Computers in Biology and Medicine 145 (2022) 105388

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Predicting glucose level with an adapted branch predictor

Tomas Koutny [a,*], Michael Mayo [b]

[a] Department of Computer Science and Engineering, University of West Bohemia, Univerzitni 8, 301 00, Plzen, Czech Republic
[b] Department of Computer Science, University of Waikato, New Zealand

ARTICLE INFO

Keywords:
Blood glucose level
Time series forecasting
Deep learning
Pattern learning
Computational costs
Interpretable models

ABSTRACT

Background and objective: Diabetes mellitus manifests as prolonged elevated blood glucose levels resulting from impaired insulin production. Such high glucose levels over a long period of time damage multiple internal organs. To mitigate this condition, researchers and engineers have developed the closed loop artificial pancreas consisting of a continuous glucose monitor and an insulin pump connected via a microcontroller or smartphone. A problem, however, is how to accurately predict short term future glucose levels in order to exert efficient glucose-level control. Much work in the literature focuses on least prediction error as a key metric and therefore pursues complex prediction methods such a deep learning. Such an approach neglects other important and significant design issues such as method complexity (impacting interpretability and safety), hardware requirements for low-power devices such as the insulin pump, the required amount of input data for training (potentially rendering the method infeasible for new patients), and the fact that very small improvements in accuracy may not have significant clinical benefit.

Methods: We propose a novel low-complexity, explainable blood glucose prediction method derived from the Intel P6 branch predictor algorithm. We use Meta-Differential Evolution to determine predictor parameters on training data splits of the benchmark datasets we use. A comparison is made between our new algorithm and a state-of-the-art deep-learning method for blood glucose level prediction.

Results: To evaluate the new method, the Blood Glucose Level Prediction Challenge benchmark dataset is utilised. On the official test data split after training, the state-of-the-art deep learning method predicted glucose levels 30 min ahead of current time with 96.3% of predicted glucose levels having relative error less than 30% (which is equivalent to the safe zone of the Surveillance Error Grid). Our simpler, interpretable approach prolonged the prediction horizon by another 5 min with 95.8% of predicted glucose levels of all patients having relative error less than 30%.

Conclusions: When considering predictive performance as assessed using the Blood Glucose Level Prediction Challenge benchmark dataset and Surveillance Error Grid metrics, we found that the new algorithm delivered comparable predictive accuracy performance, while operating only on the glucose-level signal with considerably less computational complexity.

1. Introduction

Glucose is a fuel used by cells for energy production. However, in order to utilize glucose, cells need insulin. Insufficient insulin production, therefore, leads to elevated blood glucose (BG) levels [1]. Diabetes Mellitus is a group of heterogeneous diseases, which manifest with elevated blood glucose level [2]. With Type 1 Diabetes Mellitus (T1D), the pancreas produces no or negligible traces of insulin. With Type 2 Diabetes Mellitus (T2D), cells develop insulin resistance. Due to this resistance, cells need considerably increased amounts of insulin, but the pancreas does not produce enough of it. In addition to other therapies such as multiple daily injections, both T1D and T2D can be treated with an external but bodily worn insulin pump [3,4] which delivers additional insulin to subcutaneous tissue from an external reservoir. The insulin pump may either independently deliver a specified basal rate of insulin according to a schedule specified by the patient or doctor (the so-called open loop artificial pancreas), or it may be linked wirelessly to a continuous glucose monitoring system (CGMS) and adapt the insulin pump rate in real time according to BG changes (the so-called closed loop version of artificial pancreas). Since the former approach may

* Corresponding author.
E-mail addresses: txkoutny@kiv.zcu.cz (T. Koutny), michael.mayo@waikato.ac.nz (M. Mayo).

https://doi.org/10.1016/j.compbiomed.2022.105388
Received 5 January 2022; Received in revised form 22 February 2022; Accepted 4 March 2022
Available online 19 March 2022

require additional insulin boluses at meal times along with manual basal rate adjustments depending on the circumstances (e.g. exercise), the closed loop approach is the more desirable approach of the two because of the potential for less patient/doctor intervention and monitoring, and much less need for manual insulin rate adjustments.

However, since insulin deficiency leads to elevated BG (possibly causing a hyperglycemic episode), closed loop treatments involving excessive amounts of insulin may clear too much glucose from the blood, increasing the risk of a hypoglycemic episode. Therefore, the insulin pump in the closed loop setting must deliver a precisely regulated amount of insulin to avoid both hypo- and hyperglycemia. Moreover, chronically elevated BG progressively damages the patient's organs. This creates a need for a sophisticated algorithm to precisely calculate the amount of insulin needed at any time. Moreover and adding to the challenge, insulin pump controllers are low-power devices with limited computing power and therefore algorithms that consume high amounts of energy to make calculations are not desirable.

To explain further how CGMS works, substances such as glucose and insulin diffuse across capillary walls from blood to interstitial fluid, which bathes cells of the subcutaneous tissue. CGMS measures glucose levels in the interstitial fluid of subcutaneous tissue, also known as interstitial glucose (IG). It does not measure the blood glucose directly. In particular, CGMS sensors have a tiny needle that is inserted into the subcutaneous tissue that the measures electric current produced by a chemical reaction with the IG. For technological reasons (such as the way the CGMS needle is designed) and additionally for physiological reasons, IG usually lags behind BG. Modern CGMSes therefore attempt to estimate current BG from the needle measurements in order to partially eliminate these physiological and technological lags. Nevertheless, the process is complicated as insulin moderates activity of GLUT-x glucose transporters with a non-linear diffusion rate of glucose gain and clearance from interstitial fluid [5–7]. Consequently, interstitial glucose is a delayed and distorted image of BG, making BG prediction a non-trivial task. Generally, we can assume a mean lag about 8 and 10 min for adolescents and adult patients respectively [8].

