Emerging Applications for Intelligent Diabetes Management

Cindy Marling, Matthew Wiley, Razvan Bunescu, Jay Shubrook, Frank Schwartz

■ *Diabetes management is a difficult task for patients, who must monitor and control their blood glucose levels in order to avoid serious diabetic complications. It is a difficult task for physicians, who must manually interpret large volumes of blood glucose data to tailor therapy to the needs of each patient. This paper describes three emerging applications that employ AI to ease this task: (1) case-based decision support for diabetes management; (2) machine-learning classification of blood glucose plots; and (3) support vector regression for blood glucose prediction. The first application provides decision support by detecting blood glucose control problems and recommending therapeutic adjustments to correct them. The second provides an automated screen for excessive glycemic variability. The third aims to build a hypoglycemia predictor that could alert patients to dangerously low blood glucose levels in time to take preventive action. All are products of the 4 Diabetes Support System project, which uses AI to promote the health and well-being of people with type 1 diabetes. These emerging applications could potentially benefit 20 million patients who are at risk for devastating complications, thereby improving quality of life and reducing health-care cost expenditures.*

> Managing diabetes is challenging for the physician and even more challenging for the patient. Very few diseases provide so much access to facts, yet remain so inexact. There are so many things that can affect blood glucose control.
>
> – Dr. Jay Shubrook

Type 1 diabetes (T1D) is an autoimmune disease in which the pancreas fails to produce insulin, an essential hormone needed to convert food into energy. It is more severe and less common than type 2 diabetes, but still affects approximately 20 million people worldwide. T1D is a chronic disease, which cannot be cured, but which must be treated and managed over time. At our diabetes center, more than 600 T1D patients are treated with insulin pump therapy. A mechanical pump infuses the patient with insulin, attempting to mimic and replace normal pancreatic function. The management goal is for the person with diabetes to achieve and maintain blood glucose (BG) levels close to those of a person without diabetes. It has been experimentally determined that good BG control can help delay or prevent serious long-term diabetic complications, including blindness, amputations, kidney failure, strokes, and heart attacks (Diabetes Control and Complications Trial Research Group 1993). Avoiding complications improves quality of life for patients, while reducing the financial burden of health-care cost expenditures.

Diabetes management is a challenging task for patients, who must monitor their BG levels and daily activities, and for physicians, who recommend therapeutic adjustments based on the monitoring data. (1) Task complexity stems from a wide variability among individual patients in terms of sensitivity to insulin, response to lifestyle factors, propensity for complications, adherence to physician recommendations, and response to treatment; and (2) voluminous BG data, which is automati-

Patients are routinely seen only three to four times a year by a diabetes specialist. Adjustments in insulin pump therapy occur infrequently between appointments for most T1D patients. In fact, there is data to suggest that they often do not get therapeutic advice at their appointments, either! Clinical inertia can result when there is too much information for a busy clinician to cognitively handle. An AI system that can analyze large volumes of data, make skilled clinical observations about the data, demonstrate information graphically for physician confirmation, and then suggest therapeutic solutions to identified problems could help."

– Dr. Frank Schwartz

cally collected through sensors, but which must be manually analyzed and interpreted.

The state of the art in commercially available software for T1D management is exemplified by the CareLink system.[1] Patients can upload the BG and insulin data stored in their pumps to a central site, where they and their physicians can review it. Data is displayed in logs and various graphical forms. No attempt is made to automatically interpret the data or to provide therapeutic advice. AI researchers have prototyped more proactive software tools (see, for example, Montani et al. [2003]; Duke et al. [2008]). Montani and colleagues' project combined decision support with telemedicine; it ended without a commercialization phase. Duke and colleagues present a promising machine-learning and telemedicine approach, which could yet have commercial potential.

AI Solutions to Diabetes Management Problems: Work in Progress

Since 2004, we have been conducting clinical research studies with T1D patients in order to develop and evaluate software tools for intelligent diabetes management. In the AI in medicine tradition, real-world medical problems have provided fertile ground for AI research, driving research directions in search of practical solutions. Three applications with potential clinical impact are emerging from our work: case-based decision support for diabetes management; machine-learning classification of BG plots; and support vector regression for BG prediction.

Case-based decision support was our initial focus. Clinicians faced with data overload sought ways to interpret BG data automatically. Rather than wading through pages of numbers and plots, they wanted to know (1) when patients had problems necessitating adjustments to therapy; and (2) which therapeutic adjustments would help individuals achieve and maintain good BG control. The case-based approach allowed us to capitalize on clinical experience in managing more than 600 T1D patients. Over the course of our work, new challenges arose that called for the integration of additional AI approaches and techniques. One such challenge was automatically detecting excessive glycemic variability. There is no standard way to characterize this potentially dangerous problem, and yet physicians know it when they see it in BG plots. Trying to capture this physician perception led to machine-learning classification of BG plots. Next, we aimed to anticipate problems that could be circumvented through immediate intervention. This led to work on support vector regression for BG prediction.

While these applications are still emerging, there are significant benefits to eventual deployment. Controlling diabetes helps individuals avoid devastating complications that reduce quality and length of life. These avoidable complications have a high financial cost, as well. In the United States alone, annual diabetes health-care expenditures total 174 billion dollars, and 32 percent of the entire Medicare budget is diabetes related (Juvenile Diabetes Research Foundation 2011). This work contributes insight into how AI technology can help through prototyping and evaluating promising solutions and clarifying issues standing in the way of deployment.

Case-Based Decision Support

The 4 Diabetes Support System (4DSS) aims to (1) automatically detect problems in BG control; (2) propose solutions to detected problems; and (3) remember which solutions are effective or ineffective for individual patients. It can assist busy clinicians managing multiple T1D patients, and it could be embedded in insulin pumps or smartphones to provide low-risk advice to patients in real time. We selected case-based reasoning (CBR) as the initial approach because (1) it provides support for tailored solutions based on similarity to known cases; (2) diabetes management guidelines are general in nature, requiring personalization; (3) a wide range of both physical and lifestyle factors influence BG levels; and (4) CBR has been successfully applied to managing other chronic medical conditions (Bichindaritz 2008).

The first step in developing 4DSS was to build a case base as a central knowledge repository. Although abundant BG data was initially available, usable cases were not. This is because the life events coinciding with BG levels, used by physicians to determine appropriate therapy, were not routinely recorded. To acquire contextualized cases for the system, we conducted a clinical research study involving 20 T1D patients. Each patient participated for six weeks, manually entering daily

Problem: Nocturnal hypoglycemia. BG levels are dangerously low all night. The patient reports feeling "totally out of it" when she wakes up. She does not eat anything to correct the hypoglycemia until noon. She had not eaten a bedtime snack the night before.

Solution: The patient should always have a mixed-nutrient snack before bed. She should lower her overnight basal rate. The combination of more food and less insulin will prevent overnight lows.

