Machine Learning Experiments with Noninvasive Sensors for Hypoglycemia Detection

Cindy Marling[1], Lijie Xia[1], Razvan Bunescu[1] and Frank Schwartz[2]
[1]School of Electrical Engineering and Computer Science,
Russ College of Engineering and Technology
[2]Department of Specialty Medicine,
Heritage College of Osteopathic Medicine
Ohio University, Athens, Ohio 45701, USA

Abstract

Accurate hypoglycemia detection would enable people with type 1 diabetes (T1D) to treat this dangerous condition promptly, improving health and safety. This paper presents machine learning experiments that aim to improve hypoglycemia detection by leveraging data from noninvasive sensors found in fitness bands. A middle-aged subject with T1D provided blood glucose and fitness band data for two months. Sensor data included heart rate, galvanic skin response, and skin and air temperatures. Statistical tests identified features derived from this data that could differentiate hypoglycemic from non-hypoglycemic states. Support vector machines (SVM) were then trained, using only these features, to classify instances as hypoglycemic or non-hypoglycemic. An SVM with a linear kernel was able to outperform two simple baselines. Results show proof of concept; however, system performance was limited by the size and nature of the dataset. Results are being used in ongoing work to improve the performance of overall blood glucose prediction models that use blood glucose, insulin, and life-event data.

1 Introduction and Motivation

For people with type 1 diabetes (T1D), hypoglycemia is a major health concern. Hypoglycemia is defined by low blood glucose levels, typically below 70 mg/dl. Initial symptoms, which vary from individual to individual, may include confusion, dizziness, weakness, hunger, nausea, shaking or sweating. If left untreated, severe hypoglycemia can lead to seizures, coma, or death. For this reason, low blood glucose levels are especially dangerous for individuals with hypoglycemia unawareness, or lack of symptoms, and for all individuals with T1D while they sleep.

People with T1D are advised to treat their hypoglycemia, as soon as they recognize it, by consuming a mixed-nutrient snack containing from 15 to 20 grams of carbohydrate, such as a granola bar or peanut butter crackers. This simple treatment will often restore blood glucose levels to normal within 15 minutes. The key, then, is to recognize when treatment is needed. People may recognize hypoglycemia by its symptoms; they may also learn of it through a continuous glucose monitoring (CGM) system. CGM systems, available by prescription, provide blood glucose readings every 5 minutes, in real time. The readings are derived from the interstitial fluid by sensors inserted just under the skin. While CGM data provides valuable insight for managing blood glucose levels, there are drawbacks to CGM systems. For one thing, measurements based on interstitial fluid lag actual blood glucose levels by approximately 10 minutes. There is some inherent noise, or inaccuracy, in the readings as well; frequent calibration with finger sticks is required. Furthermore, the cost of the CGM sensors may be prohibitive for some individuals.

With the recent advent of inexpensive, noninvasive, wearable physiological sensors, the opportunity arises to improve the detection and/or prediction of hypoglycemia. The goal is to be able to alert people as soon as possible to current or impending hypoglycemia, so that they can take the actions necessary to correct or prevent the problem.

The present work builds upon five years' experience building machine learning models for blood glucose level prediction. In past work, we have used features derived from CGM data, finger sticks, insulin doses, and manually recorded meals, sleep and exercise. A high-level overview of our past work is available in [Marling *et al.*, 2012]. In [Bunescu *et al.*, 2013], we introduced an adaptive model for blood glucose level prediction that outperformed predictions made by three diabetes experts, across a wide range of blood glucose levels. Recognizing that the greatest clinical need was for accurate predictions at low blood glucose levels, we reported our first experiment with hypoglycemia prediction in [Plis *et al.*, 2014]. There, we could only predict 23% of hypoglycemic episodes 30 minutes in advance. Hypothesizing that performance could be improved by incorporating features from newly available noninvasise sensors, in this paper we describe experiments on building hypoglycemia detection models that use data *only* from these sensors. In ongoing work, we leverage these results to improve the performance of the overall blood glucose level prediction models.

2 Experimental Dataset

Data was contributed by a middle-aged male who has had T1D since childhood. For two months, he wore a commercially available fitness band along with his regularly pre-

scribed medical devices, and he reported meals, sleep and exercise via a smart phone. The fitness band, a Basis Peak, provided data for heart rate (HR), galvanic skin response (GSR), and skin and air temperatures (ST and AT). The medical devices, a Medtronic insulin pump and a Dexcom CGM system, provided insulin and blood glucose data. He met with the authors on a weekly basis to review and analyze the collected data. The data was consolidated and displayed via custom-built graphical software, which allowed us to visualize and discuss it. We tried to identify visual patterns in the fitness band data corresponding to hypoglycemia observed in the CGM data.