To clarify, in this paper BG denotes glucose levels measured in a blood sample. IG denotes a CGMS estimate of BG which, as discussed above, can differ from true BG. In the practice, the T1D patient sporadically (up to 4 times a day [9]) draws a drop of blood to measure BG in order to calibrate the CGMS. The BG measurement is sporadic due to the associated pain and discomfort. Alternatively, the patient can use a factory calibrated sensor [10,11]. CGMS is minimally invasive and provides continuous readings. Therefore, IG is the suitable glucose-level time series, on which to base the prediction and against which to calculate the prediction error, using a specific error metric. Currently, the ISO 15197 and FDA 2014 standards specify the desired CGMS accuracy [12]. Specifically, one of the ISO 15197:2013 requirements is that 95% of the measurement errors do not exceed 15% relative error. Relative error is defined as absolute difference between estimated (e.g. predicted) and measured glucose level, divided by measured glucose level. Average relative error between 3.25% and 5.25% has a nearly 100% probability of satisfying this ISO 1597:2013 requirement [13].

When calculating a glucose-level prediction, the prediction horizon (PH) is the time distance of the predicted level since the current level. Longer PH suffers from the possibility of unanticipated events such as meal intake, patient-administered insulin boluses or increased physical activity. Therefore, the prediction error is directly proportional to PH. To compare different prediction methods, PH is usually 30 or 60 min in the literature [14].

To predict IG, a researcher can devise a model that either utilizes the knowledge of physiological effects (for example, a first order differential model with components corresponding to physiological compartments such as the liver etc), or alternatively make less/no use of glucose physiology and instead utilize statistical, regression and/or artificial intelligence (AI) methods. In such a case, the researcher treats the IG readings from a CGMS as a time series and assumes that the past is sufficiently similar to the future in order to make the predictions work. AI methods largely include neural networks (NN) [15–17] i) as such methods have proven the most accurate.

While direct NN based prediction from IG time series is an intriguing research problem, additional information such as the typical timing of meals, their carbohydrate content (CHO), and other information such as exercise times, can also be used; alternatively these additional physiological data can be used to correct the predictions of the AI model. This leads to a sophisticated and complex system with intriguing possibilities. In recent years, a BG-prediction challenge (BGLP) has run, with researchers competing to develop BG prediction models [14]. Each iteration of the competition attracted a substantial number of entries, most of which concerned deep learning for regression based approaches. Only a few entries focus on other ideas such as classification, such as [16]; or using alternative machine learning models.

In the challenge, many entries chose energy-intensive approaches as such as deep learning. The best BGLP 2020 method is a complex neural network [15]. While accurate, such complexity suggests that the model predictions cannot be reliably explained or understood. Explainable models, on the other hand, can be further validated using analytical methods, to increase the safety of the entire insulin-delivery system.

Diabetes healthcare professionals have already agreed on a threshold relative error, below which a BG prediction system is clinically accurate enough. Errors can be further quantified by clinical impact. Three such approaches are the Clarke Error Grid Analysis [18], Parkes error grid [19] and Surveillance Error Grids (SEG) [20]. SEG is the most recent one. Each of these error analysis approaches defines a 2D grid of predicted BG levels vs actual BG levels. The grid is then divided into different zones of prediction error, accordingly to their safety. For example, one type of prediction error may be insignificant enough that the clinical action resulting from it makes no change, regardless of what the relative prediction error is. On the other hand, prediction error may be so wrong that it leads to a harmful clinical treatment.

Utilising the SEG approach, 30% relative prediction error is known to satisfy its safe zone. The BGLP challenge uses mean relative error (MARD) and root mean square error (RMSE) to establish a ranking of the competing methods. Contrary to this approach, we trade the precision below the 30% SEG relative error for a significantly reduced computational complexity. In this paper, we prove feasibility of this approach.

Computational costs are important for real-time devices such as artificial pancreas [21]. Therefore, we adapted the Intel P6 branch predictor algorithm into a proposal of a predictive method that is explainable, has low computational costs and clinical accuracy comparable to the best BGLP 2020 method, and does not have the negative properties of the deep NN solutions to this problem. We call the resulting algorithm Pattern Prediction.

2. Methods

Patients with diabetes need BG to be kept within a desired range, to avoid hypoglycemic episodes and long periods of elevated BG, including hyperglycemic episodes. As a response to various activities such as meal ingestion, increased physical activity, insulin dosage, and internal events of the glucose metabolism, there are certain, recurring patterns of BG change. Due to the blood and interstitial glucose relationship, we can observe these recurring patterns in IG as well.

2.1. Pattern prediction

To detect ingested meals in the IG signal [22,23], classified recent IG movements as accelerating/decelerating/steady, increase/decrease or a constant pattern. In the IG signal, each pattern is defined to occur between zero IG derivatives of the first or second order.

To reduce the computational cost, we propose sampling the most recent 15 min of IG signals with three levels at 5 min intervals. The 5 min interval period matches frequently used CGMS sensor frequencies. We

derived the 15 min interval from empirical data as longer intervals did not improve the results with the proposed patterns.