Outcome: The patient reports eating mixed nuts and crackers before bed. She sets the basal rate in her pump as advised. BG data for subsequent weeks shows that the problem is resolved.

Figure 1. A Sample Case.

BG, insulin, and life-event data into an experimental database through a web-based interface. Physicians reviewed the data, detecting BG control problems and recommending therapeutic adjustments. Patients implemented the recommended adjustments (or not), and physicians reviewed subsequent data to evaluate the clinical outcomes, in an iterative cycle.

Problems, solutions, and outcomes were structured into cases and stored in the case base. We were able to acquire 50 cases over the course of the clinical research study. Figure 1 shows the high-level textual description of a sample case from 4DSS. Internally, a case is represented as an object of a hierarchical Java class containing approximately 150 data fields. Figure 2 shows part of the internal representation of the sample case displayed in figure 1. A more detailed description of the case representation is presented in Marling, Shubrook, and Schwartz (2009).

The sample case records an actual problem of nocturnal hypoglycemia. Hypoglycemia, or low BG, leads to weakness, confusion, dizziness, sweating, shaking, and, if not treated in time, seizures, coma, or death. Hyperglycemia, or high BG, contributes to long-term diabetic complications. Extremely high BG levels can cause diabetic ketoacidosis, a serious condition leading to severe illness or death. It is important to note that patients do not know when problems are impending and are frequently unaware of problems even once they occur. Problems that occur when patients are asleep, as in the sample case, are especially dangerous.

Typically in CBR systems, reasoning begins with a known problem that can be readily described and elaborated. Solving a given problem entails finding and adapting the most similar, or most useful, case in the case base. In this domain, problems are not usually given, or known a priori, but must be detected in continuous patient data. Our approach was to model automated problem-detection routines on physician problem-detection strategies. We implemented rule-based routines to detect 12 common BG control problems identified by physicians. We built a 4DSS prototype to (1) detect BG control problems in patient data; (2) display detected problems to the physician, who would select those of interest; (3) retrieve, for each selected problem, the most applicable case in the case

Problem

Description	Patient has nocturnal lows
Problem Type	Hypoglycemia
Detected As	Prewaking hypoglycemia
Evidence	BG data, sleep data
Event Details	
BG Value	50
BG Low Target	70
Symptoms	Totally out of it
Related to Time	Yes
Time Period	Bedtime to awakening
...	

Figure 2. Partial Internal Representation of the Sample Case.

AI could potentially provide the physician (and patient) with real-time decision support. It has the potential to become an electronic "coach," letting people have timely feedback that could potentially increase safety and efficacy of treatment.

– Dr. Jay Shubrook

base; and (4) display the retrieved case to the physician as decision support in determining appropriate therapy to correct the problem.

Evaluation and feedback were obtained through a patient exit survey and two structured sessions in which diabetes practitioners evaluated system capabilities. Patients indicated on the survey that they would willingly accept automated decision support, but noted that the time required for data entry was a deterrent. Panels of clinicians were shown random samples of identified problems and retrieved cases. They agreed 90 percent of the time that the problem detections would be useful to physicians, 80 percent of the time that the cases retrieved were similar to the cases of interest, and 70 percent of the time that the solutions retrieved would benefit the current patients. Physicians also noted that the integration of BG, insulin, and life-event data helped them to identify BG trends more readily and adjust therapy more effectively. Conclusions were (1) the prototype provides proof of concept that intelligent decision support can assist in diabetes management; (2) additional problem/solution/outcome cases are needed to provide solutions for more BG control problems; and (3) data entry time demands on the patient must be reduced. Results of this study were reported in Schwartz, Shubrook, and Marling (2008) and Marling, Shubrook, and Schwartz (2008, 2009).

We conducted a second clinical research study, involving 26 T1D patients, to (1) reduce patient time demands and (2) reevaluate the 4DSS prototype. BG and insulin data stored in the patient's pump was uploaded to the experimental database rather than entered by the patient. Patients were asked for their typical daily schedules, and these were used to approximate actual life events. Patients were not required to supply continuous glucose monitoring (CGM) data, but it was uploaded for patients who normally used it as part of routine care. Data that could not be automatically transferred or approximated was omitted from consideration. During evaluation, approximately half as many problems were detected per patient per week as were detected in the first clinical study. This finding was statistically significant ($p = .017$), although there were no statistically significant differences between the two patient populations and no reason to suspect that patients were actually experiencing fewer problems. As detailed in (Schwartz et al. 2010), we attribute this impaired performance to the lack of requisite data.

An adverse event that occurred during this study highlights the potential for 4DSS to affect health and well-being. A participating patient experienced a problem in which his pump failed and stopped delivering insulin. He was aware that his BG was high, and he instructed the pump to deliver more insulin. However, he did not know that the pump was not functioning, and his BG continued to climb. He went into diabetic ketoacidosis (DKA) and was admitted to the hospital, where he experienced a (nonfatal) heart attack. When his data was scanned retroactively, the system automatically detected the pump problem eight hours before the patient was hospitalized. Had the system been running in real time, the patient might have been alerted to make a simple adjustment before experiencing DKA.

Conclusions from this study were (1) lack of life-event and CGM data impairs the ability to detect clinical problems; and (2) extending system capabilities to predict and prevent problems presents new research challenges and new opportunities to improve health outcomes. Results of this study were published in Schwartz et al. (2010).

Next, we conducted a third study with the goals of scaling up the system prototype and enhancing it to adapt past solutions to the needs of current patients. Twelve T1D patients completed a 3-month protocol in which they (1) uploaded insulin pump and CGM data weekly; and (2) supplied otherwise unavailable life-event data through a Web browser on a daily basis. During this study (1) six new problem-detection routines were developed; (2) 30 new cases were added to the case base; and (3) a case adaptation module was implemented. The new problem-detection routines expand the kinds of BG control problems that can be identified. The new cases codify additional knowledge, allowing more or better solutions to be found for identified problems. The case adaptation module customizes retrieved solutions for individual patients. For example, a past solution could be to lower the rate at which the pump continuously infuses insulin from 1.0 to 0.9 units per hour. If the current patient's rate is 0.5 units, the adaptation module recommends lowering it to 0.45 units, rather than simply displaying the old case and leaving adaptation to the user. While all system developments during this study were reviewed by participating physicians, formal evaluation awaits a future study. The current structure of the 4DSS is shown in figure 3.

Machine-Learning Classification of BG Plots

During 4DSS development, we encountered a type

of BG problem that we could not readily detect by encoding physician problem-detection strategies in rules. This was excessive glycemic variability, a bouncing back and forth between hypo- and hyperglycemia. Excessive glycemic variability is illustrated by figure 4a, while figure 4b illustrates acceptable glycemic variability. Glycemic variability is an active area of current diabetes research (Ceriello and Ihnat 2010; Kilpatrick, Rigby, and Atkins 2010). Excessive glycemic variability has been linked to hypoglycemia unawareness, an acutely dangerous condition, and to oxidative stress, which contributes to long-term diabetic complications (Monnier et al. 2006). Its successful detection would enable routine screening for the 20 million individuals with T1D, a valuable clinical application in and of itself.

