Blood Glucose Level Prediction using Physiological Models and Support Vector Regression

Razvan Bunescu, Nigel Struble, Cindy Marling
School of Electrical Engineering and Computer Science
Russ College of Engineering and Technology
Ohio University, Athens, Ohio 45701, USA
Email: bunescu,ns362007,marling@ohio.edu

Jay Shubrook, Frank Schwartz
The Diabetes Institute
Heritage College of Osteopathic Medicine
Ohio University, Athens, Ohio 45701, USA
Email: shubrook,schwartf@ohio.edu

Abstract—Patients with diabetes must continually monitor their blood glucose levels and adjust insulin doses, striving to keep blood glucose levels as close to normal as possible. Blood glucose levels that deviate from the normal range can lead to serious short-term and long-term complications. An automatic prediction model that warned people of imminent changes in their blood glucose levels would enable them to take preventive action. Modeling inter-patient differences and the combined effects of insulin and life events on blood glucose have been particularly challenging in the design of accurate blood glucose forecasting systems. In this paper, we describe a solution that uses a generic physiological model of blood glucose dynamics to generate informative features for a Support Vector Regression model that is trained on patient specific data. Experimental results show that the new prediction model outperforms all three diabetes experts involved in the study, thus demonstrating the utility of using the generic physiological features in machine learning models that are individually trained for every patient.

Keywords—regression, time series forecasting, diabetes

I. INTRODUCTION AND MOTIVATION

Type 1 diabetes (T1D) is a chronic disease, which cannot be prevented or cured, but which must be managed over a lifetime. In this disease, the pancreas does not produce the insulin needed to maintain normal blood glucose (BG) levels, and so patients depend upon exogenous supplies of insulin to live. In order to avoid serious diabetic complications, patients must continually monitor their blood glucose levels and adjust insulin doses, striving to keep blood glucose levels as close to normal as possible. Blood glucose management is complicated by a wide variability among individual patients in terms of sensitivity to insulin, response to lifestyle factors, propensity for complications, and response to treatment. Furthermore, large volumes of blood glucose data that are automatically collected through sensors must be manually analyzed and interpreted by the physicians.

Blood glucose levels that are too high, or hyperglycemia, lead to longterm complications of diabetes. Extremely high levels can cause diabetic ketoacidosis, a serious condition leading to severe illness or death. Blood glucose levels that are too low, or hypoglycemia, lead to complications including weakness, confusion, dizziness, sweating, shaking, and, if not treated in time, seizures, coma, or death. It is important to note that patients do not always know when problems are impending and are frequently unaware of problems even once they occur.

Problems that occur when patients are asleep are especially dangerous. The ability to predict impending BG problems before they occur would enable preemptive intervention. This would not only improve overall BG control, but could greatly impact patient safety. For example, a sleeping patient could be awakened and advised to eat before becoming hypoglycemic.

An accurate prediction model that warned people up to an hour in advance of imminent changes in their blood glucose levels would allow plenty of time for them to take preventive action. Consequently, we have spent significant research effort on designing machine learning models aimed at predicting blood glucose levels 30 minutes and 60 minutes into the future. Since blood glucose measurements have a natural temporal ordering, we approach the task of predicting blood glucose as a time series forecasting problem. For training and evaluation, we capitalize on a database of approximately 1,400 days worth of clinical patient data, which was amassed over the course of three clinical research studies. The collected data consists of blood glucose measurements taken at five-minute intervals through a continuous glucose monitoring (CGM) system and the corresponding daily events that include insulin and life event data. Life event data is input by the patients using a smartphone interface and tracks factors known to impact blood glucose levels, including: the timing and composition of meals; the intensity, duration and competitive nature of exercise; work schedules; sleep patterns; stress; illness; alcohol; holidays and travel.

