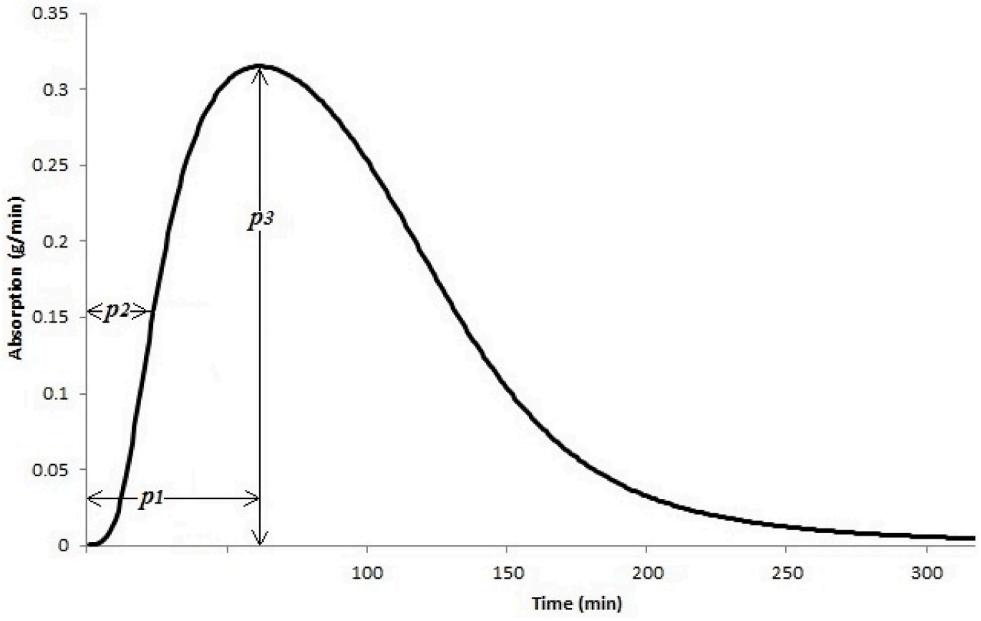

Fig. 5. Meal absorption curve represented by numerical parameters.

- p4: insulin amount [pmol/1000]
- p5: baseline BGL [mmol/l]

3.3. FNN training and evaluation

The FNN used in the training had 5 inputs (see above) and 30 to 120 outputs depending on the desired prediction horizon. As the CGM device used in the trial took a measurement every 2 min, 30 outputs were used for the 1-h, 60 outputs for the 2-h, 90 outputs for the 3-h and 120 outputs for the 4-h long predictions. For the training regime, the Quasi-Newton method was used [34]. The most important algorithmic parameters of the training were determined empirically to achieve the best results; as a result, the number of hidden layers was set to 20, the maximum number of iteration cycles to 118 and the error threshold to 10E-16, meaning that in most cases the iteration limit served as the stop condition. Using more hidden layers or more iteration cycles was found to result in over-training, and hence worse predictions.

To verify our original hypothesis that the use of an absorption model

could be beneficial for the accuracy of the FNN model, the same FNN was also trained with the raw nutrient values of the meals, i.e. without using the absorption model. In this version, the following 5 training inputs were used:

- p1: CH (g)
- p2: Lipids (g)
- p3: Fiber (g),
- p4: insulin amount [pmol/1000]
- p5: baseline BGL [mmol/l]

In order to distinguish the two versions, the code FNN-ABS will be used for the absorption model-based version and FNN-NUT for the version using raw nutrient quantities.

Though, to our best knowledge, all the NN-based BGL prediction methods proposed in the literature use raw (if any) CH values and none uses GI, a sub-version of FNN-NUT has also been implemented in which the p3 parameter was replaced with the 'summary GI' of the logged meal. This was computed as the average of the GI's of the meal's ingredients weighed by the ingredients' quantity.

When comparing the performance of the FNN-ABS to the FNN-NUT, the paired sample *t*-test was used to check for significant differences [35]. Since it can be argued that the 'predictability' of patients may differ, we also evaluated the ANOVA nonparametric test with the patient identifier as a factor, for the comparison of the FNN-NUT-GI and the FNN-ABS methods.

The following figures of merit were used in the evaluation.

- Average absolute error (MAE)
- Root mean square error (RMSE)
- Percentage of predicted values in the 'clinically acceptable' EGA classes A and B. As a 'worst case' scenario, the FNN-NUT and FNN-ABS percentages in the classes D and E were also compared.

The predictions were evaluated on the 1, 2 and 3-h horizons.

It should be noted that our primary goal is to reduce the RMSE, as the chosen model optimization method (FNN) considers only the differences between the model output and the measured CGM values, and no EGA classes. The EGA results were only evaluated in the paper for completeness and comparison to other published results that include only EGA results without RMSE values.

3.4. Data used for training and validation

8 volunteers had participated in the clinical study when the data set was finalized. The data sets were examined concerning the accuracy and completeness of the CGM and lifestyle log data. As a result, 3 patients were excluded from the study due to their lack of cooperation resulting in incomplete lifestyle logs. In total, 84 days of CGM data and lifestyle logs were available, containing 365 meals and 391 insulin injections. Table 3 shows a summary of the data available for the study.

Table 3
Properties of the datasets used for the study.