There is no definitive metric for glycemic variability; nor is there any available tool to detect excessive glycemic variability (Rodbard 2009). Yet, diabetes specialists readily recognize excessive glycemic variability when they see it in blood glucose plots, like the one shown in figure 4a. This situation motivated the use of machine-learning (ML) classification to capture physician perception in an automated screen for excessive glycemic variability. Figure 5 shows the basic architecture of our approach to glycemic variability classification.

Two physicians (JS and FS) reviewed 400 BG plots and characterized each one as excessively variable or not, based on their gestalt perceptions. They were in agreement on 262 of the plots, 187 positive examples and 75 negative. A development set of 52 randomly chosen examples was set aside for feature selection and parameter tuning for three ML algorithms. The remaining data was used in a 10-fold cross validation setting to train and evaluate the ML algorithms for the task of detecting excessive glycemic variability.

The three learning algorithms that we investigated are naive Bayes (NB), support vector machines (SVMs), and multilayer perceptrons (MPs). A naive Bayes model implements a simple probabilistic Bayesian network in which the observed features of an example are considered to be independent given the example's class label. Support vector machines (Schölkopf and Smola 2002; Vapnik 1995) are dual learning algorithms that process examples only through computing their dot-products. These dot-products between feature vectors can be efficiently computed through a kernel function, without iterating over all the corresponding features. Given the kernel function, the SVM learner tries to find a hyperplane that separates positive from negative examples and at the same time maximizes the separation (margin) between them. A multilayer perceptron (Bishop 1995) is a feedforward neural network composed of multiple layers of nodes, in which every internal node implements a logistic activation function that produces an output value by combining weighted inputs from the nodes on the previous layer. The weights of the connections between nodes on consecutive layers constitute the model's parameters and are learned efficiently using the technique of error back propagation.

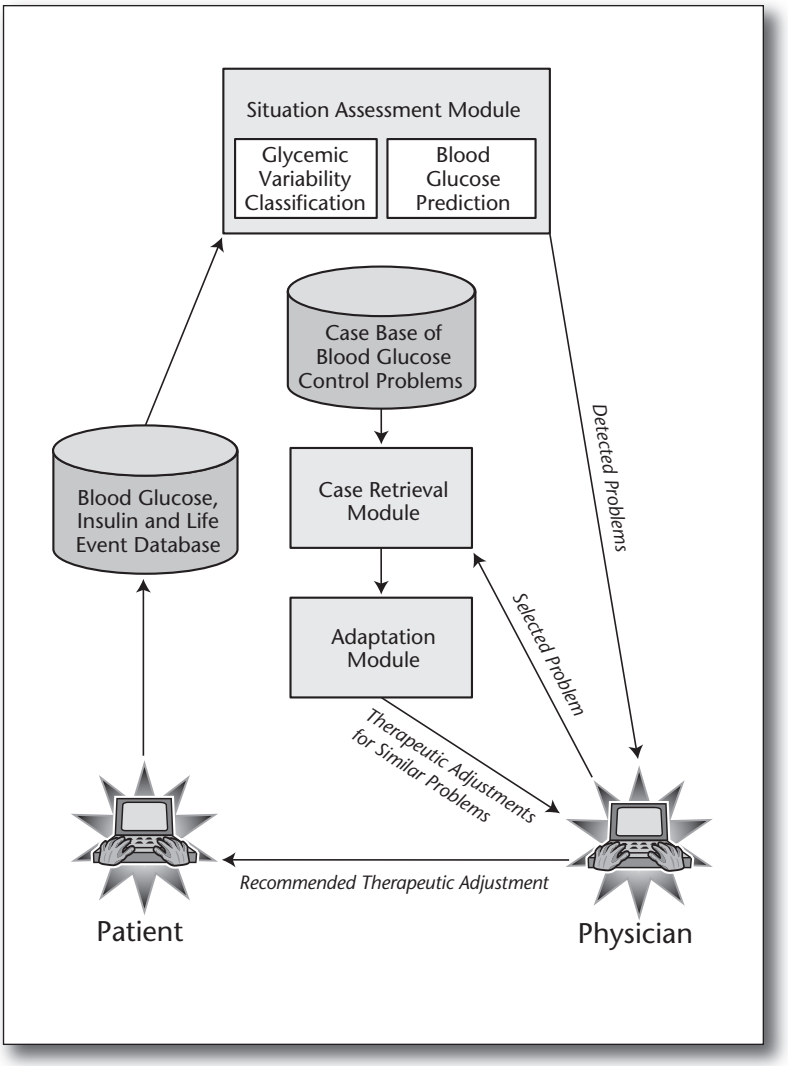

Figure 3. Overview of the 4 Diabetes Support System.

Noise Elimination

Since the CGM sensors do not record data with 100 percent accuracy, physicians implicitly smooth the original sequence of values recorded by the CGM sensor. Figure 6 shows an example plot of CGM data on which a physician has explicitly marked the actual graph used for making diagnostic decisions. We investigated a number of smoothing techniques such as simple moving averages, exponential moving averages, polynomial regression with L2 regularization, low pass discrete Fourier transform filter, and cubic spline smoothing

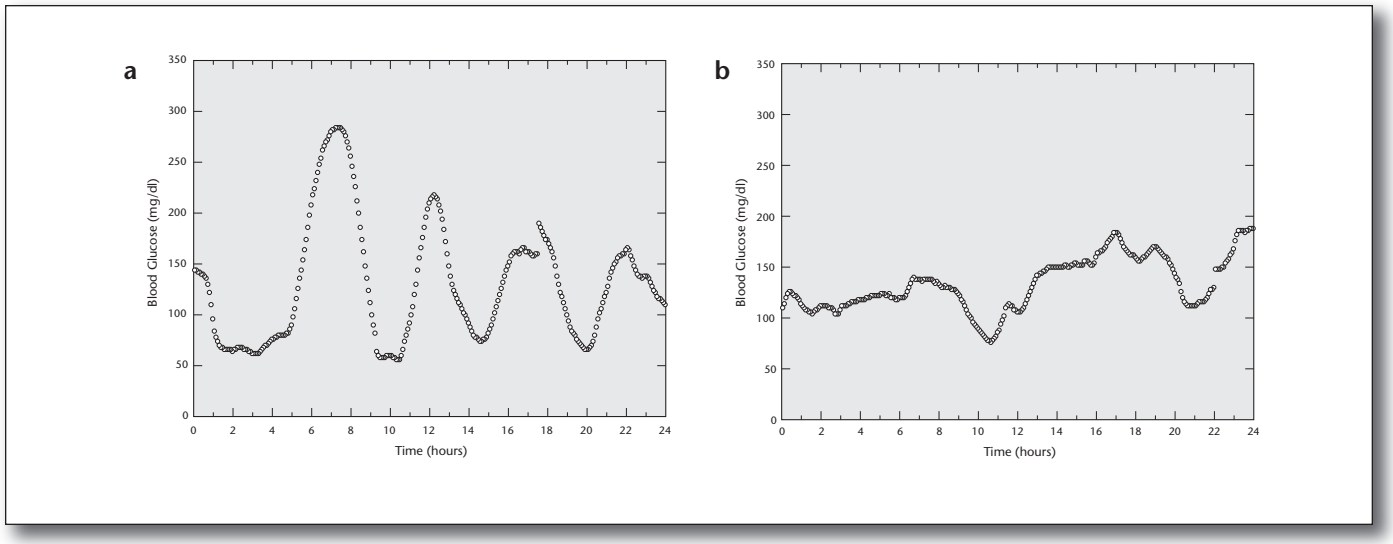

Figure 4. Blood Glucose Plots from CGM Data.