Over the course of two months, there were 34 hypoglycemic episodes lasting 10 minutes or more. From each episode, we selected the timestamp corresponding to the lowest blood glucose level. The resulting 34 data points were used in the machine learning experiments as positive examples. Negative examples were selected so as to maintain the ratio of positive to negative data points among all of the data collected, which was 1 to 37. Therefore, $34 \times 37 = 1,258$ negative examples were randomly selected for inclusion in the dataset. In selecting negative examples, points were excluded if hypoglycemia occurred within one hour or if there was more than one hour of missing data within the past day. Exclusions were intended to ensure that negative examples were truly representative, but they may also have made the classification task easier than it otherwise would have been.

3 Feature Engineering

Features were derived from the raw data using insights obtained during data reviews. First, the difference between skin and air temperature (ST-AT) was used to compute features, rather than the two individual measures. Otherwise, cold temperatures might appear to presage hypoglycemia, simply because the subject exercised outdoors during Winter. Features were then implemented to capture physiological states over the past 24 hours and the past hour. The following features were implemented for each of HR, GSR and ST-AT:

1. Mean over the past 24 hours (Mean-24hr)
2. Standard deviation over the past 24 hours (SD-24hr)
3. Distance traveled per hour over the past 24 hours (DT-24hr)
4. Mean over the past hour (Mean-1hr)
5. Standard deviation over the past hour (SD-1hr)
6. Distance traveled over the past hour (DT-1hr)
7. Difference between the current value and the value one hour ago (Diff-1hr)

Above, distance traveled (DT) refers to the sum of the absolute differences in value between each pair of consecutive data points. Intuitively, DT may be thought of as stretching the data curve out flat, as if it were a crumpled string, to see how long it is. We included it among the features, because we have found that, when applied to CGM data, it is a useful measure of variability [Marling *et al.*, 2011].

An additional feature combined HR, GSR and ST-AT:

8. Change score (CScore)

The change score is a measure of how much HR, GSR, and ST-AT change over the past 5 minutes, in relation to how much they change every 5 minutes for the past 3 hours. Only increases in HR and GSR and decreases in ST-AT contribute to the change score. The change score was inspired by, but does not replicate, the measure used to detect nocturnal hypoglycemia in hospitalized patients by Schechter et al. [2012].

Because a patient may experience hypoglycemia at predictable times of the day, such as late afternoon, another type of feature was introduced to account for time of day. A Boolean feature was implemented for each of the four-hour time intervals: [00:00, 04:00), [04:00, 08:00), [08:00, 12:00), [12:00, 16:00), [16:00, 20:00), and [20:00, 00:00). For each timestamp, the value of the interval feature containing that timestamp is true, while the value of all other interval features is false. While the choice of granularity for the time intervals was somewhat ad hoc, smaller time intervals would have resulted in too many features for the size of the dataset. We chose generic time intervals, rather than patient-specific ones (e.g. intervals that reference sleep or work times), although such features could also be useful. Even though our subject experienced hypoglycemia during all time intervals, it was most frequent overnight and during late afternoons.

Statistical tests were run to see which, if any, of these basic features could differentiate hypoglycemic episodes from non-hypoglycemia. Welch's t-test was used for the numeric physiological features, while a χ^2 test was used for the Boolean time interval features. Since no statistically significant difference was observed for any basic feature at the 0.05 level, new combination features were then implemented by masking each of the 22 physiological features by each of the 6 time interval features. Each of these 132 combination features maintains the value of its physiological feature during its specified time interval, but has a zero value during the other 20 hours of the day. So, for example, the value of the combination feature [00:00, 04:00) GSR SD-24hr is equal to the SD of the GSR over the previous 24 hours if the point to be classified has a timestamp between midnight and 4:00 AM and is equal to 0 otherwise. Welch's t-test showed statistically significant differences at a $p < 0.05$ level for 36 of the combination features, as shown in Table 1. Each example in the dataset, then, consists of 36 features, plus a label designating it as representative of hypoglycemia or not.

4 Machine Learning Experiments

Support Vector Machines (SVMs) were trained to classify instances as hypoglycemic (positive) or non-hypoglycemic (negative). SVMs are a state-of-the-art supervised learning algorithm that can effectively handle large numbers of possibly overlapping features [Schölkopf and Smola, 2002; Vapnik, 1995]. LibSVM [Chang and Lin, 2001] was used for the implementation. The discriminant function computed by the SVM is proportional to the margin, i.e. the distance between the example and the decision hyperplane, and can be used as a measure of confidence in the system classification. Once a threshold is selected (by default 0), examples with values at or above the threshold are classified as positive, while

