A Machine Learning Approach to Predicting
Blood Glucose Levels for Diabetes Management

Kevin Plis, Razvan Bunescu
and **Cindy Marling**
School of Electrical Engineering and Computer Science
Russ College of Engineering and Technology
Ohio University, Athens, Ohio 45701, USA

Jay Shubrook and **Frank Schwartz**
The Diabetes Institute
Heritage College of Osteopathic Medicine
Ohio University, Athens, Ohio 45701, USA

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. 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. The new model outperforms diabetes experts at predicting blood glucose levels and could be used to anticipate almost a quarter of hypoglycemic events 30 minutes in advance. Although the corresponding precision is currently just 42%, most false alarms are in near-hypoglycemic regions and therefore patients responding to these hypoglycemia alerts would not be harmed by intervention.

Introduction and Motivation

Of the world's 347 million people with diabetes, from 5 to 10% have type 1 diabetes (T1D), the most severe kind. In T1D, the pancreas fails to produce insulin, an essential hormone needed to convert food into energy and to regulate blood glucose (BG) levels. T1D can not be prevented or cured, but it can be effectively treated with external supplies of insulin and managed through BG control. Good BG control helps to delay or prevent serious long-term complications of diabetes (Diabetes Control and Complications Trial Research Group 1993). Patients strive to avoid both hyperglycemia, or high BG levels, and hypoglycemia, or low BG levels. Hyperglycemia can lead to long-term complications including blindness, amputations, kidney failure, strokes, and heart attacks, while hypoglycemia can cause immediate symptoms of weakness, confusion, dizziness, sweating, shaking, and, if not treated in time, seizures, coma or death.

Achieving and maintaining good BG control is a difficult task. T1D patients must monitor their BG levels throughout the day and take corrective action whenever they are hyper or hypoglycemic. Being able to accurately *predict* impending hyper or hypoglycemia would give patients time to intervene and prevent these BG excursions, improving overall health, safety, and quality of life. Such predictions would enable or facilitate numerous potential applications of benefit to T1D patients, including: (a) alerts warning of immediately impending problems; (b) recommendations of interventions to prevent problems; (c) "what if" analysis to project the effects of different lifestyle choices and treatment options; and (d) enhanced individual BG profiles that could help physicians tailor treatment to the needs of each patient.

There has been a recent explosion of interest in BG prediction due to its role in closed loop control for the Artificial Pancreas Project (Juvenile Diabetes Research Foundation 2014). In brief, an artificial pancreas consists of three components: an insulin pump; a continuous BG monitoring system; and a closed loop control algorithm to tie them together, so that insulin flow can be continuously adjusted to meet patient needs (Klonoff 2007; Dassau et al. 2012). Although still a work in progress, Time Magazine named the artificial pancreas one of the 25 best inventions of 2013 (Time Magazine 2013). Several recent publications detail related work in BG prediction (Jensen et al. 2013; Zecchin et al. 2013; Wang et al. 2014); however, a limiting factor has been the use of small and/or simulated patient datasets.

Over the past ten years, we have worked on intelligent decision support systems for patients with T1D on insulin pump therapy (Marling et al. 2012), and we have amassed a database of over 1,600 days worth of clinical patient data. We are now capitalizing on this data to build machine learning models aimed at predicting BG levels 30 and 60 minutes into the future. Since BG measurements have a natural temporal ordering, we approach the task of predicting BG levels as a time series forecasting problem. 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 impact BG levels, such as insulin, meals, exercise and sleep. In this paper, we describe our machine learning approach and report on our latest experiments, which focus on the safety-critical problem of predicting hypoglycemia. The structure of the paper is as follows: we first introduce a small dataset on which we assess the prediction performance of diabetes experts; we then describe a physiological model that is used

to generate informative features for a support vector regressor; in the following experimental results section we evaluate the performance of the new model, both on the general blood glucose level prediction and on the more focused task of hypoglycemic event prediction; the paper ends with conclusions and ideas for future work.