As the IG signal is noisy due to various physiological effects, when the CGMS sensor needle measures the electrical current, CGMS smooths the measured signal [24]. Therefore, we do not apply any additional smoothing but consider two IG levels equal, if their absolute difference is not greater than 0.1 mmol/L.

Using boolean *acc* symbol as an indicator, let us denote a detected acceleration/deceleration as following:

$$acc = |IG_{t-10} - IG_{t-5}| < |IG_{t-5} - IG_{t-0}| \quad (1)$$

Let us now define the following further boolean indicators, which will simplify the pattern definitions that follow:

$$aeb = |IG_{t-10} - IG_{t-5}| <= 0.1 \quad (2)$$

$$alb = (IG_{t-10} < IG_{t-5}) \ \& \ !aeb \quad (3)$$

$$agb = (IG_{t-10} > IG_{t-5}) \ \& \ !aeb \quad (4)$$

$$bec = |IG_{t-5} - IG_{t}| <= 0.1 \quad (5)$$

$$blc = (IG_{t-5} < IG_{t}) \ \& \ !bec \quad (6)$$

$$bgc = (IG_{t-5} > IG_{t}) \ \& \ !bec \quad (7)$$

In this set of indicator definitions, *a*, *b* and *c* refer to a sequence of three IG sensor readings, while *e*, *l* and *g* refer a relationship between a pair of these readings. For example, *alb* defines an indicator that is true whenever the first sensor reading is less than the second sensor reading by more than a margin of 0.1 mmol/L.

Next, we define the patterns that we use. These patterns are similar to those in Ref. [22]; but we additionally define convex and concave patterns. Each pattern can be defined intuitively as a logical combination of the indicators defined in Equations (1)–(7), and the exact pattern definitions are given in Table 1. Examining this table, we can see that all of the patterns besides "Steady" can be grouped into corresponding pairs. For example, the pattern "Acceleration" (defined as an accelerating increase across all three time points) and the pattern "Deceleration" are inverses. Similarly, "Convex" (defined as a peak at time $t-5$) can be paired with "Concave" (defined as a trough at $t-5$). We also make subtle distinctions between IG movements in the same direction. For example, both the "Deceleration" and "Decrease" patterns indicate decreasing IG readings, but with different degrees of rapidity in both cases.

Fig. 1 illustrates different patterns and the principle as a particular combination of pattern and discretized IG identifies particular IG levels, from which the pattern-prediction method predicts future $IG_{t+\ PH}$.

Future $IG_{t+\ PH}$ follows IG_t with the PH delay. To predict it, we adopted the two-level adaptive branch-predictor algorithm described by Ref. [25]. Superscalar processors such as Intel P6 use this predictor to guess future program branching. In a computer program's machine code, there are conditional jumps and call instructions at specific

Table 1
Detected IG patterns.

Pattern	Condition	Lookup Ordinal Number
Deceleration	agb & bgc & acc	0
Decrease	agb & bgc & !acc	1
Convex	agb & !bgc	2
Steady-Decrease	aeb & bgc	3
Steady-Increase	aeb & blc	4
Concave	alb & !blc	5
Increase	alb & blc & !acc	6
Acceleration	alb & blc & acc	7
Steady, no significant change	None of the above conditions apply.	8

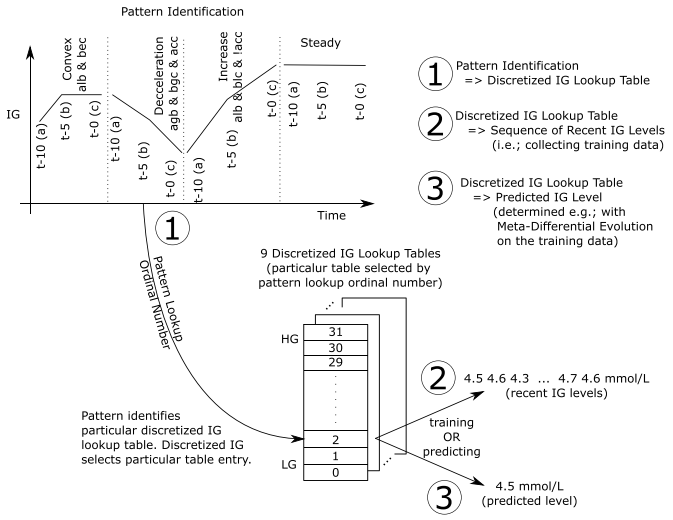

Fig. 1. Illustration of pattern identification to glucose-level prediction.

addresses. For such an address, the processor maintains a sequence of taken/not taken branch history. Such a sequence makes an integer of an arbitrary length. The processor then interprets this integer as an index into a pattern table. The pattern table comprises 2-bit saturating counters, which determine branch prediction per each pattern.

With a healthy subject, the body keeps BG within the desired range by applying various regulation mechanisms such as insulin and glucagon triggered reactions to the changing BG. With diabetes, either the patient or artificial pancreas regulates BG by following certain rules on insulin dosing. This regulation process has a strong resemblance to a computer code with jumps conditioned on specific BG levels and its changes. Therefore, we consider the proposed approach as feasible.