Patient ID	P01	P02	P03	P04	P05	Total
Gender	Female	Male	Male	Male	Male	–
Age	56	47	69	63	23	–
Type	T2DM	T2DM	T2DM	T2DM	T1DM	–
HbA1c [%]	9.3	11.0	6.7	9.0	7.7	–
Height [cm]	169	175	160	183	197	–
Weight [kg]	77	133	50	97	82	–
Log length [days]	24	23	15	12	10	84
# meals	97	109	75	39	45	365
Breakfast	20	23	15	10	7	75
Lunch	22	23	14	11	9	79
Dinner	19	23	13	11	9	75
Other	36	40	33	7	20	136
# insulin	87	92	107	58	47	391
# CGM records	5523	14,558	5420	7692	2311	35,504

In the next step, meals were excluded that had no corresponding pre-meal insulin dosing information in the log and those whose total CH content was less than 5 g. A meal was also excluded if the time to the patient's next meal was less than our minimal prediction horizon, 1 hour. After this process, a total of 167 meals were left that could be used in the study. Table 4 shows the patient-wise distribution of the meals used for training (2/3) and validation (1/3). The separation of the training data from the validation data was performed on the basis of the date of the meal: the first 2/3 at the beginning of the trial was used for training and the last 1/3 at the end for validation.

In order to check the effect of the data separation scheme between training and validation data, a 3-fold cross validation was performed for the proposed FNN-ABS method. This meant three training sessions for each patient, with another 2/3 of the data. If the performance of the three trained models is similar, then the approach can be considered robust with respect to data selection.

Since the considerable differences in the number of meals available for training may have an effect on the patient-wise performance of the models, a test to compare the FNN-ABS and FNN-NUT method was also run in which the number of training meals was set to that of P03 (17, the lowest number). The 17 training meals in this test were selected from the available meals for the P01, P02, P04 and P05 patients at random.

3.5. The clinical study

The clinical study was performed at the Cardiac Rehabilitation Institute of the Military Hospital, Balatonfüred, Hungary. The study included insulin-dependent T1DM and T2DM patients taking part in 3-week rehabilitation courses between April and August 2019, with daily activities similar to everyday life. The patients were under continuous medical and dietary supervision and an informed consent was obtained from the patients as a prerequisite to enter the trial. The CGM system used was the Medtrum's S7 EasySense CGM System [36], which registered subcutaneous glucose values every 2 min. Mérykék 800 fingertip BGL sensors were used to record the baseline BGL.

3.6. Ethical considerations

The study protocol was approved on 9 April 2018 by the National Institute of Pharmacy and Nutrition (OGYÉI), Budapest, Hungary, chaired by Péter Bunyitai, under the submission number OGYÉI/4778/2018. The protocol was designed and implemented in compliance with the World Medical Association Declaration of Helsinki on Ethical Principles for Medical Research Involving Human Subjects.

3.7. Data processing tools

Mongodb database technology was used for storing the dietary and insulin logs [37]. For calculating the ingredient quantity and the GI values from the dietary logs as required by the absorption model, the Lavinia application and the MenuGene dietary expert database was used

Table 4
Number of monitoring days and meals used for FNN training and validation, for each patient.

Patient	# days	# meals (training)	# meals (validation)	# all meals
P01	23	19	10	29
P02	22	28	15	43
P03	14	17	9	26
P04	11	22	12	34
P05	15	22	13	35
SUM	85	108	59	167

[38].

The absorption model was implemented according to the original paper due to Arleth [30], in the form of a custom desktop application for BGL modeling [5].

The FNN used for the study was the OpenNN library, a feed-forward, multilayer perceptron network implemented as a C++ open-source library [39].

Microsoft Excel 2013 was used for statistical analysis and visualization.

4. Results

Table 5 shows the results achieved by the three training method versions, with the best results highlighted in bold face, while Table 6 shows the FNN-ABS results in 3-fold cross validation for comparison.

4.1. Comparison of FNN-NUT/FNN-NUT-GI to FNN-ABS

It was found that the FNN-NUT performed worse on all horizons and with all figures of merit, though the difference was not always statistically significant. Table 7 shows the average gain of the prediction accuracy as a percentage, in favor of FNN-ABS.

FNN-NUT was found to produce more EGA D and E class (worst case) scenarios than FNN-ABS as shown in Table 8.

The FNN-NUT-GI version performed slightly better than the FNN-NUT. Table 9 shows the average gain of its prediction accuracy as a percentage, in favor of FNN-ABS.

Detailed results of the test with the same number of meals (i.e. 17 meals only) used for training in the FNN-ABS and FNN-NUT methods are shown in Table 10.

Table 11 shows the summary comparison of the above test.

Finally, for a qualitative visual comparison, Fig. 6 shows the first 60 min of the BGL measured by the CGM and predicted by the NUT and ABS methods, belonging to a typical meal of the P03 patient.

5. Discussion

Table 5 shows that there are considerable differences among the 5 patients concerning the accuracy of the predictions. This may be partly due to the quality of the input data or even more to the length of the training sample: P02 and P04 had much more samples than P01 and P03, and performed ca. 30% better. A NN is naturally expected to produce a better model if a longer training sequence is available. However, results for P05 were much worse than those for P04 despite the nearly same number of their logged meals. Also, the test with the same number

Table 6
Results of the 3-fold cross-validation for the FNN-ABS. The V1 version used the first 2/3 time period for training (also shown in detail in Table 5). V2 used the first and last 1/3, while V3 used the last 2/3 for training.

Version	Figure of merit	1-h	2-h	3-h
V1	MAE	0.965	1.550	1.870
	RMSE	1.120	1.755	2.176
	EGA acceptable	96.46%	98.13%	96.03%
V2	MAE	1.253	1.932	1.971
	RMSE	1.666	2.316	2.518
	EGA acceptable	95.31%	95.88%	93.80%
V3	MAE	1.020	1.608	1.937
	RMSE	1.244	1.859	2.290
	EGA acceptable	98.95%	98.55%	97.73%

Table 7
Prediction accuracy gain of FNN-ABS over FNN-NUT. P-values show the significance level of the paired sample *t*-test.