Blood Glucose Prediction Dataset

Using data collected from 5 T1D patients, we created an evaluation dataset of 200 timestamps, 40 points per patient, with the aim of measuring the prediction performance of expert physicians. The 200 evaluation points were manually selected to reflect the diverse set of situations the predictor would encounter in practice: 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 optima were either in the past or in the future; or in the vicinity of inflection points.

To better understand the role of daily events in the dynamics of blood glucose levels, we asked three diabetes experts to label this evaluation dataset with their 30 and 60 minute predictions. 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 the two predictions were made for any given point, the system displayed the real BG behavior, so that the doctors could further fine tune their prediction strategy. The dataset annotation was done separately with each of the three doctors.

The root mean square error (RMSE) results of the three physicians are summarized in Table 1, together with the performances of two baselines on the same dataset of 200 timestamps. The first baseline, t_0, simply assumes that BG does not change, i.e., stays at its t_0 level. The second is an Auto Regressive Integrated Moving Average (ARIMA) (Box, Jenkins, and Reinsel 2008) 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 best performing doctors outperform the simple t_0 baseline. 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

Table 1: 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

Table 2: Physician prediction cost vs. baselines.

Horizon	t_0	ARIMA	$Phys_1$	$Phys_2$	$Phys_3$
30 min	0.40	0.27	**0.20**	0.21	0.29
60 min	0.38	0.33	**0.27**	0.30	0.28

typically overestimated the magnitude of 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 $C(L, S) = 0.5$, $C(H, S) = 0.5$, and $C(L, H) = 1$. The average costs for each doctor over the same evaluation dataset are listed in Table 2. 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.

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 number 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 explore the utility of physiological models in the extraction of informative features for an adaptive regression model.

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 the overall dynamics into three compartments: meal absorption dynamics, insulin dynamics, and glucose dynamics (Andreassen et al. 1994; Lehmann and Deutsch 1998; Briegel and Tresp 2002). Since they are based on the same data, the equations used in the literature to model the
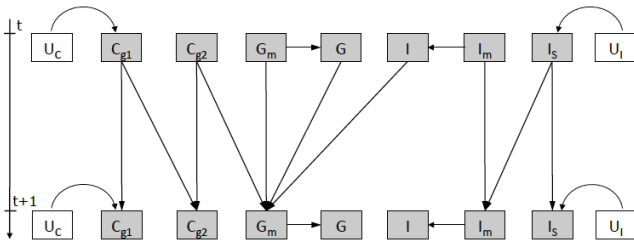
Figure 1: Variables and dependencies in the physio model.

underlying processes are almost identical (Briegel and Tresp 2002; Duke 2009). For our physiological model, we used these equations based on the description in Duke's PhD thesis (Duke 2009), 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 bolus and basal rate information. The state transition function captures dependencies among variables in the model, as illustrated in Figure 1.

The state transition equations are parameterized 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:

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

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. 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 (Simon 2006), as described in more detail in (Bunescu et al. 2013).

SVR Model with Physiological Features

The state vector computed by the physiological model is $X(t) = [C_{g_1}(t), C_{g_2}(t), I_S(t), I_m(t), I(t), G_m(t), G(t)]$. While the transition equations reflect generic blood glucose behavior, the actual parameters of these equations differ among patients. Instead of tuning these parameters directly for each patient, we opted for a simpler approach, using the state variables to create features for a Support Vector Regression (SVR) model (Smola and Scholkopf 1998) that was individualized for each patient, as follows. First, the extended Kalman filter was 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 Δ_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 Δ_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.

Table 3: SVR prediction results vs. baselines and highest scoring doctor results.

Horizon	t_0	ARIMA	$Phys_1$	$SVR\phi$	$SVR\phi$+A
30 min	27.5	22.9	19.8	19.6	**19.5**
60 min	43.8	42.2	38.4	36.1	**35.7**

Table 4: SVR vs. ARIMA performance on hypo prediction.