While a superscaler processor predicts a boolean value when executing a conditional computer code, we predict a floating-point number that represents glucose level. As this makes our task harder, we made the following considerations:

- As the instruction address, we used discretized IG_t.
- As the branching history pattern, we use the pattern lookup ordinal number as described in Table 1.
- To enable floating-point number prediction, we replaced each saturating counter with a single prediction - i.e.; parameters which we can determine using an external solver.

The IG discretization is necessary to limit computational and memory requirements of the pattern prediction method. Arbitrarily, let us consider a glycemic range between a low glucose level threshold of 3.0 mmol/L (LG) and a high glucose level threshold of 13.0 mmol/L (HG). Further, let us discretize the LG-HG range to 30 uniformly spaced slots. As a result, we effectively discretize the overall glucose-level range into 32 different slots - any glucose level equal or below LG, 30 glucose levels between LG and HG, and any glucose level equal or greater than HG.

Glucose levels below LG or greater than HG require immediate attention, which has to lead to a corrective action to avoid serious hypo- or hyperglycemia. In such cases, the entire system cannot be left unsupervised as it has already failed to keep the patient safe. Therefore, we limit the prediction to indicate levels between these two dangerous states only, because any autonomous function of the system is no longer trusted outside of these extreme levels.

The pattern prediction method has 9 patterns per 32 discretized IG levels. This makes $9 \times 32 = 288$ parameters, which we can determine with an external solver such as Meta-Differential Evolution (MetaDE) [26,27]. For the fitness function to be optimised by MetaDE, we used average relative error plus an unbiased estimation of standard deviation

of relative errors to quantify the difference between the measured and predicted glucose-level signal.

2.2. Experimental setup

We used the SmartCGMS framework [28] as the platform to implement the pattern-prediction method. SmartCGMS is a continuous glucose monitoring and controlling software framework. It is capable of reading various input signals such as IG, temperature, accelerometer, heart rate etc. Data can come from either real physical wearable devices or from simulated devices replaying previously recorded sensor readings. Stacks of further iterative processing are then applied to transform the input signals into other signals. To illustrate with a use case relevant to this work, SmartCGMS can periodically read the current IG signal to predict future IG, then use this prediction signal as an input to produce insulin pump control commands.

First, we executed the best BGLP 2020 method [15] to obtain state-of-the-art benchmark results, which could be viewed as an upper limit on IG predictive performance without regard to the interpretability or energy consumption of the model. We downloaded the original public implementation of this method by its author.

The BGLP 2020 dataset comprises data from years 2018 and 2020. We executed the state of the art baseline method on data of both years at specified lookaheads. This produced a series of timestamped glucose level predictions for the test data for each prediction horizon.

For each patient in the BLGP dataset, there are two files of multivariate time series data. Respectively, they are the training and testing data with the testing data occurring chronologically immediately after the training data. The approximate sizes of these data splits are 134,790 training examples and 31,671 testing examples, with CGM sensor readings every 5 min for most of the period covered by the data, though gaps in the time series do exist and must be accounted for [14,29]. further describe the BGLP dataset in detail. We created a SmartCGMS configuration that replayed the training glucose levels only and from this data constructed pattern prediction models for each patient. The models effectively store the most recent glucose levels for each pattern as discretized IG.

Next, we configured SmartCGMS to merge BGLP testing data with the predictions of the best BGLP2020 method. Then, we replayed the merged data to compute IG predictions with the pattern prediction method. To obtain minimum required prediction accuracy, we also shifted the IG signal in time by the respective PH. Eventually, this SmartCGMS configuration produced a single log, which contained timestamp, IG, time-shifted-IG and IG prediction by both methods - the BGLP2020 best method and the pattern prediction method. We repeated this procedure for PH from five to 60 min, stepping the PH up in increments of 5 min.

A key advantage of the BGLP competition is the availability of the public dataset. The BGLP organizers acquired the dataset with no prior consideration of any of the competing methods. This allows a fair comparison of different methods. With SmartCGMS, we strengthened this as SmartCGMS allows a strict isolation of the prediction method from the input data. With SmartCGMS, the method receives the input signals continuously without exposing the entire dataset at once. It behaves exactly as would a real CGMS sensor.

Eventually, we processed the log with all the predictions. In the IG signal, there are discontinuities caused by various reasons. For example, there could be signal loss or the CGMS-sensor needle may loose its grip. SmartCGMS ensures that all discontinuities are handled properly and in exactly the same manner for all predictive methods. Moreover, this particular SmartCGMS configuration ensures that all predicted levels less than LG and greater than HG are considered equal. As a result, we achieve a completely fair comparison of different methods.

3. Results

When recording predictions of the evaluated prediction methods, we produced a sorted sequence of tuples of the form (time stamp, IG, time shifted IG, BGLP2020 best method prediction, pattern-prediction method prediction). All the levels share exactly the same time stamp. Then, we calculated relative errors for each prediction method and sorted them by prediction method. Relative error as a metric was discussed previously, and the specific definition of relative error for our application is given by Equation (8).

$$relative\_error = \begin{cases} 0 & \text{if } IG_{measured} < LG \;\&\; IG_{predicted} < LG \\ 0 & \text{if } IG_{measured} > HG \;\&\; IG_{predicted} > HG \\ \dfrac{|IG_{measured} - IG_{predicted}|}{IG_{measured}} & \text{otherwise} \end{cases} \tag{8}$$

By sorting the predictions of the different methods by relative errors, we can then calculate the empirical cumulative distribution function of relative error (ECDF) for each prediction method [30]. In Fig. 2, we plotted percentiles of a particular relative error as a function of PH. In particular, we plotted the percentiles for the following relative errors: 10%, 15%, 20% and 30%. Examination of the figure shows that for relative errors under 30%, the deep learning approach has the best predictive performance, followed by the pattern prediction method, followed by the time shift method. As the relative error increases, however, the gap between the NN method and pattern prediction narrows. By the 30% relative error mark, pattern prediction outperforms the deep learning approach with clinically useful predictions being made approximately 38 min in advance in contrast to the deep learning approach with useful predictions up to approximately 34 min in the future.