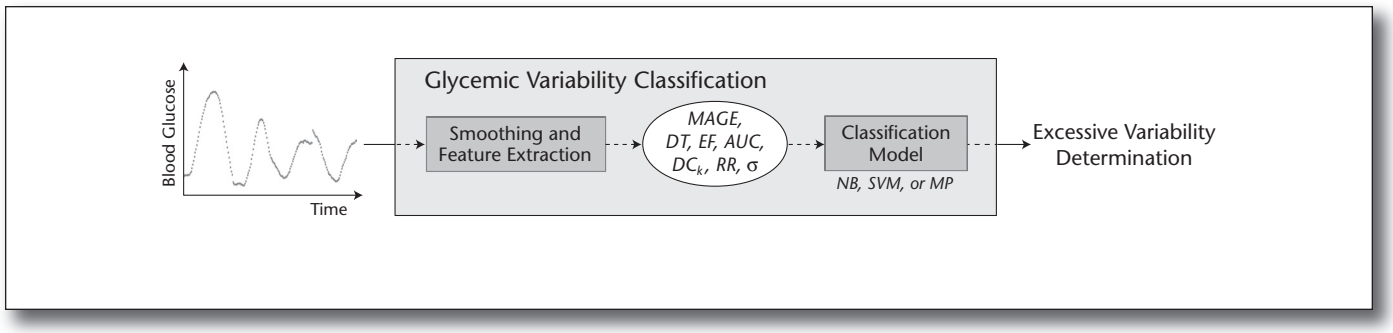

Figure 5. Glycemic Variability Classification Pipeline.

(Wiley et al. 2011). A modified version of regularized cubic spline smoothing that passes close to points of significant local optima was identified by the physicians to best correspond with their implicit smoothing process. Figure 7 shows the result of running this version of cubic spline smoothing on the same CGM graph marked by the physician in figure 6.

Excessive Glycemic Variability Detection Features

For the purpose of automatic classification, each CGM plot is represented as a vector of domain-dependent and domain-independent features that are automatically extracted from the corresponding sequence of BG levels. The domain-specific features, described in brief below and more fully in (Marling et al. 2011), are Mean Amplitude of Glycemic Excursion, Distance Traveled, and Excursion Frequency. We included domain-independent pattern-recognition features with the aim of modeling aspects of blood glucose variability not captured by the domain-dependent metrics (Wiley et al. 2011). The domain-independent features we investigated were standard deviation, area under the curve, roundness ratio, bending energy, eccentricity, direction codes, central image moments, and amplitudes of discrete Fourier transform frequencies. The features that were automatically selected by running feature selection algorithms on the development data set are summarized below.

Mean Amplitude of Glycemic Excursion (MAGE): This was the first glycemic variability metric (Service et al. 1970), and it remains the most respected (Rodbard 2009). MAGE calculates the mean distance between the local minima and maxima of a blood glucose plot. Only distances exceeding the standard deviation of the blood glucose values are included in the aggregate.

Distance Traveled (DT): Distance traveled is the sum of the distances between each pair of consecutive data points; it captures overall fluctuation.

Excursion Frequency (EF): This feature counts the number of blood glucose excursions between the local minima and maxima. Only distances greater than 75 mg/dl are included in the count.

Standard Deviation (σ): The sample standard deviation is computed over the set of BG levels in a CGM graph. The intuition is that an excessively variable day will have a higher standard deviation than an acceptable day.

Area under the Curve (AUC): This feature is computed as the total area between the CGM graph and a horizontal line corresponding to the minimum blood glucose level measured for that day. The intuition is that a larger area correlates with increased glycemic variability.

Roundness Ratio (RR): This feature is proportional to the ratio between the square of the perimeter of the CGM graph and its area. In the general case of two-dimensional objects, this feature will take the value of 1 for a perfect circle, and larger values as the objects deviate more from a circular shape. The intuition is that an excessively variable day, with its more jagged perimeter, will have a larger roundness ratio than an acceptable day.

Direction Codes (DC_k): A direction code is the absolute difference between the values of two consecutive blood glucose readings. Consequently, a CGM plot with n blood glucose measurements has $n - 1$ direction codes. Direction code bins record the percentage of direction codes having values within set ranges. The bins used to define the DC features for this application are $b_1 = [0, 3)$, $b_2 = [3, 6)$, and $b_3 = [6, 9)$. We may expect more blood glucose spikes in the third bin to reflect greater variability.

Experimental Results

We performed automatic feature selection on the development data set for both raw and smoothed CGM data using two methods: filtering with Welch's *t*-test (Filter) and greedy backward elimination (Greedy). We trained and evaluated three learning algorithms: naive Bayes, support vector machines, and multilayer perceptrons. The naive Bayes model assumes that the features are independent of each other given the class label. One advantage of this assumption is that it leads to a reduced number of parameters that can be computed very efficiently. Since all the glycemic variablility features are continuous, their conditional probabilities are represented as Gaussian distributions, with the mean and standard deviation estimated using maximum likelihood. The disadvantage of the conditional independence assumption is that it is often violated in practice, which may lead to suboptimal performance. The features EF and DT, for example, are not independent given the label, since each excursion greater than 75 mg/dl represents a large distance traveled. Multilayer perceptrons and support vector machines, on the other hand, can accommodate many overlapping features. MPs that are implemented as a backpropagation network with enough hidden nodes can approximate any decision surface (Hornik, Stinchcombe, and White 1989). Similarly, an SVM with a Gaussian kernel is a very flexible learning model, as it can approximate highly nonlinear decision boundaries. SVMs are known to be resilient to overfitting and to have good generalization performance, due to the max-margin criterion used during optimization. Furthermore, while the MP solution may be only a local optimum, the SVM is guaranteed to converge to a global optimum due to the corresponding convex optimization formulation.