System	TPR	FPR	P	F_1
SVR	**23.0%**	0.8%	42.7%	**29.9%**
ARIMA	9.9%	**0.3%**	**44.1%**	16.2%

Table 5: SVR vs. ARIMA vs. t_0 performance.

	RMSE			Average Cost	
System	All	30m	60m	30m	60m
SVR	**24.4**	**22.6**	**35.8**	**0.27**	**0.27**
ARIMA	27.1	24.9	39.6	0.31	0.33
t_0	28.7	26.6	41.7	0.39	0.36

Experimental Results

We used the evaluation dataset previously described 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 above. Two versions of this system were evaluated: with ($SVR\phi$+A) and without ($SVR\phi$) ARIMA features.

The RMSE results are summarized in Table 3, in which, for comparison purposes, we repeat the first three columns of results from Table 1. The new SVR system that is trained with physiological features outperforms the t_0 and ARIMA baselines. 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. To determine whether the overall competitive RMSE performance can translate to practical utility, we conducted an evaluation of the SVR system on the more focused task of hypoglycemia prediction, as described in the next section.

Hypoglycemia Prediction

Hypoglycemia prediction is the most critical application of BG prediction from a clinical perspective. Untreated hypoglycemia is a safety hazard for patients, especially for sleeping patients, who are at risk for the "dead in bed" syndrome. Being able to accurately predict that BG will drop to under 70 mg/dl 30 minutes in advance would give patients time to intervene and prevent the drop.

The evaluation dataset described above contains 200 evaluation points, of which only 13 belong to hypoglycemic regions. In order to estimate the system's performance on predicting hypoglycemia, we constructed a much larger dataset of 5,816 test points as follows. First, consecutive CGM readings under 70 mg/dl were combined into hypo events. Events that lasted less than 20 minutes, i.e., had 3 or fewer sensor readings, were ignored. The remaining hypo events were checked for errors that might suggest recent sensor failure, such as missing data shortly before or after the event. For each hypo event, a prediction time t_0 was selected such that $t_0 + 30$ was the first point where blood glucose dropped below 70. This resulted in 152 hypo events and corresponding prediction times. A set of non-hypo points was then sampled automatically such that the overall evaluation dataset contains a mix of hypo and non-hypo points reflecting the true distribution. First, the ratio of hypo to non-hypo readings was calculated for each patient. Then, the corresponding number of non-hypo points for that patient was sampled

at random from the patient's CGM data. For example, if a patient had 64 hypo events and the ratio of non-hypo to hypo points was 25 to 1, then 1600 non-hypo points were selected at random from the patient's CGM history. Each sampled point was checked to make sure that the BG level did not fall under 70 mg/dl in the next hour. This procedure resulted in a new evaluation dataset of 5,816 test points.

During evaluation, each of the 5,816 test points served as the prediction time t_0. The SVR system was trained on data preceding time t_0, and predictions were made for future timestamps in 5 minute increments, from $t_0 + 5$ to $t_0 + 60$. If any one of the 12 predictions was less than 70 mg/dl, the system was considered to have predicted a hypo event. Otherwise, the system was considered to have predicted a non-hypo event.

Table 4 shows the SVR and ARIMA results on the task of predicting hypoglycemic events, in terms of sensitivity/recall (TPR), $1 -$ specificity (FPR), precision (P), and F-measure (F_1). By definition, the baseline t_0 would never predict hypoglycemia; therefore, it is not shown in this table. Table 5 compares the SVR system with the ARIMA and t_0 baselines in terms of their RMSE and average ternary classification costs at 30 and 60 minutes.

The results in Table 4 show that the SVR system is able to predict 23% of the hypoglycemic events with a false positive rate under 1%. All 47 false positives were for BG readings under 140 mg/dl, with 32 of them for near-hypoglycemic readings of 80 mg/dl or lower. Were the model to be used in practice, patients responding to hypoglycemia alerts at these BG levels would not be harmed by intervention. While we are still working to improve sensitivity, results to date provide proof of concept that the system could alert patients to impending dangerous situations.