II. OUTLINE OF RESEARCH ON BG PREDICTION

In previous work (Section III), we have introduced a Support Vector Regression (SVR) model [1] that incorporates a set of features based on past BG behavior and daily event data. This approach was shown to outperform a simple baseline in which blood glucose levels were assumed to remain constant. In a subsequent effort to improve the utility of the BG time series, we experimented with Auto Regressive Integrated Moving Average (ARIMA) [2] models and discovered that the SVR model, which was using both blood glucose and daily event data, did not perform better than an ARIMA model, which was using only BG data. This result challenged our belief that incorporating daily event data should lead to an improvement in prediction performance. Two alternative hypotheses were considered at this stage: either 1) the daily event data was not sufficiently accurate to have an impact on the predictive performance; or 2) the way daily event data was used in our

initial SVR model was suboptimal. The first hypothesis lost support after an annotation exercise (Section IV) in which three diabetes experts were shown plots of past BG behavior superimposed with daily event data and asked to predict the BG levels 30 and 60 minutes into the future. The physicians, who had access to daily event data, outperformed the prediction models that were using only BG data. Furthermore, live feedback from the physicians during the annotation exercise contained numerous references to daily events, reinforcing our belief in their explanatory power. Consequently, we focused on finding better ways to extract informative features from the daily event data (Section V). We implemented existing equations that attempt to model the impact that insulin and carbohydrate consumption have on the dynamics of blood glucose. Due to high inter-patient variability, the resulting generic physiological model does not have a good predictive performance. However, when incorporated as features in the SVR model (Section VI), it leads to significant improvements in accuracy. In particular, we show that the new SVR model, when augmented with physiological features, outperforms our diabetes experts on both 30 minute and 60 minute prediction (Section VII).

III. PREVIOUS WORK

In [3], [4], we conducted a preliminary experimental evaluation in which an SVR model was trained to predict the BG levels of a T1D patient. An arbitrary pivot date was selected about one month into the experimental data. Then 7 days before the pivot date were used to create training data, while test data was created from the 3 days following the pivot date. Since BG measurements are recorded by CGM systems every 5 minutes, one day may contribute up to 288 training or testing examples. Two separate SVR models were trained and tested to predict the BG levels for 30 and 60 minutes into the future. Each training and testing example was represented using features computed from past BG levels, as well as bolus dosages, basal rates, meal carbohydrate amounts and intensity of exercise. When trained with a linear kernel and default parameter settings, the SVR models were shown to substantially outperform a simple baseline that used the present BG level as the predicted value. For 30 minutes prediction, the SVR model obtained a Root Mean Square Error (RMSE) of 18.0 which compares favorably with the baseline RMSE of 25.1. For 60 minutes prediction, the SVR model had an RMSE of 30.9, substantially lower than the baseline RMSE of 43.2.

In subsequent work [5], the SVR system was refined to use a more comprehensive set of features and then evaluated on a much larger dataset. The feature templates used therein are as follows:

1) The BG level of patient x at present time t_0.
2) Exponential moving averages over $k = 3, 6, 9, 12$ previous BG levels, with an exponential decrease coefficient $\lambda = 0.9$.
3) Rates of change in the BG level over $k = 3, 6, 9, 12$ previous observations, with an exponential smoothing factor $\lambda = 0.25$.
4) Bolus insulin totals for every 10 minute interval in the past 60 minutes.
5) Basal insulin totals for every 10 minute interval in the past 60 minutes.
6) How much the basal rate in each 10 minute interval deviates from the average basal rate in the past 60 minutes.
7) Carbohydrate amounts ingested in each of the 10 minute intervals from the past hour.
8) Amounts of time spent on exercise, at work, or sleeping in multiple intervals in the past 60 minutes.

Data from 10 different T1D patients was used for evaluation this time. For each patient, a pivot date was chosen such that it was always a Sunday about 1 month into the patient's study. The feature templates were tuned using a grid search with 2 weeks of training data and 1 week of development data prior to the pivot point. This gave ten unique feature vectors tailored for each individual at each prediction horizon. The next 14 days after this pivot date were used for testing, while the 14 days prior to the pivot date were used for training an SVR model with a Gaussian kernel. The pivot date itself was used for tuning the model parameters. The SVR system was then compared against two baselines: the simple t_0 baseline that uses the present BG level as the prediction, and an ARIMA model trained on 4 days of CGM data – an exploratory data analysis had shown that 4 days gave the lowest RMSE for the ARIMA model. Model identification for ARIMA was completed with the R statistical function *auto.arima*, which uses the Bayes information criterion to determine the orders p and q of the autoregressive and moving average components, and the Phillips-Perron unit root test for determining the order d of the difference component. The RMSE results are shown in Table I, in which SVR_1 stands for the SVR system trained with all the features (including ARIMA), whereas SVR_2 stands for the SVR system trained without insulin or life event data.