All three algorithms obtained the best performance when run on the smoothed CGM graphs. The best performing configuration of features for each ML algorithm together with the corresponding performance results are shown in tables 1 and 2. In terms of accuracy, the best results are obtained by a multilayer perceptron trained on smoothed CGM plots with an ordered set of features automatically selected using greedy backward elimination. These results provide proof of concept, but refinement is

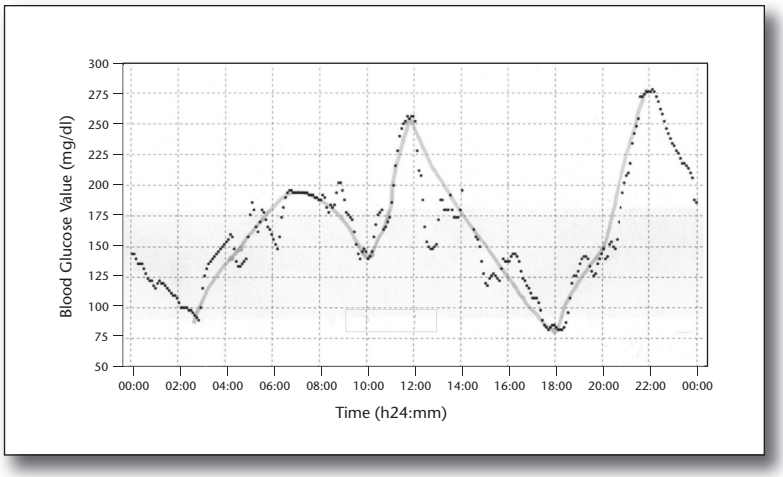

Figure 6. CGM Plot Smoothed by a Physician.

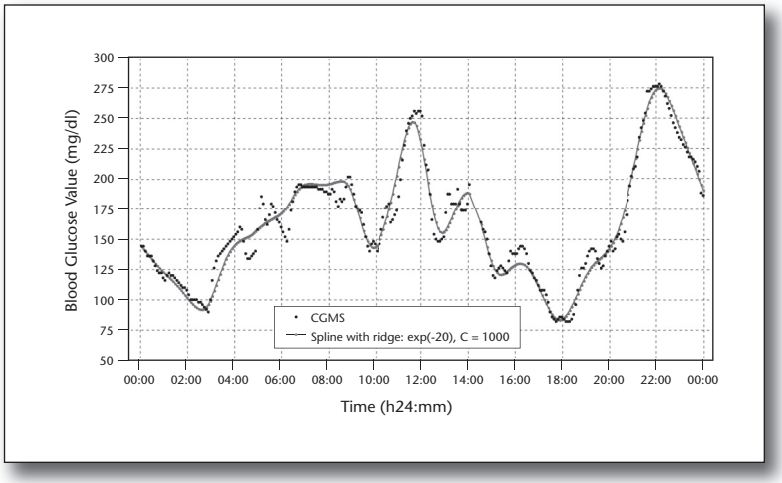

Figure 7. CGM Plot with Regularized Cubic Spline Smoothing.

Model	Selection	Features
NB	*Filter*	DC_1, *DT*, *AUC*, DC_3, *EF*, σ
SVM	*Filter*	DC_1, *DT*, AUC, DC_3, EF, σ
MP	*Greedy*	*DT*, *RR*, DC_3, *MAGE*, σ

Table 1. Best Configuration Features for Each ML Algorithm.

Model	Accuracy	Sensitivity	Specificity
NB	91.9 percent	91.6 percent	92.0 percent
SVM	92.8 percent	85.0 percent	96.0 percent
MP	93.8 percent	86.6 percent	96.6 percent

Table 2. Best Configuration Results for Each ML Algorithm.

needed to build a clinically viable tool. In particular, we must minimize the number of false positives so that clinicians can use the tool with confidence. We are currently working to improve the results by acquiring additional training examples from multiple clinicians and running a new set of experiments.

Support Vector Regression for BG Prediction

Detecting BG problems, as in 4DSS and the screen for excessive glycemic variability, allows corrective action to be taken. 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 affect patient safety. For example, the sleeping patient in the sample case (figure 1) could be awakened and advised to eat before becoming hypoglycemic. Then she would not lie in a dangerous state all night long. In fact, the most clinicially significant application for this technology is a hypoglycemia predictor that would reside in a patient's pump and sound an alarm before hypoglycemia would otherwise occur. Such a predictor would aid not only patients, but their families as well. Parents of young children with T1D frequently get up in the night to check for low BG levels. A reliable hypoglycemia alarm would allow families to rest more comfortably.

A significant part of our current research effort is directed toward designing ML models that can be trained on patient data to predict BG levels. Figure 8 shows the basic architecture of our approach to BG level prediction. Since BG measurements have a natural temporal ordering, we approach the task of predicting BG levels as a time-series forecasting problem. In time-series prediction, the task is to estimate the future value of a target function based on current and past data samples. Numerous prediction problems in a wide array of domains ranging from finance (for example, stock market [Kim 2003]), to medicine (for example, sleep apnea [Aguirre, Barros, and Souza 1999]), environment (for example, air quality [Perez and Reyes 2001]), or power systems (for example, electric utility load [Chen, Chang, and Lin 2004]) have been approached in the past as time-series forecasting problems.

We have conducted a preliminary experimental evaluation in which a support vector regression (SVR) model (Smola and Schölkopf 1998) 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 (and including) 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. We trained and tested two separate SVR models to predict the BG levels for 30 and 60 minutes into the future. These intervals would allow enough time to intervene to prevent predicted problems.

We represented training and testing examples as feature vectors using the following set of features:

The BG level (BG_0) of patient x at present time t_0.

A simple moving average (*MA*) over four past points from, and including, t_0.

An exponentially smoothed rate of change (*RC*) in BG level over fours past points from, and including, t_0.

Bolus dosage totals starting 30 minutes before prediction time, computed for durations of 30 minutes and 10 minutes, respectively. The bolus dosage refers to insulin that is injected before meals and/or to correct for hyperglycemia.

Basal rate averages starting right before prediction time, over 5 or 15 minute time intervals. This is the rate at which insulin is slowly and continuously infused into the patient by the pump. The basal rate changes throughout the day to accommodate changing insulin needs.

Meal carbohydrate amounts starting 30 minutes before prediction time, for durations of 30 minutes and 15 minutes, respectively.