Conclusion

We have described a machine learning approach to predicting blood glucose levels and presented the results of recent experiments in hypoglycemia prediction. An SVR model, informed by a physiological model and trained on patient specific data, has outperformed diabetes experts at predicting blood glucose levels and can predict 23% of hypoglycemic events 30 minutes in advance. We are working

to improve the overall accuracy of our model and its sensitivity to impending hypoglycemia, in part through incorporating additional daily events that impact blood glucose levels into the model. For example, sleep is known to impact blood glucose levels, but it is not factored into current physiological models. Our database includes patient-entered sleep records which we are now incorporating. Further, we are experimenting with newly available, low-cost, noninvasive and unobtrusive physiological sensors to aid in the automated identification of relevant daily events. Diabetes management presents many opportunities for clinicians and artificial intelligence researchers to collaborate on challenging research that could potentially improve health, safety and quality of life for millions of people living with diabetes.

Acknowledgments

References

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

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

Briegel, T., and Tresp, V. 2002. A nonlinear state space model for the blood glucose metabolism of a diabetic. *Automatisierungstechnik* 50:228–236.

Bunescu, R.; Struble, N.; Marling, C.; Shubrook, J.; and Schwartz, F. 2013. Blood glucose level prediction using physiological models and support vector regression. In *Proceedings of the IEEE 12th International Conference on Machine Learning and Applications (ICMLA)*. Miami, FL: IEEE.

Dassau, E.; Lowe, C.; Barr, C.; Atlas, E.; and Phillip, M. 2012. Closing the loop. *International Journal of Clinical Practice* 66:20–29.

Diabetes Control and Complications Trial Research Group. 1993. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *New England Journal of Medicine* 329(14):977–986.

Duke, D. L. 2009. *Intelligent Diabetes Assistant: A Telemedicine System for Modeling and Managing Blood Glucose*. Ph.D. Dissertation, Carnegie Mellon University, Pittsburgh, Pennsylvania.

Jensen, M. H.; Christensen, T. F.; Tarnow, L.; Seto, E.; Johansen, M. D.; and Hejlesen, O. K. 2013. Real-time hypoglycemia detection from continuous glucose monitoring data of subjects with type 1 diabetes. *Diabetes Technology and Therapeutics* 15(7).

Juvenile Diabetes Research Foundation. 2014. Artificial pancreas project research. http://jdrf.org/research/treat/artificial-pancreas-project/, accessed April, 2014.

Klonoff, D. C. 2007. The artificial pancreas: How sweet engineering will solve bitter problems. *Journal of Diabetes Science and Technology* 1(1):72–81.

Lehmann, E., and Deutsch, T. 1998. Compartmental models for glycaemic prediction and decision-support in clinical diabetes care: promise and reality. *Computer Methods and Programs in Biomedicine* 56:133–204.

Marling, C.; Wiley, M.; Bunescu, R. C.; Shubrook, J.; and Schwartz, F. 2012. Emerging applications for intelligent diabetes management. *AI Magazine* 33(2):67–78.

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

Smola, A. J., and Scholkopf, B. 1998. A tutorial on support vector regression. Technical report, NeuroCOLT2 Technical Report Series.

Time Magazine. 2013. The 25 best inventions of the year 2013. http://techland.time.com/2013/11/14/the-25-best-inventions-of-the-year-2013/slide/the-artificial-pancreas/, accessed April, 2014.

Wang, Q.; Molenaar, P.; Harsh, S.; Freeman, K.; Xie, J.; Gold, C.; Rovine, M.; and Ulbrecht, J. 2014. Personalized state-space modeling of glucose dynamics for type 1 diabetes using continuously monitored glucose, insulin dose, and meal intake: An extended kalman filter approach. *Journal of Diabetes Science and Technology* 8(2):331–345.

Zecchin, C.; Facchinetti, A.; Sparacino, G.; and Cobelli, C. 2013. Reduction of number and duration of hypoglycemic events by glucose prediction methods: A proof-of-concept in silico study. *Diabetes Technology & Therapeutics* 15(1):66–77.