Figure of merit	1-h	2-h	3-h	ALL horizons together
MAE	31.69% P < 0.01	24.56% p<0.01	7.65% p>0.05	**21.30%**
RMSE	38.35% p < 0.01	30.76% p<0.01	15.81% p>0.5	**28.31%**
EGA A/B	1.02% p > 0.05	3.28% p = 0.017	2.97% p > 0.05	**2.42%**

Table 8
FNN-ABS and FNN-NUT percentage of predictions in the EGA D and E classes. Patient-wise results are omitted for brevity.

Version	1-h	2-h	3-h
FNN-ABS	3.59%	1.87%	2.95%
FNN-NUT	4.18%	5.08%	6.48%
Improvement	14.14%	63.30%	54.52%

Table 5
Results of the FNN-ABS, FNN-NUT and FNN-NUT-GI methods for the five patients (best results in bold) P05 is the only T1DM patient, the others are T2DM.

		FNN-ABS			FNN-NUT			FNN-NUT-GI		
Patient	Figure of merit	1 h	2 h	3 h	1 h	2 h	3 h	1 h	2 h	3 h
P01	AVG	1.651	2.095	3.070	1.450	2.633	2.795	1.430	2.346	2.736
	RMSE	1.908	2.426	3.568	2.096	3.402	3.838	2.006	2.982	3.738
	EGA acceptable	86%	95%	87%	93%	91%	89%	90%	92%	88%
P02	AVG	0.768	1.208	1.504	1.421	1.548	1.628	1.084	1.524	1.598
	RMSE	0.887	1.434	1.721	1.647	1.996	2.019	1.456	**1.812**	2.004
	EGA acceptable	98%	99%	99%	98%	99%	99%	98%	99%	99%
P03	AVG	1.060	1.900	2.147	2.245	2.550	2.617	1.885	2.240	2.478
	RMSE	1.237	2.283	2.484	3.214	3.641	3.398	2.684	3.211	3.418
	EGA acceptable	96%	100%	95%	83%	83%	81%	87%	86%	83%
P04	AVG	0.477	0.961	1.349	1.032	1.623	1.501	0.973	1.572	1.481
	RMSE	**0.540**	**1.126**	**1.571**	**1.293**	**1.967**	**1.906**	**1.232**	1.843	**1.736**
	EGA acceptable	100%	100%	99%	100%	100%	99%	99%	100%	99%
P05	AVG	1.047	1.582	1.660	1.146	1.969	1.965	1.132	1.753	1.912
	RMSE	1.236	1.825	1.976	1.312	2.248	2.336	1.341	2.203	2.273
	EGA acceptable	100%	97%	98%	99%	97%	92%	98%	98%	94%
All datasets										
All datasets	AVG	0.965	1.550	1.870	1.412	2.054	2.025	1.253	1.932	1.971
	RMSE	**1.120**	**1.755**	**2.176**	1.816	2.535	2.585	1.666	2.316	2.518
	EGA acceptable	96%	98%	96%	95%	95%	93%	95%	96%	94%

Table 9
Prediction accuracy gain of FNN-ABS over FNN-NUT-GI. P-values show the significance level of the paired sample *t*-test and the one-way ANOVE test.

Figure of merit	1-h	2-h	3-h	ALL horizons together
MAE	22.25%	19.79%	5.09%	**15.71%**
t-test	$p < 0.01$	$p<0.01$	$p>0.05$	
ANOVA	$p<0.05$	$p>0.05$	$p > 0.05$	
RMSE	32.80%	24.22%	13.59%	**23.54%**
t-test	$p < 0.01$	$p<0.01$	$p>0.05$	
ANOVA	$p = 0.01$	$p = 0.05$	$p > 0.05$	
EGA A/B	1.19%	2.30%	2.33%	**1.94%**
t-test	$p > 0.05$	$p > 0.05$	$p > 0.05$	
ANOVA	$p > 0.05$	$p > 0.05$	$p > 0.05$	

of meals (17) used for training showed that, as it could be expected, the overall prediction performance decreased, but still the P04 and P02 results were the best, though the differences decreased (Tables 10 and 11). This phenomena may be due to fundamental differences in the 'predictability' of humans: our model missed several factors that are hard to quantify, but which are known to influence BGL, such as emotions, and it can be stipulated that those patients for whom the impact of such factors is relatively stronger are harder to predict.

Personal variations in predictability naturally call for a larger number of patients to validate a prediction approach. As Table 1 shows, this number varies between 2 and 37 in the reported studies. These numbers are relatively low compared to other clinical research fields, which may be explained by the difficulties associated with the acquisition of high quality dietary and especially CGM data from volunteers in a properly managed clinical trial.

Since the BGL of T1DM patients usually shows more variations and more extreme BGL values than T2DM patients, it could be expected that the BGL prediction performance of a NN will also be worse for T1DM. There was only one T1DM patient (P05) in our study, for whom the FNN-ABS results were 'average' i.e. worse than P02 and much worse than P04, yet slightly better than P03 and much better than P01. Based on these results, the above expectation can neither be confirmed nor dismissed. More patients would obviously be needed, especially T1DM, to make a statistically confirmed statement about the connection between the DM type and prediction performance.

The cross-validation was performed to track the effect of training data selection on the performance of the trained model (Table 6). The best results (1.120, 1.755, 2.176 mmol/l RMSE on the 1-h/2-h/3-h horizons) were produced by the default scheme that used the first 2/3 time period, yet the worst version (V2 with 1.244, 1.859, 2.290 mmol/l RMSE) fell behind by only less than 0.13 mmol/l (ca. 12%). This shows that the FNN could quite robustly learn the behavior of the system.

As expected, the FNN-NUT-GI version outperformed the FNN-NUT version (see Table 5), showing that the 'summary' GI computed for the meal was more valuable information for the learning than the fiber content, at least with this limited amount of data, though fiber is known to have an effect on GI.