TABLE I.   PREVIOUS BG PREDICTION RESULTS [5].

Horizon	t_0	ARIMA	SVR_1	SVR_2
30 min	19.5	4.5	4.7	4.5
60 min	35.5	17.9	17.7	17.4

Table I shows an unexpected result: the incorporation of daily event data in the SVR model (SVR_1) does not necessarily lead to improvements over an SVR model that uses only CGM data (SVR_2) or the simpler ARIMA model. One possible cause of this behavior was thought to be the suboptimal use of daily event data in our feature templates. Consequently, we explored the utility of using physiological models of glucose dynamics for the extraction of informative features from daily event data such as insulin and carbohydrates, as described in Sections V and VI later in this paper.

IV. PHYSICIAN PERFORMANCE DATA

In order to better understand the role of daily events in the dynamics of blood glucose levels, we asked three diabetes experts to label an evaluation dataset with their 30 and 60 minute predictions. There are a total of 200 timestamps in the dataset, collected from 5 T1D patients, 40 points per patient. The 200 evaluation points were manually selected to reflect a diverse set of situations: different times of day or night; close to or far from each of the possible types of daily events; on rising, decreasing, or flat BG curves; close to or far from local minima or maxima of the BG curve – where the local
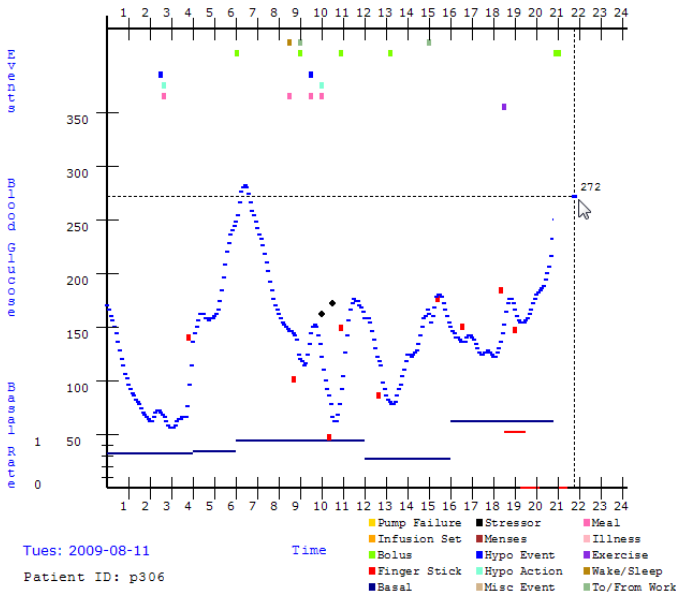
Fig. 1. The core frame of the annotation GUI with a sample CGM graph.

TABLE II. PHYSICIAN PREDICTION RMSE VS. BASELINES.

Horizon	t_0	ARIMA	$Phys_1$	$Phys_2$	$Phys_3$
30 min	27.5	22.9	**19.8**	21.2	34.1
60 min	43.8	42.2	**38.4**	40.0	47.0

results were obtained on CGM data that was smoothed using regularized cubic spline interpolation, whereas the results in Table II and henceforth are computed on raw CGM data. Most important, however, is the comparison between the doctors and ARIMA. Our highest scoring doctors, who use both the CGM data and daily events to perform prediction, outperform the ARIMA model that is using only CGM data. Regarding the third doctor's performance, we noticed during the annotation sessions that the doctor was fairly accurate in guessing the trend of the BG levels, but typically overestimated the increase or decrease in BG levels. To quantify this behavior, we defined a ternary classification task with the following three classes:

1) Same (S): the future BG value is within a threshold τ of the current value.
2) Lower (L): the future BG value is more than τ lower than the current value.
3) Higher (H): the future BG value is more than τ higher than the current value.