Fig. 3 gives average relative error, average relative error plus standard deviation of relative errors and relative error at 95% percentile. These markers relate to the ISO 15197 and FDA 2014 standards, as discussed in the first section. This plot tells a similar story to that shown in Fig. 2: pattern prediction and the best BGPL competition method are close in predictive performance at short PHs but diverge after the 30 min PH mark (with the exception of the average metric), and both methods consistently outperform timeshift.

Tables 2 and 3 break PH and 30% relative-error percentile down to the individual patients. Table 2 gives the percentage of predictions with relative error less than 30% for 25, 30 and 35-min PHs. Table 3 gives the maximum possible PH, for which at least 95% of predicted IG has

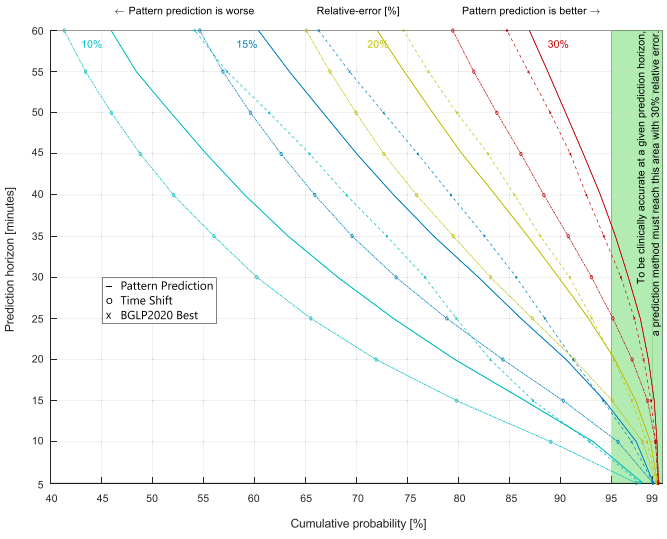

Fig. 2. Sliced ECDFs of the particular prediction signals.

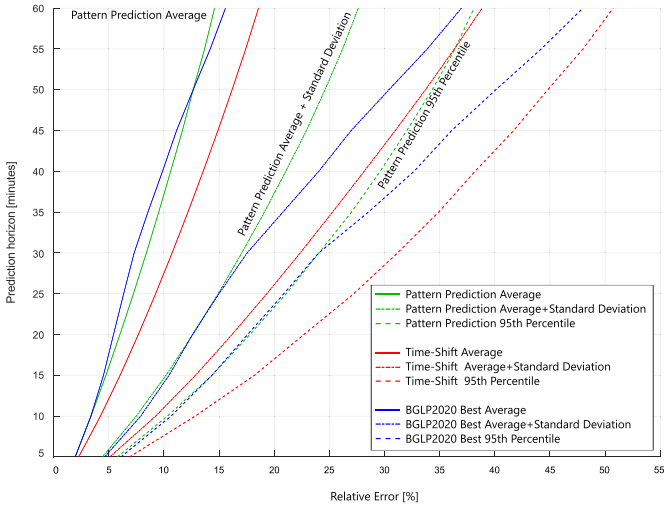

Fig. 3. Relative-error markers.

Table 2
Cumulative probabilities [%] of 30% relative error with 25, 30 and 35-min prediction horizons.
(greater is better).

Patient ID	Time-Shift	Pattern Prediction	BGLP Best
BGLP 2018			
559	96–95 – 93	98–97 – 96	98–96 – 94
563	98–97 – 96	99–98 – 98	99–98 – 98
570	99–99 – 98	100–99 – 99	99–99 – 98
575	93–90 – 88	97–95 – 94	96–95 – 93
588	98–96 – 95	99–98 – 98	99–98 – 98
591	93–90 – 87	97–95 – 93	96–95 – 93
BGLP 2020			
540	91–87 – 83	97–95 – 92	96–94 – 90
544	98–96 – 94	99–98 – 97	97–96 – 94
552	96–93 – 90	99–98 – 97	99–98 – 96
567	93–90 – 88	98–97 – 96	98–95 – 93
584	96–95 – 93	98–97 – 96	98–97 – 95
596	96–94 – 92	98–98 – 96	97–96 – 95
Average	96–93 – 91	98–97 – 96	98–96 – 95
Minimum	91–87 – 83	97–95 – 92	96–94 – 90
Count of ≥95	8–6 – 3	12–12 – 10	12–11 – 6

Table 3
Maximum prediction horizon [minute] with relative error less or equal to 30% for at least 95% of predicted levels (greater is better).