The superior performance of the ABS over the NUT/NUT-GI methods, shown in Tables 7 and 8, verified our startup hypothesis that additional domain knowledge formulated in an absorption model will improve the predictability of a complex system such as the human absorption system combined with the BGL regulation system. This conclusion may seem to contradict the results of Zecchin et al., who did not find much difference with respect to whether CH data was included in the training scenario [17]—however, they did not use an absorption model, only 'raw' CH values.

Table 7 also shows that in general, while the accuracy of both methods decreases naturally for longer horizons, the performance advantage of the ABS method over the NUT also decreases (RMSE 1-h: 38.35%, 2-h: 30.76%, 3-h: 15.81%, for MAE 31.69%, 24.56%, 7.65%, respectively). This phenomenon may be explained by the accumulation of 'noise' i.e. error due to not modeled factors, in the prediction error as time passes by. As the BGL curve becomes harder to explain by the absorption model, the ABS method loses its power over the simpler NUT method. EGA errors do not follow this rule as the differences are very small (1–3%) and not significant, but one should not forget that during

Table 11
Prediction accuracy gain of FNN-ABS over FNN-NUT with the same number of training meals. P-values show the significance level of the paired sample *t*-test.

Figure of merit	1-h	2-h	3-h	ALL horizons together
MAE	29.70% $p < 0.01$	26.22% $p<0.01$	10.58% $p>0.05$	**23.70%**
RMSE	35.99% $p < 0.01$	30.63% $p<0.01$	15.63% $p>0.05$	**28.31%**
EGA A/B	1.30% $p < 0.01$	4.49% $p = 0.013$	4.04% $p > 0.05$	**3.28%**

Table 10
Results for the same number of meals for each patient.

		FNN-ABS			FNN-NUT		
Patient	Figure of merit	1-h	2-h	3-h	1 h	2-h	3-h
P01	AVG	1.752	2.225	3.153	1.583	2.883	3.055
	RMSE	2.020	2.575	3.677	2.206	3.583	4.042
	EGA acceptable	85.15%	94.39%	85.69%	92.59%	90.00%	87.78%
P02	AVG	0.983	1.433	1.998	1.957	2.278	2.327
	RMSE	1.131	1.681	2.247	2.044	2.540	2.548
	EGA acceptable	96.33%	98.18%	98.02%	96.32%	97.96%	98.40%
P03	AVG	1.060	1.900	2.147	2.245	2.550	2.617
	RMSE	1.237	2.283	2.484	3.214	3.641	3.398
	EGA acceptable	95.56%	99.81%	95.06%	82.96%	82.78%	81.48%
P04	AVG	0.977	1.331	1.598	1.425	1.820	1.842
	RMSE	**1.055**	**1.417**	**1.846**	1.662	**2.120**	**2.178**
	EGA acceptable	100.00%	100.00%	98.52%	100.00%	100.00%	98.64%
P05	AVG	1.306	1.983	2.202	1.436	2.493	2.569
	RMSE	1.442	2.106	2.349	**1.631**	2.620	2.772
	EGA acceptable	99.94%	96.34%	97.16%	98.89%	96.06%	88.97%
All datasets							
All datasets	AVG	1.216	1.774	2.219	1.729	2.405	2.482
	RMSE	1.377	2.012	2.521	2.151	2.901	2.988
	EGA acceptable	95.40%	97.75%	94.89%	94.15%	93.36%	91.05%

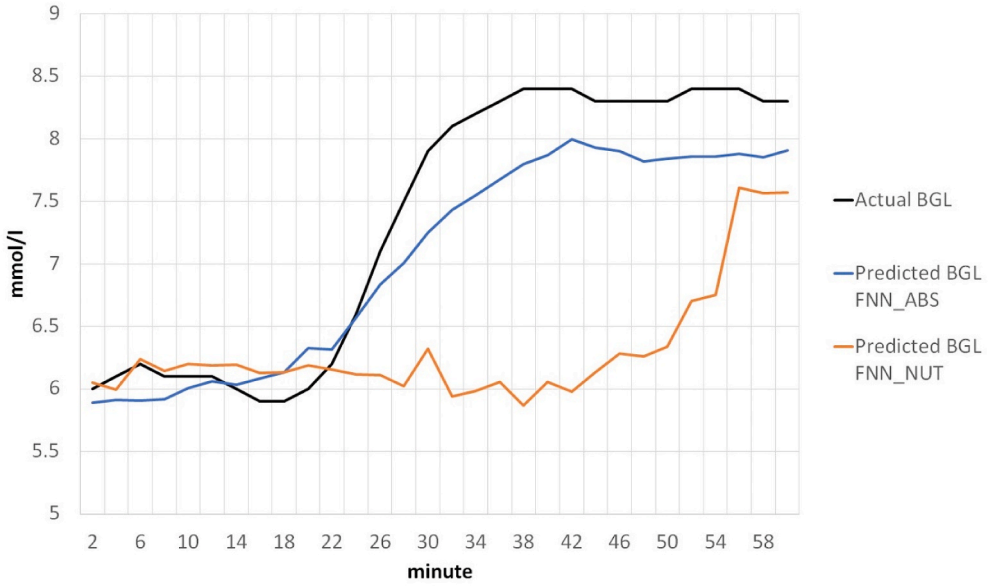

Fig. 6. A typical meal's measured BGL (black line) and predicted (FNN-ABS: blue, FNN-NUT: orange) BGL values of the P03 patient for 1 h. The abscissa shows the measured and predicted BGL curves in mmol/l of the subject P03, in the function of time (shown in minutes on the x-axis).

training, an FNN always tries to minimize the difference between the measured and predicted values, which in our case was the MAE error and *not* the best EGA classification. The apparent contradiction that P02, P04 and P05 have nearly the same EGA yet quite different MAE or RMSE values can be explained by the quite wide bands of the EGA classes that do not punish an error until it crosses a region boundary. The RMSE shows a much clearer picture of the real power of the model. The improvement in terms of EGA are more apparent in the D and E (worst case) categories, as shown by Table 8.