We used a threshold $\tau = 5mg/dl$ for 30 minute prediction and $\tau = 10mg/dl$ for 60 minute prediction. We then evaluated performance using a symmetric cost matrix C with a zero diagonal in which $C(L,S) = 1$, $C(H,S) = 1$, and $C(S,H) = 2$. The total costs for each doctor over the same evaluation dataset are listed in Table III. While the first physician still obtains the best results (lowest costs), the third physician now has results on the 60 minute task that are better than the baselines and the second physician.

TABLE III. PHYSICIAN PREDICTION COST VS. BASELINES.

Horizon	t_0	ARIMA	$Phys_1$	$Phys_2$	$Phys_3$
30 min	159	107	**81**	86	115
60 min	151	131	**106**	119	113

Additional support for the utility of daily event data came from the live feedback that the doctors generated during the annotation sessions. In making their predictions, the doctors would regularly refer to the presence of daily events and their properties, such as: the timing of meal events, the amount of carbohydrates and meal composition, the frequency of the bolus events and their type and dosage. This further reinforced our belief that daily event data is important in BG prediction, motivating us to look for better ways of using them to extract features for the automatic prediction models.

optima were either in the past or in the future; or in the vicinity of inflection points. The annotation exercise was performed using an in-house graphical user interface (GUI) through which the doctors could see only the patient data up to the present time t_0. The doctors were able to use the interface to navigate to any day in the past in order to make generalizations about BG level behavior. Before each annotation exercise, the doctors were trained to use the annotation user interface on a separate development dataset of ten points. For each point t_0 in the dataset, the doctors made two predictions: one for 30 minutes, one for 60 minutes, in this order. After making the two predictions for any given point, the system displays the real BG behavior, so that the doctors could further fine tune their prediction strategy. The annotation of the dataset was done separately with each of the three doctors. Figure 1 shows part of the user interface displaying the CGM data of a patient up to time $t_0 = 20:40$. The vertical dotted line corresponds to 60 minutes into the future, for which the doctor is supposed to make a prediction. Whenever the doctor clicks on the graph, a horizontal line is displayed, together with the corresponding BG level. The doctor can use the mouse to perform multiple exploratory clicks. The annotation is final only after pressing a "next" button. The same graph also shows two black diamond shaped points at about 10 and 10:30, corresponding to predictions made by the same doctor for $t_0 = 9:30$. Daily event data is superimposed in the upper part of the frame. Detailed information about each event can be displayed by clicking on the corresponding point.

The RMSE results of the three physicians are summarized in Table II, together with the ARIMA and t_0 baseline performance on the same dataset of 200 timestamps. Note that the t_0 and ARIMA baseline RMSE values are significantly larger than the previous results reported in Table I. This was expected, as the new evaluation dataset, which contains many points close to local optima, is much more difficult than the previous dataset where most of the test points lie on quasi-linear segments of the BG graph. Furthermore, previous

V. PHYSIOLOGICAL MODEL

Physiological models try to capture the dynamics of glucose relevant variables within different systems in the body. For example, equations have been introduced in the literature for tracking the carbohydrate intake as it is converted to blood glucose which then interacts with the kidneys, liver, muscles, and other body systems. Most physiological models characterize
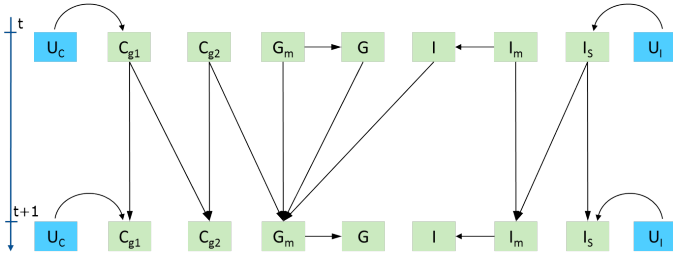
Fig. 2. Dependencies between variables in the physiological model.

the overall dynamics into three compartments: meal absorption dynamics, insulin dynamics, and glucose dynamics [6], [7], [8]. Since they are based on the same data, the equations used in the literature to model the underlying processes are almost identical [8], [9]. For our physiological model, we used these equations based on the description in Duke's PhD thesis [9], with a few adaptations in order to better match published data and feedback from our doctors.