Exercise intensity averages starting right before prediction time, over 5-, 30-, or 60-minute time intervals. Exercise tends to amplify the effect of insulin. This effect influences BG levels during and after exercise; the length of the effect depends on the

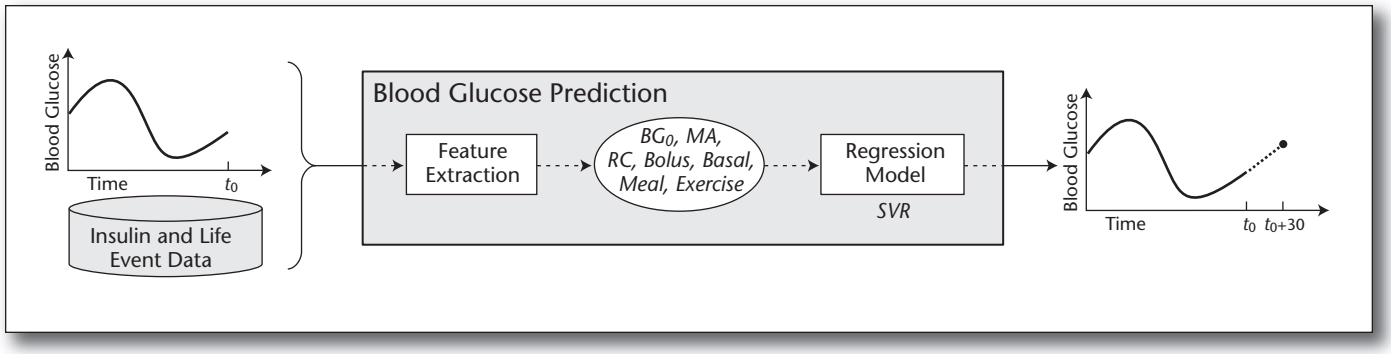

Figure 8. BG Level Prediction Pipeline.

length and the intensity of the exercise.

All data used to build the feature vectors was extracted from the 4DSS database. The BG data came from CGM sensors, which supply a data point every 5 minutes. The bolus and basal insulin data was recorded by the patient's insulin pump. Meal and exercise data was entered by the patient.

The influence that each type of event exerts on the BG level is known to vary with time. This specific time-dependent variability was taken into account through the offset and the length of the various time intervals that were used to define the features above. For example, the effects of exercise are strongest while the patient is exercising, but they may persist for several hours, especially if exercise is intense. This is why exercise features are computed in shorter 5-minute intervals close to the time of exercise, with intervals lengthening to 30 and 60 minutes as exercise recedes into the past. The SVR models were trained with a linear kernel, using a capacity parameter $C = 100$, and a default tube width $\epsilon = 1.0$. We used the LIBSVM implementation of SVMs for regression.[2]

In table 3, we compare the performance of the SVR models trained to predict BG level for 30 and 60 minutes into the future with the simple baseline $BGL(x, t_0)$ that uses the present BG level to predict any future BG level value. We use this simple baseline for comparison only because it was found to outperform more complex moving average and rate of change baselines.

We report the root mean square error E_{RMS}, the coefficient of determination R^2, and the percentage of predictions falling in the five areas from A to E in the Clarke Error Grid Analysis (CEGA) (Kovatchev et al. 2004). CEGA is a standard for evaluating the accuracy of BG measurement that is normally used to assess the quality of blood glucose sensors. As can be seen in figures 9 and 10, the Clarke Error Grid breaks a scatter plot into five regions:

(A) Points within 20 percent of the actual BG value; (B) Points that are more than 20 percent off but that would not lead to inappropriate treatment; (C) Points leading to unnecessary, but not harmful, treatment; (D) Points that obscure hypoglycemia or hyperglycemia, leading to a lack of necessary treatment; and (E) points misclassifying hypoglycemia as hyperglycemia, or vice versa, leading to harmful treatment.

30-Minute Predictions							
Method	E_{RMS}	R^2	A	B	C	D	E
SVR	**18.0**	**0.92**	**93.0**	7.0	0.0	0.0	0
$BGL(x, t_0)$	25.1	0.84	87.8	11.8	0.0	0.4	0
60-Minute Predictions							
Method	E_{RMS}	R^2	A	B	C	D	E
SVR	**30.9**	**0.76**	**81.0**	18.1	0.4	0.5	0
$BGL(x, t_0)$	43.2	0.52	74.5	21.5	2.2	1.8	0

Table 3. SVR and Baseline $BGL(x, t_0)$ Results.

The SVR models are promising, as they outperform the baselines on all performance measures. The two CEGA plots in figures 9 and 10 show the performance of the $BGL(x, t_0)$ baseline and SVR, respectively, for the 60-minute prediction time. The plots clearly show that, overall, the learned SVR model makes predictions that are closer to the ideal diagonal line. We believe that features based on more sophisticated time-series analysis models such as autoregressive moving averages (Box, Jenkins, and Reinsel 2008) have the potential to further improve the performance, and as such they are an important part of our current research efforts.

To account for individual patient differences, a predictive model is trained for each patient. We also plan to explore transfer learning approaches that effectively exploit data coming from multiple patients in order to improve the model predictions, which will be especially useful for patients

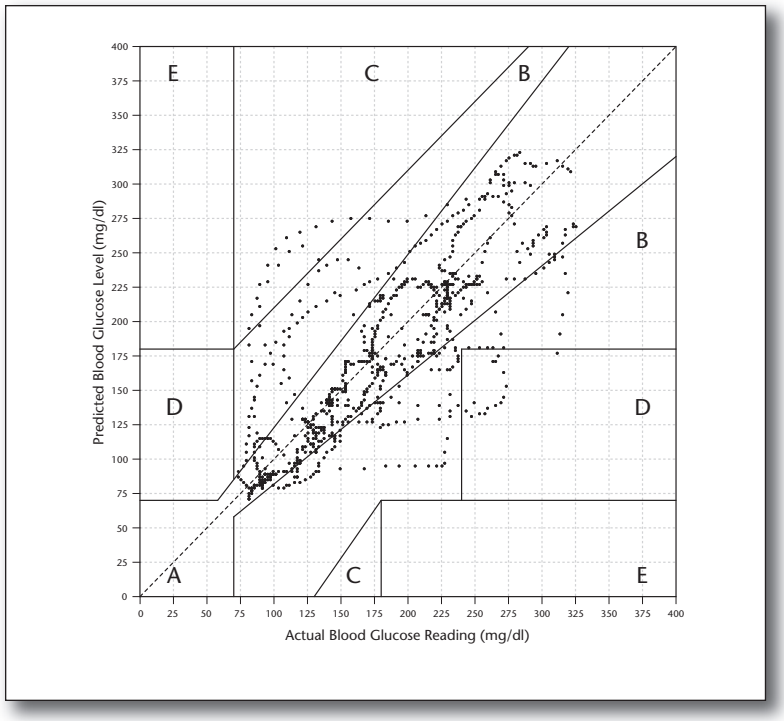

Figure 9. Performance of Baseline for 60-Minute Prediction.