5.1. Comparison to related work

First, the new results can be compared with our own earlier results using a state-of-the-art BGL regulation model, the parameters of which were trained (personalized) with various methods of optimization [4]. That study used a similar clinical protocol, the same dietary database, and it was supervised by the same medical team as this trial. The 60-min RMSE result with the best algorithmic setup was 1.62 mmol/l, considerably worse than the 1.12 mmol/l of the FNN-ABS, proving that in the previous study the possibly over-simplified BGL model itself was a limitation.

When our results are numerically compared with those of other studies, one should not forget that for a fair comparison, the various methods should be run on the same data sets which are not always available for sharing, due to restrictions of the clinical trials. In our specific case, dietary log data from other trials could not be used anyway, because our method uses fiber, lipid, GI etc. values which are not included in other trials and which can only be computed from a localized (culture-specific) dietary expert database. That being said, it can be stated that the best new result (1.12 mmol/l) is very promising compared to the 60-min, outpatient and CH-insulin based RMSE results of the literature survey, (Li: 1.85, Mirshekarian: 2.11, Mathiyazhagan: 3.16 mmol/l). Though there are far better results than this as well on the 45/60-min horizon (cf. Ali: 0.5, Frandes: 0.24, Daskalaki: 0.3 on 45-min, Zarkogianni: 1.26), but these models do not consider the (possibly hectic) CH input of the patient, so a direct comparison would not be fair. Another point that one must bear in mind is that CGM-only predictors require a continuous CGM data input *even when the model is already trained*, which is not possible for a large part of the DM outpatient community for financial reasons. It could also be argued that it is not fair to compare our mostly T2DM-based results to studies with only T1DM patients. The only relevant study that was found in the literature with T2DM patients, due to Kim et al., had worse RMSE results (2.3 mmol/l) on the 45-min horizon than any of our patients on the 60-min horizon, including the single T1DM patient. For a more direct comparison, more T1DM patients would be needed.

Published RMSE results of *simulated datasets* are considerably better than our results [14,21]. However, BGL is influenced by such factors as the mental state, emotions, sudden movement and environmental changes etc., which form an inherent part of an outpatient's daily life, but which even sophisticated simulators cannot consider. However, the effect of these factors appears as a 'noise' imposed on the real-life measured BGL curve, which makes the accuracy of predictions for real patients worse than those validated on simulators (see these differences in e.g. Ref. [21]). Also, simulators cannot account for the significant variances in the personal parameters of the metabolism. Therefore, the direct comparison of simulator vs. real patients' results would naturally be biased in favor of the simulators.

Figs. 7 and 8 show a graphical evaluation of the RMSE results. Fig. 7 compares the FNN-ABS best RMSE value to our previous best results (Math. model GA, [4]).

There are only a few published RMSE results for the 120-min and no results for the 180-min horizons, making it hard to evaluate our results (1.75 mmol/l and 2.75 mmol/l, respectively) on these horizons. For 120 min, Mathiyazhagan reported 5.71 mmol/l using CH input and Zarkogianni 2.08 mmol/l, the latter being close to our result (1.75 mmol/l). These values are graphically compared in Fig. 8 with the performance of the NUT methods.

As for the EGA evaluation, though CG-EGA classes are not exactly comparable to EGA classes, our 96.46% result on the 60-min horizon for the 'clinically acceptable' classes compares very favorably to Pappada's 92.3% (75-min, EGA), Daskalaki's 89% (93% in the hypoglycemic range, 45-min, CG-EGA), and Zarkogianni's 73.3% (60-min, CG-EGA). It also a strength of our model that the EGA accuracy does not decrease significantly on the 120-min and 180-min horizons.

6. Conclusion and future work

The paper presented a new outpatient BGL predicting method that is based on the application of an absorption model to generate training input for a neural network. For the successful training of the network, a good quality dietary and insulin log as well as the CGMS data is needed

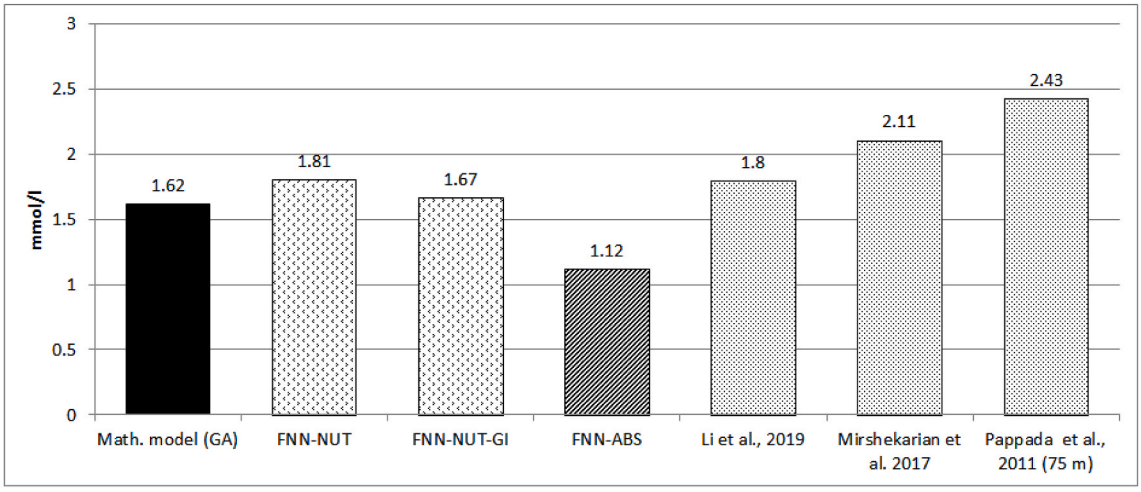

Fig. 7. Prediction results compared with the NUT methods and results from other studies on the 1-h prediction horizon (75 min for Pappada). Other studies are identified by the author and year. The abscissa shows the RMSE in mmol/l of the methods enlisted on the x-axis.