A physiological model of glucose dynamics can be seen as a continuous dynamic model that is described by its state variables X, input variables U, and a state transition function that computes the next state given the current state and input variables i.e. $X_{t+1} = f(X_t, U_t)$. The vector of state variables X is organized according to the three compartments as follows:

1) Meal Absorption Dynamics:
   - $C_{g_1}(t)$ = carbohydrate consumption (g).
   - $C_{g_2}(t)$ = carbohydrate digestion (g).
2) Insulin Dynamics:
   - $I_S(t)$ = subcutaneous insulin (μU).
   - $I_m(t)$ = insulin mass (μU).
   - $I(t)$ = level of active plasma insulin (μU/ml).
3) Glucose Dynamics:
   - $G_m(t)$ = blood glucose mass (mg).
   - $G(t)$ = blood glucose concentration (mg/dl).

The vector of input variables U contains the carbohydrate intake $U_C(t)$, measured in grams (g), and the amount of rapid acting insulin $U_I(t)$, measured in insulin units (U). The value $U_I(t)$ at any time step t is computed automatically from the bolus events and the basal rate information. The state transition function captures dependencies among variables in the model, as illustrated in Figure 2.

The state transition equations are parametrized with a set of parameters α and are listed below for each compartment. For the meal absorption compartment, the equations are:

- $C_{g_1}(t+1) = C_{g_1}(t) - \alpha_{1C} * C_{g_1}(t) + U_C(t)$ [consumption]
- $C_{g_2}(t+1) = C_{g_2}(t) + \alpha_{1C} * C_{g_1}(t) - \alpha_{2C}/(1 + 25/C_{g_2}(t))$ [digestion]

The equations for the insulin compartment are:

- $I_S(t+1) = I_S(t) - \alpha_{fi} * I_S(t) + U_I(t)$ [injection]
- $I_m(t+1) = I_m(t) + \alpha_{fi} * I_S(t) - \alpha_{ci} * I_m(t)$ [absorption]

The general equation for the glucose compartment is $G_m(t+1) = G_m(t) + \Delta_{abs} - \Delta_{ind} - \Delta_{dep} - \Delta_{clr} + \Delta_{egp}$, where:

- $\Delta_{abs} = \alpha_{3C} * \alpha_{2C}/(1 + 25/C_{g_2}(t))$ [absorption]
- $\Delta_{ind} = \alpha_{1ind} * \sqrt{G(t)}$ [insulin independent utilization]
- $\Delta_{dep} = \alpha_{1dep} * I(t) * (G(t) + \alpha_{2dep})$ [insulin dependent utilization]
- $\Delta_{clr} = \alpha_{1clr} * (G(t) - 115)$ [renal clearance, only when $G(t) > 115$]
- $\Delta_{egp} = \alpha_{2egp} * exp(-I(t)/\alpha_{3egp}) - \alpha_{1egp} * G(t)$ [endogenous liver production]

Finally, the glucose and insulin concentrations depend deterministically on their mass equivalents as follows, where bm refers to the body mass and IS refers to the insulin sensitivity:

- $G(t) = G_m(t)/(2.2 * bm)$
- $I(t) = I_m(t) * IS/(142 * bm)$

In order to account for the noise inherent in the CGM data and the input variables, the state transition equations were used in an extended Kalman filter (EKF) model [10]. Since the CGM data comes every 5 minutes, the Kalman filter ran a state prediction step every 1 minute and a correction step every 5 minutes. The parameters α used in the physiological model were tuned to match published behavior and further refined based on feedback from the doctors, who were shown plots of the dependencies between variables in the model. This generic EKF model can be used on its own to make predictions about the BG level, however results from initial experiments on the reference dataset from Section IV were not better than the simple t_0 baseline. This was somewhat expected, given that patients with diabetes can vary significantly in how they respond to insulin or carbohydrate intake. One possibility for improving performance is to use a grid search to tune the model parameters α and the insulin sensitivity in order to fit training data from any given patient, an approach followed by Duke in [9]. However, using a grid search on all the parameters is unfeasible; even tuning just a subset of parameters can still be prohibitive in terms of time complexity. Furthermore, incorporating other types of life events in the EKF model requires significant engineering effort and may lead to a further increase in the number of parameters. Given the drawbacks associated with making the EKF model useful on its own for prediction, we decided to use it instead to generate informative physiological features for our SVR model.