Patient ID	Time-Shift	Pattern Prediction	BGLP Best
BGLP 2018			
559	30	35	30
563	35	50	50
570	45	60	45
575	20	30	30
588	35	45	45
591	20	30	30
BGLP 2020			
540	20	30	25
544	30	40	30
552	25	40	40
567	20	35	30
584	30	35	35
596	25	40	35
20 PH Count	4	0	0
25 PH Count	2	0	1
≥30 PH Count	6	12	11

relative error less or equal to 30%. The tables show that pattern prediction on average is equivalent to the NN method at the 25 min PH, but exceeds the NN method at the 30 and 35 min PHs.

Fig. 4 explains the source of prediction errors for a single patient, comparing the pattern prediction algorithm's prediction to the best BGLP 2020 method's predictions to the actual IG readings. This demonstrates the method-design issues of trading the accuracy for low computational complexity. From this figure, we can see that the NN approach tracks the IG levels reasonably well, but occasionally wrongly anticipates rises in BG levels where none occur. The pattern prediction method, on the other hand, tends to oscillate its predictions more, most likely due to the effect of the glucose level discretization. Both methods however correctly signal transitions of IG across the high glucose level threshold.

4. Discussion

Our results (shown in Figs. 2 and 3) indicate that the BGLP2020 best method outperforms the pattern-prediction method for the 10%, 15% and 20% relative errors. With 30% relative error, the pattern-prediction method is better than BGLP2020 best method for all PHs with the respect to SEG. A consensus of healthcare professionals, reflected in SEG, defined 30% relative error as clinically accurate enough.

Interestingly, time-shifted IG produced prediction results that could often be used reliably up to 25-min PH. This confirms the necessity to include time-shifted IG as a baseline method giving a minimal required predictive performance when evaluating any IG prediction method.

The pattern prediction method exhibits a certain oscillation due to the discretization effect, as mentioned above (see Fig. 4). Fig. 5 depicts the possibility of post-processing the prediction signal, particularly with weighted moving average. Suitable post-processing is a subject to future research, including error rate grid analysis.

To summarise, we successfully reused an existing body of a cross-domain knowledge on branch prediction for IG prediction. We adopted a relatively early approach to branch prediction from the Intel P6 processor due to its simplicity. Over time, more effective predictors appeared for processors. However, not all of the newer branch predictors can be adopted due to the taken/not taken branch nature of the problem. Nevertheless, some of the ideas in the more advanced branch predictors could possibly be used to improve the prediction accuracy of our pattern prediction algorithm further, such as indirect-branch predictors.

Furthermore, the pattern prediction method uses the IG signal only. Thus, it is resilient to human behavior such as unannounced meals and wrongly calculated or misreported insulin boluses that can often occur when patients are required to frequently enter information.

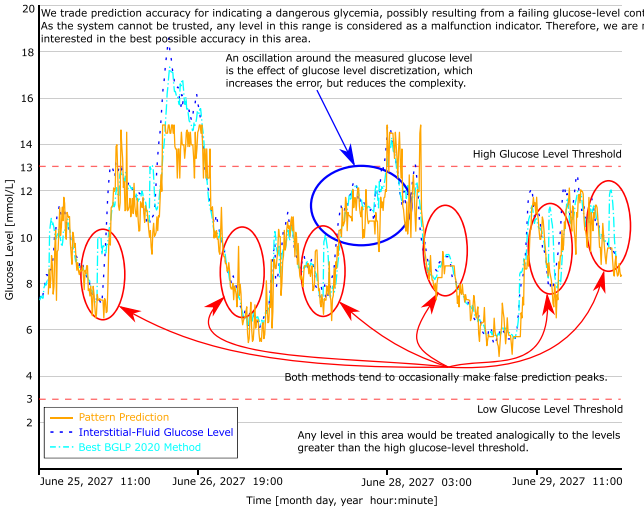

Fig. 4. Subject 544, 30-min prediction horizon.

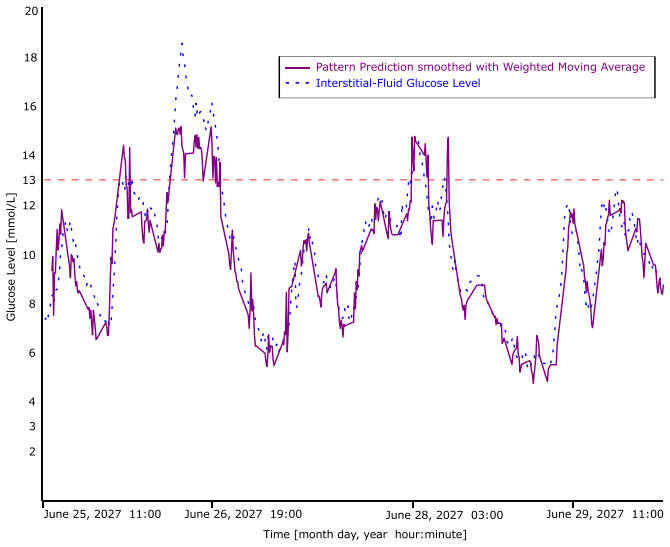

Fig. 5. Subject 544, 30-min Prediction Horizon with smoothed prediction.

When comparing the methods in the terms of hardware requirements, the BGLP2020 best NN has total of 55,524 parameters. The pattern prediction method has 288 parameters plus an additional 10 wt used for smoothing Weighted Moving Average signal. This makes 298 parameters, which is approximately a half percent of the quantity of floating-point parameters required by the NN. Additionally, in the BGLP Challenge dataset, there are a total of only 31,671 test examples. Given the number of parameters in the NN, there is a high risk that a NN could easily be overfit to the data. On the other hand, we are sure that our pattern prediction model is not overfit due to the order of magnitude less parameters.