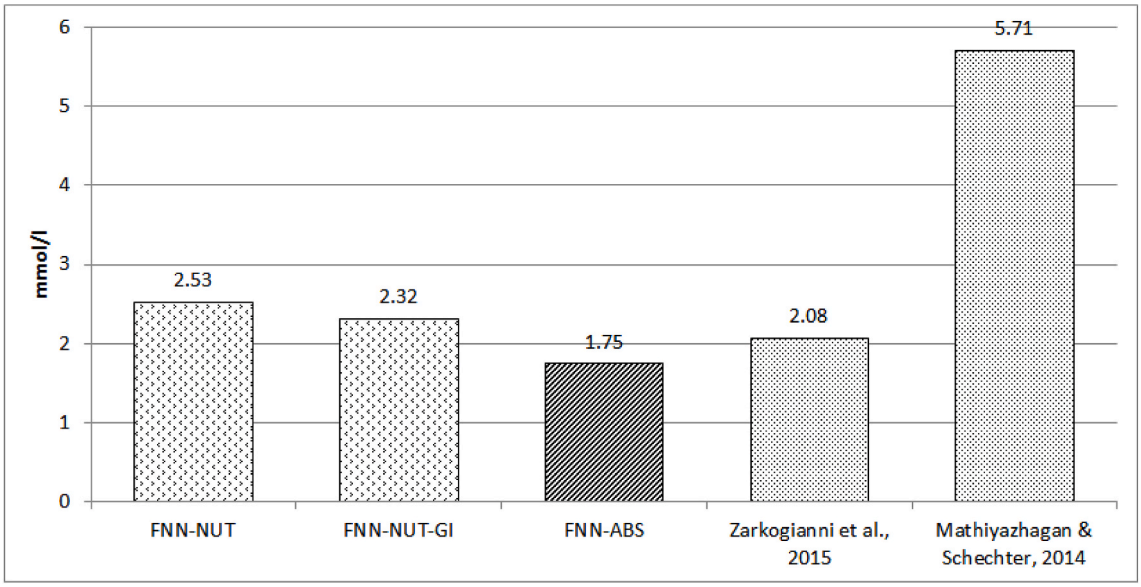

Fig. 8. Prediction results compared with the NUT methods and results from other studies on the 2-h prediction horizon. Other studies are identified by the author and year. The abscissa shows the RMSE in mmol/l of the methods enlisted on the x-axis.

for a period of ca. one week. The trained model uses only a startup (fingertip) BGL, and the dietary and insulin log for a 60-min to 180-min prediction, therefore it is applicable in practice for outpatients without continuous access to a CGMS device. The RMSE and EGA accuracy of the prediction (60-min: 1.12 mmol/l, 96.46% clinically acceptable) is better than those published results to which our method is directly comparable, and it also surpasses the authors' previous results using personalized BGL control models. The study also showed that the application of the absorption model has significantly decreased the RMSE prediction error at least on the 60- and 120-min horizons compared to a CH-only version, so the integration of dietary science has indeed contributed to the success of the model.

Future research in this field must include, most of all, new trials with more patients to verify these promising results. On a larger sample, the inclusion of insulin types (basal vs. bolus insulin), the physical activity, and the presence of emotional or mental stress as training inputs are also expected to improve the accuracy of the prediction—if these factors could be monitored in a reliable way.

Data availability

The detailed, anonymized data sets used to train the neural network are available at request from the corresponding author and as a supplement to this article. Other data related to the clinical trial may be released upon application to the institutional Ethical Committee of the Military Hospital, which can be contacted at Magyar Honvédség Egészségügyi Központ Intézményi és Regionális Kutatásetikai Bizottsága, Róbert Károly körút 44, 1134 Budapest, Hungary.

Authors' contributions

All authors have contributed to the conception and design of the study, the acquisition, analysis and interpretation of data, and the drafting of the article. Main roles are as follows. Rebaz Karim: data processing, neural network training and analysis; István Kósa: medical supervision and clinical advice, István Vassányi: information technology supervision and advice, manuscript editing. All authors have approved the final article.

Declaration of competing interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

The authors wish to express their gratitude to Brigitta Szálka for her help in the execution of the clinical trial and also for her valuable advice on dietary science, and to Péter Gyuk for his continuous technical advice and for his groundbreaking work in model-based BGL prediction that paved the way for this study. The Medtrum S7 EasySense CGM System and the fingertip BGL sensors were made available for the clinical study by Di-Care Inc., Hungary.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Abbreviations

NN	Artificial Neural Network
ARNN	Autoregressive Artificial Neural Network
MAE	Mean Absolute Error
BGL	Blood Glucose Level
CGMS	Continuous Glucose Monitoring System
CH	Carbohydrates
DM	Diabetes Mellitus
EGA	Error Grid Analysis
FNN	Feed-forward Artificial Neural Network
GA	Genetic Algorithms
RMSE	Root Mean Square Error
RNN	Recurrent Artificial Neural Network
T1DM	Type 1 Diabetes Mellitus
T2DM	Type 2 Diabetes Mellitus

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2020.103956.