VI. SVR MODEL WITH PHYSIOLOGICAL FEATURES

The state vector computed by the physiological model is $X(t) = [C_{g_1}(t), C_{g_2}(t), I_S(t), I_m(t), I(t), G_m(t), G(t)]$. In order to create features for the SVR model, the extended Kalman filter was first run up to the training/test point t_0, with a correction step every 5 minutes, including a correction at t_0. This resulted in a state vector $X(t_0)$. The EKF model was then run in prediction mode for 60 more minutes, and the state vectors at 30 minutes $X(t_0 + 30)$ and 60 minutes $X(t_0 + 60)$ were selected for feature generation. The actual physiological features were as follows: all the state variables in the vectors $X(t_0 + 30)$, $X(t_0 + 60)$, and the difference vectors $X(t_0 + 30) - X(t_0)$, $X(t_0 + 60) - X(t_0)$. The $4 * 7 = 28$ physiological features were augmented with a set of 12 $delta_i$ features that were meant to encode information about the trend of the CGM plot in the hour before the test point t_0.

Each $delta_i$ feature was computed as the difference between the BG level at time t_0 and the BG level at i time steps in the past, i.e., $delta_i = BG(t_0) - BG(t_0 - 5i)$. Furthermore, whenever ARIMA was used to generate features, it was trained on the 4 days before t_0 and then used to forecast the 12 values at 5 minute intervals in the hour after t_0. These values were used as additional features in one version of the SVR system.

For a given test point t_0 in the dataset, the SVR systems were trained on the week of data preceding t_0, using a Gaussian kernel. The width γ of the kernel, the width ϵ of the ϵ-insensitive tube, and the capacity parameter C were tuned using grid search on the data preceding the training week. This training and tuning scheme was problematic, however, for points in the evaluation dataset that were early in the history of a patient, and thus had insufficient training and/or tuning data. Since the evaluation dataset contained data from 5 patients, we used the data from a 6th patient to create a set of generic parameters to use in these early cases, as follows: a set of 50 points was selected from the 6th patient history; for each parameter combination in the grid search, SVR models were trained for all 50 points using the week before as training data, and performance was averaged across the 50 test points; the parameter setting that resulted in the best average performance was then selected as the generic set of parameters. We used these parameters whenever one of the test points in the evaluation dataset did not have at least 1,000 CGM measurements before the training week. For the few early points in the patient history that had less than one week of training data, we trained only on the data that was available.

VII. Experimental Results

We used the evaluation dataset described in Section IV to compare the following BG prediction systems:

1) ARIMA and the simple t_0 baseline.
2) The SVR system that uses physiological features, as described in Section VI above. Two versions of this system were evaluated: with (SVRϕ+A) and without (SVRϕ) ARIMA features.
3) The previous SVR system (SVRπ+A) [5] that uses the CGM and daily event features described in Section III, as well as ARIMA features.

The previous SVR system (SVRπ) was tuned, trained, and tested using the same scenario described for SVRϕ. However, we noticed that it obtained better performance when it used the same generic parameters for all points in the dataset; therefore we report these better results for SVRπ. A possible reason for this behavior may be the fact that the generic parameters were tuned by averaging performance over 5 weeks of training on a different patient, as opposed to using only one week of training from the given patient. While the new SVRϕ, due to its better physiological features, obtained good performance using one week of training, the old SVRπ system, which used a poorer feature representation for daily events, was less resilient to overfitting when trained with the same amount of data. The RMSE results are summarized in Table IV, in which, for comparison purposes, we repeat the first three columns of results from Table II. The new SVR system that is trained with physiological features beats the baselines and the previous SVR results (all with $p < 0.01$ in a one-tailed T-test of

TABLE IV.   SVR PREDICTION RESULTS VS. BASELINES AND HIGHEST SCORING DOCTOR RESULTS.

Horizon	t_0	ARIMA	$Phys_1$
30 min	27.5	22.9	19.8
60 min	43.8	42.2	38.4

Horizon	SVRπ+A	SVRϕ	SVRϕ+A
30 min	22.2	19.6	**19.5**
60 min	41.3	36.1	**35.7**

statistical significance). Most importantly, it also outperforms the best predictions from our three diabetes experts, thus demonstrating the utility of physiological modeling for the engineering of features in machine learning models for blood glucose level prediction.