Finally, when making predictions and due to the high number of parameters, the NN requires many floating-point operations to predict a single glucose level. The proposed pattern prediction is merely a two-level look-up table, and the weighted moving average would be the most complex computation done.

Due to the difficult interpretability of long-short term memory NNs, Cappon et al. adopt Shapley Additive exPlanations (SHAP) [31] to interpret deep learning model predictions. Using SHAP, they were able to identify impact of each input feature on model output obtained. Nevertheless, they approach is still NN and therefore the pattern-prediction approach is considerably less computationally intensive. In our opinion, the pattern-prediction approach is also intuitive to healthcare professionals and patients.

5. Limitations

One of the practical issues are the missing glucose levels due to sensor dropouts (for example, when the sensor battery runs out). One possible approach is to incorporate prior regularization in the model parameters identifications, which can handle the missing levels [32]. In our approach, SmartCGMS automatically detects discontinuities due to sensor gaps and applies the proposed model only when there is a contiguous time series of sensor reading inputs available. This obviates the need to impute missing values, and therefore we do not need to be concerned with the effectiveness of different imputation methods. Methods with a larger input window would have a more pressing need to handle missing values, however. Since the pattern prediction model requires only three past values, the prediction restarts after 15 min since the last discontinuity. Such a short period should not lead to an unexpected, rapid change of glucose level unless the patient is already in a critical state, when the patient should already be under a medical supervision anyway as its insulin-pump controller cannot be no longer trusted.

A period of CGMS sensor dropout is dynamic. It ranges from minutes to hours. With smaller periods, it is possible to reconstruct the missing glucose levels [26]. With larger periods, the reconstruction reliability decreases. By letting SmartCGMS to handle the discontinuities as described, it leads to a less complex program code. Otherwise, we would have to implement a specific decision-making based on the period length, thus increasing the program code complexity.

6. Conclusion

To conclude, we have demonstrated that a simple, explainable method can produce acceptable clinical accuracy. As insulin pump controllers are low power devices with reduced computational power, we argue that we need to pay more attention to the practical design issues once we reach the acceptable clinical accuracy. Such issues include input data size, energy required for model inference, memory and computing power requirements, and explainable predictions. Such characteristics enable a static analysis of the model that in turn enables patient's safety by design. NNs, on the other hand, are less safe because the only issue usually addressed is accuracy.

In the near future, we expect an emerging need for a future miniaturization of medical devices. While such a miniaturization would improve patient satisfaction with easy hiding of the CGM sensor and insulin pump, it would also require new algorithms that would consume less power. Our approach points towards this direction.

Declaration of competing interest

None declared.

Acknowledgement

This work was partially supported by the institutional long-term strategic development of the University of West Bohemia. It was published with the financial support of the European Union, as part of the project entitled Development of capacities and environment for boosting the international, intersectoral and interdisciplinary cooperation at the University of West Bohemia, project reg. No. CZ.02.2.69/0.0/0.0/18_054/0014627.

References

[1] L. Poretsky, Principles of Diabetes Mellitus, vol. 21, Springer, 2010.
[2] J.L. Jameson, A.S. Fauci, D.L. Kasper, S.L. Hauser, D.L. Longo, J. Loscalzo, Harrison's Principles of Internal Medicine, twentieth ed., vols. 1 & 2, McGraw-Hill Education/Medical, 2018.
[3] L. van den Boom, B. Karges, M. Auzanneau, B. Rami-Merhar, E. Lilienthal, S. von Sengbusch, N. Datz, C. Schröder, T. Kapellen, M. Laimer, et al., Temporal trends and contemporary use of insulin pump therapy and glucose monitoring among children, adolescents, and adults with type 1 diabetes between 1995 and 2017, Diabetes Care 42 (2019) 2050–2056.
[4] Z. Landau, I. Raz, J. Wainstein, Y. Bar-Dayan, A. Cahn, The role of insulin pump therapy for type 2 diabetes mellitus, Diabetes.Metabol.Res. Rev. 33 (2017), e2822.
[5] J.E. Hall, Guyton and Hall Textbook of Medical Physiology, 13e (Guyton Physiology), Saunders, 2015.
[6] P. Schrangl, F. Reiterer, L. Heinemann, G. Freckmann, L. Del Re, Limits to the evaluation of the accuracy of continuous glucose monitoring systems by clinical trials, Biosensors (Basel) 8 (2018).
[7] T. Koutny, Blood glucose level reconstruction as a function of transcapillary glucose transport, Comput. Biol. Med. 53 (2014) 171–178.
[8] M. Sinha, K.M. McKeon, S. Parker, L.G. Goergen, H. Zheng, F.H. El-Khatib, S. J. Russell, A comparison of time delay in three continuous glucose monitors for adolescents and adults, J Diabetes Sci. Technol. 11 (2017) 1132–1137.
[9] A.E. Minder, D. Albrecht, J. Schafer, H. Zulewski, Frequency of blood glucose testing in well educated patients with diabetes mellitus type 1: how often is enough? Diabetes Res. Clin. Pract. 101 (2013) 57–61.
[10] S. Alva, T. Bailey, R. Brazg, E.S. Budiman, K. Castorino, M.P. Christiansen, G. Forlenza, M. Kipnes, D.R. Liljenquist, H. Liu, Accuracy of a 14-day factory-calibrated continuous glucose monitoring system with advanced algorithm in pediatric and adult population with diabetes, J Diabetes Sci. Technol 16 (1) (2020), https://doi.org/10.1177/1932296820958754.