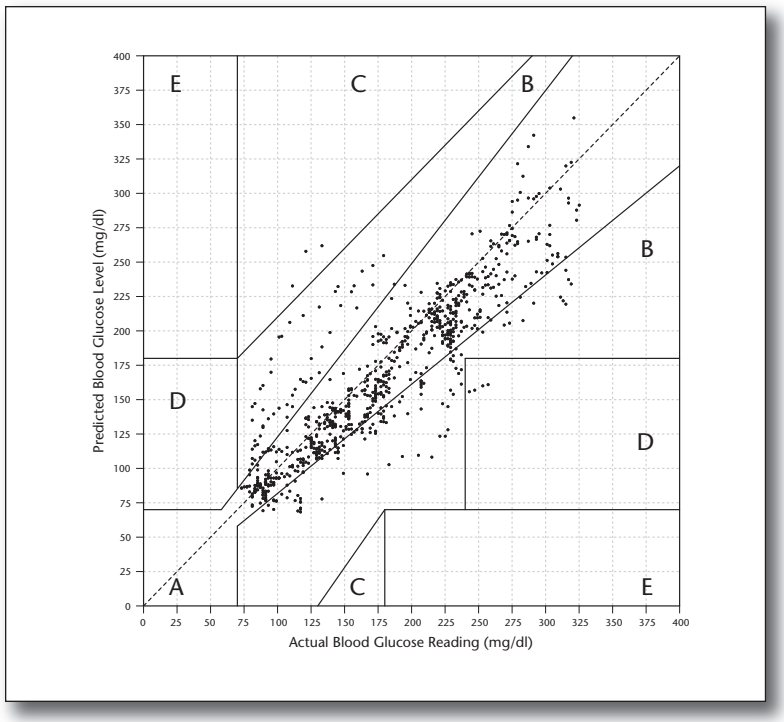

Figure 10. Performance of SVR for 60-Minute Prediction.

with limited historical data. Trained prediction models will be stored in a new case base of models, so that we may further consider the possibilities of adapting past models to bootstrap predictions for new patients. Since the patient data is often inaccurate or incomplete, we are investigating learning methods that are robust in the presence of missing or uncertain data and that can also identify data anomalies automatically.

Technology Deployment Challenges

In this section, we share observations on three nontechnical issues we encountered in making the transition from intelligent diabetes management tools in the university setting to the real world. We (CM and FS) do this from the perspective of faculty members intent upon keeping our "day jobs" of conducting academic research, educating students, and treating patients (FS). We recognize that issues differ for AI researchers working for companies or starting up their own companies, and that issues may be more easily raised than resolved.

University Technology Transfer

The goal of the university technology transfer office (TTO) is to facilitate commercialization of intellectual property (IP) and to ensure that the university benefits financially from the ideas of its faculty. Working with the TTO on technology transition, however, presents challenges for academics. Economic goals do not always harmonize with the faculty-held tenet of broad dissemination of knowledge. Faculty members have to "publish or perish," writing articles like this one to share their findings with the greater research community. Protecting IP, on the other hand, means not sharing findings until appropriate legal steps have been taken. One of us (FS) had a paper withheld from publication for three years while IP was being secured, effectively rendering the findings obsolete. Our approach has been to communicate regularly with the TTO, aiming to leverage its expertise while minimizing its impact on our academic productivity.

Patents

The TTO filed a patent application on our behalf. Should software be patented? This is a current controversy, with debaters arguing both sides of the question (Computer History Museum 2011). Issues range from the moral (for example, "Shouldn't users be entitled to free software?") to the methodological (for example, "Wouldn't copyright be a better mechanism for protecting software?") to the utilitarian (for example, "Do patents actually promote or discourage innovation in practice?"). In our case, the TTO offered two compelling arguments that led to the patent filing: (1) companies would not invest in unpatented technologies; and (2) freeware is not suitable for safety-critical medical applications.

Safety

When AI technology is to be used directly by patients in the United States, it must first be approved by the U.S. Food and Drug Administration (FDA). While critical for ensuring patient safety, the FDA approval process entails extensive investments of time and money, making it infeasible for academics. Consequently, we have focused primarily on tools that help physicians manage patients, rather than tools used directly by patients. This ensures patient safety by keeping health-care professionals in the loop, removing FDA concerns but limiting the avenues of application. One way to feasibly navigate the FDA review process would be to partner with a medical device company, and we have held exploratory conversations with several.

Forging Ahead

We are tackling these challenges and forging ahead with plans to make intelligent diabetes management a reality for patients and physicians. We have a waiting list of patients who have volunteered to participate in clinical research studies. They are counting on us to translate the research into practical tools they can use. We envision a number of potential avenues of commercialization and use. The software could be marketed directly to physicians for office use; it could be included in electronic health record (EHR) systems; it could be embedded in insulin pumps and smartphones for patient use; it could be incorporated in continuous glucose monitoring (CGM) systems, so that all BG plots would come with associated analyses; and BG control centers could be established, where data could be uploaded, analyzed, and monitored by advanced practice nurses, who would forward appropriate findings to physicians and patients. In summary, diabetes management is more than a challenging domain for AI research. It is an opportunity for AI applications to positively impace the health and well-being of people with diabetes.

Acknowledgments

This work is supported, in part, with funding from the National Science Foundation under the Smart Health and Wellbeing program (award IIS-1117489). We gratefully acknowledge support from Medtronic, the Ohio University Russ College Biomedical Engineering Fund, the Ohio University Heritage College of Osteopathic Medicine Research and Scholarly Affairs Committee, and the Ohio University Diabetes Research Initiative. We also thank our graduate research assistants, our research nurses, and more than 50 anonymous patients with T1D for their contributions to this work.

Beyond detecting problems and suggesting solutions is the potential to intervene and prevent problems. A machine learning model could reside in the patient's pump/sensor device, assess glucose recordings continuously, and predict hypoglycemia 30 minutes prior to its occurrence, thereby preventing it. I imagine that a similar system, monitoring cardiac heart rhythm tracings, could learn to predict dangerous arrhythmias before they occur, since there are usually prodromal changes in tracings prior to adverse events.

– Dr. Frank Schwartz

Note

1. See the Carelink Personal Software page at the Medtronic website, www.medtronicdiabetes.net/products/carelinkpersonalsoftware.

2. See C.-C. Chang and C.-J. Lin's LIBSVM: A Library for Support Vector Machines. Software available at www.csie.ntu.edu.tw/~cjlin/libsvm.

References

Aguirre, L. A.; Barros, V. C.; and Souza, A. V. 1999. Nonlinear Multivariable Modeling and Analysis of Sleep Apnea Time Series. *Computers in Biology and Medicine* 29(3): 207–228.

Bichindaritz, I. 2008. Case-Based Reasoning in the Health Sciences: Why It Matters for the Health Sciences and for CBR. In *Advances in Case-Based Reasoning: 9th European Conference,* ECCBR, ed. K.-D. Althof, R. Bergmann, M. Minor, and A. Hanft, 1–17. Berlin: Springer.

Bishop, C. M. 1995. *Neural Networks for Pattern Recognition*. New York: Oxford University Press.

Box, G.; Jenkins, G.; and Reinsel, G. 2008. *Time Series Analysis: Forecasting and Control*. Hoboken, NJ: John Wiley.

Ceriello, A., and Ihnat, M. A. 2010. Glycaemic Variability: A New Therapeutic Challenge in Diabetes and the Critical Care Setting. *Diabetic Medicine* 27(8): 862–867.