VIII. Related Work

Attempts to model blood glucose levels for the purpose of determining insulin dosage date back to the 1960s [11]. While no definitive model exists as of yet, it should be noted that early efforts were hindered by the lack of CGM data, which first became available in 1999. Before CGM data was available, blood glucose values were obtained from patient finger sticks, typically yielding only four to six blood glucose values per day. Most blood glucose models developed to date are mathematical formalisms of physiological processes. A well-known, freely available, model is AIDA[1] [12]. The authors note that the model is not powerful enough to simulate individual patients and restrict its use to education and demonstration. Another influential, but proprietary, mathematical model was developed at the University of Virginia. This model was developed for testing control algorithms for an artificial pancreas, which could someday supplant the diabetic patient's own deficient pancreatic function [13].

The desire to build an artificial pancreas is the impetus behind much current blood glucose modeling research. In brief, an artificial pancreas consists of three components: an insulin pump; a continuous glucose monitoring system; and a closed loop control algorithm to tie them together, so that insulin flow can be continuously adjusted to meet patient needs [14], [15]. Some control algorithms are of the traditional proportional-integral-derivative (PID) variety, in which case, no blood glucose modeling is used. Others, however, use model predictive control (MPC), which employs mathematical/physiological models, or neural networks [16]. Many of these models consider the effects of food, but as they are intended for use without patient intervention, they do not consider patient recorded life-events.

Closest to our work is the research reported in [9], in which Gaussian Process regression with a Gaussian kernel was shown to outperform patients and other published results on the task of predicting post-meal blood glucose levels. The Gaussian Process approach in [9] was focused on patient decisions at meal times, and was evaluated on predicting the blood glucose level two hours after a meal. Shorter prediction times of 15 or 45 minutes were modeled differently, using autoregressive models or physiological models that were extended to exploit

---

[1]http://www.2aida.net

exercise information. The results were mixed, showing that autoregressive models are better for near future predictions, while physiological models perform better at prediction times 45 minutes or more into the future. In our work, the BG prediction models are trained and evaluated on a much larger dataset to predict BG levels in a wide variety of situations, at any time of day. Most importantly, our best performing model was shown to make better predictions than all three diabetes experts involved in the study.

More recently, Jensen et al. [17] trained an SVM model for detecting hypoglycemic events based on seven features derived from CGM and insulin data. The model was tested on 17 examples extracted from patients whose hypoglycemic events were induced artificially with insulin injections. While their trained system is able to predict hypoglycemia 14 to 20 minutes before it happens, it needs to be evaluated on a larger dataset, with CGM data from patients with spontaneous hypoglycemic events. Similarly, Zecchin et al. [18] evaluate a CGM prediction system using data generated with a type 1 diabetes simulator, and show that the system can be used to significantly decrease the time spent in hypoglycemia.

IX. Conclusions and Future Work

We introduced an adaptive model for blood glucose prediction that uses past BG behavior from CGM data and daily events such as insulin boluses and meals. We have shown how to integrate state variables computed by a physiological model as features into an SVM model for regression. Experimental results on 30 and 60 minute prediction confirmed the utility of the physiological features: when incorporated into an existing SVR model, they led to a significant improvement in RMSE performance. Furthermore, the SVR model with physiological features made better predictions than each of the three diabetes experts involved in the study. The results demonstrate the significant utility of physiological modeling for the engineering of features in machine learning models for BG level prediction.

An interesting aspect of the results is that the adaptive BG prediction model, using only a subset of the daily events available to the diabetes experts, was able to outperform them. Based on the feedback obtained from the doctors during the annotation exercise (Section IV), we know that physicians use all types of events in their predictions. For example, the time of day was deemed to be important due to dawn phenomena [19] and other time related patterns. Therefore, we plan to engineer features that capture the impact of all relevant life events on the blood glucose dynamics. We are also investigating unobtrusive sensing solutions for physiological parameters that will be less demanding on the patient and consequently less prone to incorrect or missing data.