References

[1] O.P. Adams, The impact of brief high-intensity exercise on blood glucose levels, Diabetes. Metab. Syndr. Obes. 6 (2013) 113–122, https://doi.org/10.2147/DMSO.S29222.

[2] V.L. Goetsch, D.J. Wiebe, L.G. Veltum, B. van Dorsten, Stress and blood glucose in type II diabetes mellitus, Behav. Res. Ther. 28 (1990) 531–537, https://doi.org/10.1016/0005-7967(90)90140-E.

[3] I. Ajmera, M. Swat, C. Laibe, N. Le Novère, V. Chelliah, The impact of mathematical modeling on the understanding of diabetes and related complications, CPT Pharmacometrics Syst. Pharmacol. 2 (2013) 54, https://doi.org/10.1038/psp.2013.30.

[4] P. Gyuk, I. Vassanyi, I. Kosa, Blood glucose level prediction for diabetics based on nutrition and insulin administration logs using personalized mathematical models, J. Healthc. Eng. (2019) 8605206, https://doi.org/10.1155/2019/8605206, 2019.

[5] P. Gyuk, T. Lorincz, R.A.H. Karim, I. Vassanyi, Diabetes lifestyle support with improved glycemia prediction algorithm, in: Proc. Seventh Int. Conf. EHealth, Telemedicine, Soc. Med. (ETELEMED 2015), 2015, pp. 95–100. https://www.iaria.org/conferences2015/eTELEMED15.html.

[6] R.A.H. Karim, I. Vassányi, I. Kósa, Long-acting insulin management for blood glucose prediction models, Biomed. Res. 30 (2019) 42–47, https://doi.org/10.5281/zenodo.3922332.

[7] M.A. Nielsen, Neural Networks and Deep Learning, Determination press, San Francisco, CA, USA, 2015. http://neuralnetworksanddeeplearning.com.

[8] W.L. Clarke, The original Clarke error grid analysis (EGA), Diabetes Technol. Therapeut. 7 (2005) 776–779, https://doi.org/10.1089/dia.2005.7.776.

[9] B.P. Kovatchev, L.A. Gonder-Frederick, D.J. Cox, W.L. Clarke, Evaluating the accuracy of continuous glucose-monitoring sensors, Diabetes Care 27 (2004) 1922, https://doi.org/10.2337/diacare.27.8.1922. LP – 1928.

[10] W. Sandham, D. Nikoletou, D. Hamilton, K. Paterson, A. Japp, C. MacGregor, Blood glucose prediction for diabetes therapy using a recurrent artificial neural network, in: 9th Eur. Signal Process. Conf, 1998, pp. 1–4 (EUSIPCO 1998), https://dblp.org/rec/conf/eusipco/1998.html.

[11] A.Z. Woldaregay, E. Årsand, S. Walderhaug, D. Albers, L. Mamykina, T. Botsis, G. Hartvigsen, Data-driven modeling and prediction of blood glucose dynamics: machine learning applications in type 1 diabetes, Artif. Intell. Med. 98 (2019) 109–134, https://doi.org/10.1016/j.artmed.2019.07.007.

[12] J. Ben Ali, T. Hamdi, N. Fnaiech, V. Di Costanzo, F. Fnaiech, J.-M. Ginoux, Continuous blood glucose level prediction of type 1 diabetes based on artificial neural network, Biocybern. Biomed. Eng. 38 (2018) 828–840, https://doi.org/10.1016/j.bbe.2018.06.005.

[13] M. Frandes, B. Timar, D. Lungeanu, A risk based neural network approach for predictive modeling of blood glucose dynamics, Stud. Health Technol. Inf. 228 (2016) 577–581, https://doi.org/10.3233/978-1-61499-678-1-577.

[14] E. Daskalaki, A. Prountzou, P. Diem, S.G. Mougiakakou, Real-Time adaptive models for the personalized prediction of glycemic profile in type 1 diabetes patients, Diabetes Technol. Therapeut. 14 (2012) 168–174, https://doi.org/10.1089/dia.2011.0093.

[15] K. Zarkogianni, K. Mitsis, E. Litsa, M.-T. Arredondo, G. Fico, A. Fioravanti, K. S. Nikita, Comparative assessment of glucose prediction models for patients with type 1 diabetes mellitus applying sensors for glucose and physical activity monitoring, Med. Biol. Eng. Comput. 53 (2015) 1333–1343, https://doi.org/10.1007/s11517-015-1320-9.

[16] S.M. Pappada, B.D. Cameron, P.M. Rosman, R.E. Bourey, T.J. Papadimos, W. Olorunto, M.J. Borst, Neural network-based real-time prediction of glucose in patients with insulin-dependent diabetes, Diabetes Technol. Therapeut. 13 (2011) 135–141, https://doi.org/10.1089/dia.2010.0104.

[17] C. Zecchin, A. Facchinetti, G. Sparacino, C. Cobelli, How much is short-term glucose prediction in type 1 diabetes improved by adding insulin delivery and meal content information to CGM data? A proof-of-concept study, J. Diabetes Sci. Technol. 10 (2016) 1149–1160, https://doi.org/10.1177/1932296816654161.

[18] C. Zecchin, A. Facchinetti, G. Sparacino, C. Cobelli, Jump neural network for real-time prediction of glucose concentration. Artif. Neural Networks, 2014, pp. 245–259, https://doi.org/10.1007/978-1-4939-2239-0.

[19] M. V Jankovic, S. Mosimann, L. Bally, C. Stettler, S. Mougiakakou, Deep prediction model: the case of online adaptive prediction of subcutaneous glucose, in: 2016 13th Symp. Neural Networks Appl., IEEE, 2016, pp. 1–5.