Chen, B.-J.; Chang, M.-W.; and Lin, C.-J. 2004. Load Forecasting Using Support Vector Machines: A Study on EUNITE Competition 2001. *IEEE Transactions on Power Systems* 19(4): 1821–1830.

Computer History Museum. 2011. Software Patent Debate. Mountain View, CA. Video available at www.computerhistory.org/events/video/?videoid=f6Dh5NjlZMk.

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.; Thorpe, C.; Mahmoud, M.; and Zirie, M. 2008. Intelligent Diabetes Assistant: Using Machine Learning to Help Manage Diabetes. In *Proceedings of the IEEE/ACS International Conference on Computer Systems and*

Applications, 913–914. Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Hornik, K.; Stinchcombe, M.; and White, H. 1989. Multilayer Feedforward Networks are Universal Approximators. *Neural Networks* 2(5): 359–366.

Juvenile Diabetes Research Foundation. 2011. General Diabetes Facts. New York: JDRF (www.jdrf.org/index.cfm?page id=102586).

Kilpatrick, E. S.; Rigby, A. S.; and Atkins, S. L. 2010. For Debate. Glucose Variability and Diabetes Complication Risk: We Need to Know the Answer. *Diabetic Medicine* 27(8): 868–871.

Kim, K.-J. 2003. Financial Time Series Forecasting Using Support Vector Machines. *Neurocomputing* 55: 307–319.

Kovatchev, B. P.; Gonder-Frederick, L. A.; Cox, D. J.; and Clarke, W. L. 2004. Evaluating the Accuracy of Continuous Glucose-Monitoring Sensors: Continuous Glucose-Error Grid Analysis Illustrated by TheraSense Freestyle Navigator Data. *Diabetes Care* 27(8): 1922–1928.

Marling, C. R.; Shubrook, J. H.; Vernier, S. J.; Wiley, M. T.; and Schwartz, F. L. 2011. Characterizing Blood Glucose Variability Using New Metrics with Continuous Glucose Monitoring Data. *Journal of Diabetes Science and Technology* 5(4): 871–878.

Marling, C.; Shubrook, J.; and Schwartz, F. 2008. Case-Based Decision Support for Patients with Type 1 Diabetes on Insulin Pump Therapy. In *Advances in Case-Based Reasoning: 9th European Conference,* ECCBR, ed. K.-D. Althof, R. Bergmann, M. Minor, and A. Hanft, 325–339. Berlin: Springer.

Marling, C.; Shubrook, J.; and Schwartz, F. 2009. Toward Case-Based Reasoning for Diabetes Management: A Preliminary Clinical Study and Decision Support System Prototype. *Computational Intelligence* 25(3): 165–179.

Monnier, L.; Mas, E.; Ginet, C.; Michel, F.; Villon, L.; Cristol, J.; and Colette, C. 2006. Activation of Oxidative Stress by Acute Glucose Fluctuations Compared with Sustained Chronic Hyperglycemia in Patients with Type 2 Diabetes. *Journal of the American Medical Association* 295(14): 1681–1687.

Montani, S.; Magni, P.; Bellazzi, R.; Larizza, C.; Roudsari, A. V.; and Carson, E. R. 2003. Integrating Model-Based Decision Support in a Multi-Modal Reasoning System for Managing Type 1 Diabetic Patients. *Artificial Intelligence in Medicine* 29(1–2): 131–151.

Perez, P., and Reyes, J. 2001. Prediction of Particulate Air Pollution Using Neural Techniques. *Neural Computing and Applications* 10(2): 165–171.

Rodbard, D. 2009. Interpretation of Continuous Glucose Monitoring Data: Glycemic Variability and Quality of Glycemic Control. *Diabetes Technology and Therapeutics* 11(s1): S–55–S–67.

Schölkopf, B., and Smola, A. J. 2002. *Learning with Kernels—Support Vector Machines, Regularization, Optimization and Beyond.* Cambridge, MA: MIT Press.

Schwartz, F. L.; Vernier, S. J.; Shubrook, J. H.; and Marling, C. R. 2010. Evaluating the Automated Blood Glucose Pattern Detection and Case-Retrieval Modules of the 4 Diabetes Support System. *Journal of Diabetes Science and Technology* 4(6): 1563–1569.

Schwartz, F. L.; Shubrook, J. H.; and Marling, C. R. 2008. Use of Case-Based Reasoning to Enhance Intensive Management of Patients on Insulin Pump Therapy. *Journal of Diabetes Science and Technology* 2(4): 603–611.

Service, F.; Molnar, G.; Rosevear, J.; Ackerman, E.; Gatewood, L.; and Taylor, W. 1970. Mean Amplitude of Glycemic Excursions, a Measure of Diabetic Instability. *Diabetes* 19(9): 644–655.

Smola, A. J., and Schölkopf, B. 1998. A Tutorial on Support Vector Regression. Technical Report TR-98-030, NeuroCOLT2 Technical Report Series.

Vapnik, V. N. 1995. *The Nature of Statistical Learning Theory.* Berlin: Springer-Verlag.

Wiley, M.; Bunescu, R.; Marling, C.; Shubrook, J.; and Schwartz, F. 2011. Automatic Detection of Excessive Glycemic Variability for Diabetes Management. In *Proceedings of the 10th International Conference on Machine Learning and Applications,* 148–154. Los Alamitos, CA: IEEE Computer Society.

Cindy Marling is an associate professor of computer science and adjunct associate professor of specialty medicine at Ohio University. She is director of the SmartHealth Lab, which promotes interdisciplinary research at the intersection of AI and medicine. Her research interests include case-based reasoning, hybrid reasoning architectures, and clinical decision support systems.

Matthew Wiley received his undergraduate and master's degrees in computer science from Ohio University. He is currently a PhD student in computer science at the University of California, Riverside. His research interests include machine learning, data mining, and health informatics.

Razvan Bunescu is an assistant professor of computer science at Ohio University. He received his PhD from the University of Texas at Austin in 2007 with a thesis on machine-learning models for information extraction. His current research is focused on applications of machine learning in computational linguistics and biomedical informatics.

Jay Shubrook, DO, is an associate professor in the department of family medicine and director of clinical research at the Ohio University Heritage College of Osteopathic Medicine. He is a practicing diabetologist at the Appalachian Rural Health Institute Diabetes and Endocrine Center. His research focuses on advanced technology for diabetes management and on new ways to reach the medically underserved.

Frank Schwartz, MD, is a professor of endocrinology and the J. O. Watson chair for diabetes research at the Ohio University Heritage College of Osteopathic Medicine. He is director of the Appalachian Rural Health Institute Diabetes and Endocrine Center, which serves more than 600 patients with type 1 diabetes on insulin pump therapy. He is a practicing endocrinologist with more than 25 years of clinical experience. His research encompasses both basic and applied work in diabetes, ranging from drug discovery to computing and biomedical technology.