References

[1] A. J. Smola and B. Scholkopf, "A tutorial on support vector regression," NeuroCOLT2 Technical Report Series, Tech. Rep., 1998.

[2] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control, 4th Edition*. Wiley, 2008.

[3] C. Marling, M. Wiley, R. Bunescu, J. Shubrook, and F. Schwartz, "Emerging applications for intelligent diabetes management," in *Proceedings of the Twenty-Third Innovative Applications of Artificial Intelligence Conference (IAAI-11)*. San Francisco, CA, USA: AAAI Press, August 2011.

[4] C. Marling, M. Wiley, R. C. Bunescu, J. Shubrook, and F. Schwartz, "Emerging applications for intelligent diabetes management," *AI Magazine*, vol. 33, no. 2, pp. 67–78, 2012.

[5] M. Wiley, "Machine learning for diabetes decision support," Master's thesis, Ohio University, Athens, OH, August 2011.

[6] S. Andreassen, J. J. Benn, R. Hovorka, K. G. Olesen, and E. R. Carson, "A probabilistic approach to glucose prediction and insulin dose adjustment: Description of metabolic model and pilot evaluation study," *Computer Methods and Programs in Biomedicine*, vol. 41, pp. 153–165, 1994.

[7] E. Lehmann and T. Deutsch, "Compartmental models for glycaemic prediciton and decision-support in clinical diabetes care: promise and reality," *Computer Methods and Programs in Biomedicine*, vol. 56, pp. 133–204, 1998.

[8] T. Briegel and V. Tresp, "A nonlinear state space model for the blood glucose metabolism of a diabetic," *Automatisierungstechnik*, vol. 50, pp. 228–236, 2002.

[9] D. L. Duke, "Intelligent diabetes assistant: A telemedicine system for modeling and managing blood glucose," Ph.D. dissertation, Carnegie Mellon University, Pittsburgh, Pennsylvania, October 2009.

[10] D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. Wiley, 2006.

[11] A. Boutayeb and A. Chetouani, "A critical review of mathematical models and data used in diabetology," *BioMedical Engineering OnLine*, vol. 5, no. 43, 2006.

[12] E. Lehmann and T. Deutsch, "A physiological model of glucose-insulin interaction in type 1 diabetes mellitus," *Journal of Biomedical Engineering*, vol. 14, no. 3, pp. 235–242, 1992.

[13] B. P. Kovatchev, M. Breton, C. DallaMan, and C. Cobelli, "In silico preclinical trials: A proof of concept in closed-loop control of type 1 diabetes," *Journal of Diabetes Science and Technology*, vol. 3, no. 1, pp. 44–55, 2009.

[14] D. C. Klonoff, "The artificial pancreas: How sweet engineering will solve bitter problems," *Journal of Diabetes Science and Technology*, vol. 1, no. 1, pp. 72–81, January 2007.

[15] E. Dassau, C. Lowe, C. Barr, E. Atlas, and M. Phillip, "Closing the loop," *International Journal of Clinical Practice*, vol. 66, pp. 20–29, 2012.

[16] B. W. Bequette, "A critical assessment of algorithms and challenges in the development of a closed-loop artificial pancreas," *Diabetes Technology & Therapeutics*, vol. 7, no. 1, pp. 28–47, February 2005.

[17] M. H. Jensen, T. F. Christensen, L. Tarnow, E. Seto, M. D. Johansen, and O. K. Hejlesen, "Real-time hypoglycemia detection from continuous glucose monitoring data of subjects with type 1 diabetes," *Diabetes Technology and Therapeutics*, vol. 15, no. 7, 2013.

[18] C. Zecchin, A. Facchinetti, G. Sparacino, and C. Cobelli, "Reduction of number and duration of hypoglycemic events by glucose prediction methods: A proof-of-concept in silico study," *Diabetes Technology & Therapeutics*, vol. 15, no. 1, pp. 66–77, January 2013.

[19] M. I. Schmidt, A. Hadji-Georgopoulos, M. Rendell, S. Margolis, , and A. Kowarski, "The dawn phenomenon, an early morning glucose rise: implications for diabetic intraday blood glucose variation," *Diabetes Care*, vol. 4, no. 6, p. 579585, 1981.