[20] N. Mathiyazhagan, H.B. Schechter, Soft computing approach for predictive blood glucose management using a fuzzy neural network, in: IEEE Conf. Norbert Wiener 21st Century Driv. Technol. Futur. 21CW 2014 - Inc. Proc. 2014 North Am. Fuzzy Inf. Process. Soc. Conf. NAFIPS 2014, Conf. Proc., 2014, pp. 1–3, https://doi.org/10.1109/NORBERT.2014.6893906, 2014.

[21] K. Li, J. Daniels, C. Liu, P. Herrero-Vinas, P. Georgiou, Convolutional recurrent neural networks for glucose prediction, IEEE J. Biomed. Heal. Informatics (2019) 603–613, https://doi.org/10.1109/JBHI.2019.2908488.

[22] S. Mirshekarian, R. Bunescu, C. Marling, F. Schwartz, Using LSTMs to learn physiological models of blood glucose behavior, Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS (2017) 2887–2891, https://doi.org/10.1109/EMBC.2017.8037460.

[23] D. Kim, H.-B. Lee, Y. Kim, S. Kim, S.-J. Lee, S. Chun, 944-P: blood glucose prediction for type 2 diabetic patients using machine learning, Diabetes 68 (2019) 944-P, https://doi.org/10.2337/db19-944-P.

[24] J.I. Hidalgo, M. Botella, J.M. Velasco, O. Garnica, C. Cervigón, R. Martínez, A. Aramendi, E. Maqueda, J. Lanchares, Glucose forecasting combining Markov chain based enrichment of data, random grammatical evolution and Bagging, Appl. Soft Comput. 88 (2020) 105923, https://doi.org/10.1016/j.asoc.2019.105923.

[25] Y. Marcus, R. Eldor, M. Yaron, S. Shaklai, M. Ish-Shalom, G. Shefer, N. Stern, N. Golan, A.Z. Dvir, O. Pele, M. Gonen, Improving blood glucose level predictability using machine learning, Diabetes. Metab. Res. Rev. n/a (2020) e3348, https://doi.org/10.1002/dmrr.3348.

[26] C. Liu, J. Vehí, P. Avari, M. Reddy, N. Oliver, P. Georgiou, P. Herrero, Long-term glucose forecasting using a physiological model and deconvolution of the continuous glucose monitoring signal, Sensors 19 (2019) 4338, https://doi.org/10.3390/s19194338.

[27] A. Makroglou, J. Li, Y. Kuang, Mathematical models and software tools for the glucose-insulin regulatory system and diabetes: an overview, Appl. Numer. Math. 56 (2006) 559–573, https://doi.org/10.1016/J.APNUM.2005.04.023.

[28] S. Andreassen, O.K. Hejlesen, R. Hovorka, D.A. Cavan, The Diabetes Advisory System-an IT approach to the management of insulin dependent diabetes, Stud. Health Technol. Inf. (1996) 1079–1083, https://doi.org/10.3233/978-1-60750-878-6-1079.

[29] C.D. Man, R.A. Rizza, C. Cobelli, Meal simulation model of the glucose-insulin system, IEEE Trans. Biomed. Eng. 54 (2007) 1740–1749, https://doi.org/10.1109/TBME.2007.893506.

[30] T. Arleth, S. Andreassen, M. Orsini-Federici, A. Timi, M. Massi-Benedetti, A model of glucose absorption from mixed meals, IFAC Proc 33 (2000) 307–312, https://doi.org/10.1016/S1474-6670(17)35533-7.

[31] K. Foster-Powell, S.H.A. Holt, J.C. Brand-Miller, International table of glycemic index and glycemic load values, Am. J. Clin. Nutr. 76 (2002) 5–56, https://doi.org/10.1093/ajcn/76.1.5, 2002.

[32] T. Arleth, S. Andreassen, M. Orsini-Federici, A. Timi, M. Massi-Benedetti, Optimisation of a Model of Glucose Absorption from Mixed Meals, Alborg Uni, 2005, pp. 1–28.

[33] I. Vassányi, I. Kósa, A.H.K. Rebaz, M. Nemes, B. Szálka, Effectiveness of mobile personal dietary logging, in: P. Kommers, P. Isaías (Eds.), 13th Int. Conf. E-Society, IADIS Press, 2015, ISBN 978-989-8533-32-6, pp. 288–293. http://esociety-conf.org/oldconferences/2015/.

[34] C.G. Broyden, Quasi-Newton methods and their application to function minimisation, Math. Comput. 21 (1967) 368–381, https://doi.org/10.2307/2003239.
[35] M. Xu, D. Fralick, J.Z. Zheng, B. Wang, X.M. Tu, C. Feng, The differences and similarities between two-sample T-test and paired T-test, Shanghai Arch. Psychiatry 29 (2017) 184–188, https://doi.org/10.11919/j.issn.1002-0829.217070.
[36] Medtrum, Medtrum S7 EasySense® CGM system. https://www.medtrum.com/S7.html, 2019. (Accessed 13 April 2020).
[37] MongoDB, MongoDB. Database. https://www.mongodb.com, 2020. (Accessed 10 April 2020).
[38] I. Kósa, I. Vassányi, M. Nemes, K.H. Kálmánné, B. Pinter, L. Kohut, A fast, android based dietary logging application to support the life change of cardio-metabolic patients, in: Glob. Telemed. EHealth Updat. Knowl. Resour., International Society for Telemedicine & eHealth Grimbergen, Belgium, 2014, pp. 553–556. https://www.isfteh.org/media/global_telemedicine_and_ehealth_updates_knowledge_resources._vol.7_2014.
[39] P. Martin, C. Barranquero, J. Sanchez, G. Gestoso, J.A. Andres, D. Refoyo, C. Garcia, A. Fernadez, R. Lopez, OpenNN: Open Neural Network Library, 2019 accessed April 2, 2019), https://www.opennn.net/.