[11] R.P. Wadwa, L.M. Laffel, V.N. Shah, S.K. Garg, Accuracy of a factory-calibrated, real-time continuous glucose monitoring system during 10 days of use in youth and adults with diabetes, Diabetes Technol. Therapeut. 20 (2018) 395–402.
[12] G. Freckmann, S. Pleus, M. Grady, S. Setford, B. Levy, Measures of accuracy for continuous glucose monitoring and blood glucose monitoring devices, J Diabetes Sci. Technol 13 (2019) 575–583.
[13] S. Pardo, D.A. Simmons, The quantitative relationship between ISO 15197 accuracy criteria and mean absolute relative difference (MARD) in the evaluation of analytical performance of self-monitoring of blood glucose (SMBG) systems, J Diabetes Sci. Technol. 10 (2016) 1182–1187.
[14] C. Marling, R.C. Bunescu, The OhioT1DM dataset for blood glucose level prediction, in: KHD@ IJCAI, 2018, pp. 60–63.
[15] J. Freiburghaus, A. Rizzotti-Kaddouri, F. Albertetti, A deep learning approach for blood glucose prediction and monitoring of type 1 diabetes patients, in: 5th International Workshop on Knowledge Discovery in Healthcare Data, KDH), 2020.
[16] M. Mayo, T. Koutny, Neural multi-class classification approach to blood glucose level forecasting with prediction uncertainty visualisation, in: 5th International Workshop on Knowledge Discovery in Healthcare Data, KDH), 2020.
[17] Cappon, G., Meneghetti, L., Prendin, F., Pavan, J., Sparacino, G., Del Favero, S., Facchinetti, A., . A Personalized and Interpretable Deep Learning Based Approach to Predict Blood Glucose Concentration in Type 1 Diabetes .
[18] W.L. Clarke, The original clarke error grid analysis (ega), Diabetes Technol. Therapeut. 7 (2005) 776–779.
[19] J.L. Parkes, S.L. Slatin, S. Pardo, B.H. Ginsberg, A new consensus error grid to evaluate the clinical significance of inaccuracies in the measurement of blood glucose, Diabetes Care 23 (2000) 1143–1148.
[20] D.C. Klonoff, C. Lias, R. Vigersky, W. Clarke, J.L. Parkes, D.B. Sacks, M.S. Kirkman, B. Kovatchev, E.G. Panel, The surveillance error grid, J Diabetes Sci. Technol 8 (2014) 658–672.
[21] X. Yu, M. Rashid, J. Feng, N. Hobbs, I. Hajizadeh, S. Samadi, M. Sevil, C. Lazaro, Z. Maloney, E. Littlejohn, et al., Online glucose prediction using computationally efficient sparse kernel filtering algorithms in type-1 diabetes, IEEE Trans. Control Syst. Technol. 28 (2018) 3–15.
[22] S. Samadi, K. Turksoy, I. Hajizadeh, J. Feng, M. Sevil, A. Cinar, Meal detection and carbohydrate estimation using continuous glucose sensor data, IEEE.J.Biomed. Health Inf. 21 (2017) 619–627.
[23] S. Samadi, M. Rashid, K. Turksoy, J. Feng, I. Hajizadeh, N. Hobbs, C. Lazaro, M. Sevil, E. Littlejohn, A. Cinar, Automatic detection and estimation of unannounced meals for multivariable artificial pancreas system, Diabetes Technol. Therapeut. 20 (2018) 235–246.
[24] W. Gao, G.A. Brooks, D.C. Klonoff, Wearable physiological systems and technologies for metabolic monitoring, J. Appl. Physiol. 124 (2018) 548–556.
[25] L. Gwennap, New algorithm improves branch prediction, Microprocess. Rep. 9 (1995) 17–21.
[26] T. Koutny, Glucose-level interpolation for determining glucose distribution delay, in: XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013, Springer, 2014, pp. 1229–1232.
[27] T. Koutny, A. Della Cioppa, I. De Falco, E. Tarantino, U. Scafuri, M. Krcma, De–randomized meta-differential evolution for calculating and predicting glucose levels, in: 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), IEEE, 2019, pp. 269–274.
[28] T. Koutny, M. Ubl, Smartcgms as a testbed for a blood-glucose level prediction and/or control challenge with (an fda-accepted) diabetic patient simulation, Procedia Comput. Sci. 177 (2020) 354–362.
[29] C. Marling, R. Bunescu, The Ohiot1dm Dataset for Blood Glucose Level Prediction: Update 2020, KHD@ IJCAI, 2020.
[30] R.M. Dudley, in: Uniform Central Limit Theorems (Cambridge Studies in Advanced Mathematics), 2 ed., Cambridge University Press, 2014.
[31] S. Lundberg, S.I. Lee, A Unified Approach to Interpreting Model Predictions, 2017 arXiv preprint arXiv:1705.07874.
[32] X. Sun, M. Rashid, N. Hobbs, M.R. Askari, R. Brandt, A. Shahidehpour, A. Cinar, Prior informed regularization of recursively updated latent-variables-based models with missing observations, Control Eng. Pract. 116 (2021), 104933.