Table 1: Statistically Significant Features for Hypoglycemia Detection

Time Interval	Physiological Feature	t	p
[20:00, 00:00)	CScore	-7.842	<0.001
[16:00, 20:00)	GSR SD-24hr	-6.399	<0.001
[16:00, 20:00)	GSR Mean-24hr	-5.971	<0.001
[16:00, 20:00)	GSR DT-24hr	-5.933	<0.001
[12:00, 16:00)	CScore	-6.036	<0.001
[04:00, 08:00)	GSR SD-24hr	-5.244	<0.001
[04:00, 08:00)	GSR Mean-24hr	-5.174	<0.001
[04:00, 08:00)	GSR DT-24hr	-5.153	<0.001
[20:00, 00:00)	ST-AT SD-1hr	-5.119	<0.001
[20:00, 00:00)	ST-AT SD-24hr	-4.081	<0.001
[20:00, 00:00)	HR SD-24hr	-3.940	<0.001
[20:00, 00:00)	ST-AT DT-24hr	-3.893	<0.001
[20:00, 00:00)	ST-AT Mean-24hr	-3.882	<0.001
[20:00, 00:00)	ST-AT Mean-1hr	-3.786	<0.001
[20:00, 00:00)	ST-AT DT-1hr	-3.786	<0.001
[20:00, 00:00)	HR DT-24hr	-3.531	0.001
[20:00, 00:00)	HR Mean-24hr	-3.530	0.001
[20:00, 00:00)	HR DT-1hr	-3.424	0.001
[20:00, 00:00)	HR Mean-1hr	-3.424	0.001
[20:00, 00:00)	HR SD-1hr	-3.284	0.002
[20:00, 00:00)	GSR Mean-24hr	2.630	0.010
[16:00, 20:00)	ST-AT Mean-1hr	2.604	0.013
[16:00, 20:00)	ST-AT DT-1hr	2.604	0.013
[20:00, 00:00)	GSR DT-24hr	-2.523	0.014
[20:00, 00:00)	ST-AT Diff-1hr	2.458	0.016
[16:00, 20:00)	HR DT-1hr	2.317	0.026
[16:00, 20:00)	HR Mean-1hr	2.317	0.026
[16:00, 20:00)	HR DT-24hr	2.256	0.030
[16:00, 20:00)	HR Mean-24hr	2.255	0.030
[16:00, 20:00)	HR SD-24hr	2.247	0.030
[16:00, 20:00)	ST-AT Mean-24hr	2.244	0.031
[16:00, 20:00)	ST-AT DT-24hr	2.239	0.031
[16:00, 20:00)	ST-AT SD-24hr	2.206	0.033
[00:00, 04:00)	GSR SD-24hr	2.106	0.042
[08:00, 12:00)	GSR Mean-1hr	-1.980	0.048
[08:00, 12:00)	GSR DT-1hr	-1.980	0.048

those with values below the threshold are classified as negative.

To use an SVM to detect hypoglycemia in practice, the threshold would be selected to achieve a desired trade-off between sensitivity and specificity. For experimental purposes, the behavior of the SVM across the entire spectrum of possible thresholds is examined. Since the label distribution is skewed towards negative examples, it is useful to view results in terms of precision and recall. Note that "recall" is equivalent with "sensitivity;" it is also known as the "true positive rate." "Precision" is also known as "positive predictive value." Specificity and overall accuracy would not be fair metrics to use with this dataset; due to the large ratio of negative to positive examples, they would overstate the goodness of performance.

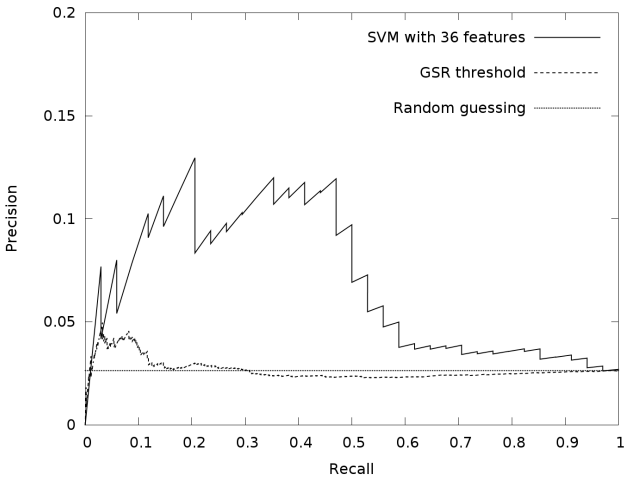

Figure 1: Precision-Recall Curve for the Linear SVM with 36 Features. Note that the y-axis has been truncated, as precision is always very low (<0.15).

To train and evaluate the SVM models, the 1,292 examples were grouped by day and then partitioned into folds for 25-fold cross-validation. We used 25 folds because the 34 hypoglycemic episodes occurred on 25 different days. Grouping the data by day ensures that the data used to train and to test an SVM always comes from different days. Each fold contains one day with hypoglycemic events, and a number of days with no hypoglycemic events chosen such that the ratio of negative to positive examples in the fold is close to the observed ratio of 37 to 1. The kernel parameters were tuned for each test fold on a separate validation fold. Because of the imbalance in the number of positive and negative examples, the weight parameter for positive examples was set to 37, the ratio of negative to positive points. Linear, Gaussian, and quadratic kernels were all used in early experiments. However, due to the relatively large number of features (36) with respect to the small number of positive examples (34), the Gaussian and quadratic kernels overfit and did not perform well. Later experiments used only the linear kernel.

Initial results come from a linear SVM trained on all 36 features. The performance of this SVM is compared to that of two simple baselines: random guessing and GSR threshold. In the random guessing baseline, the probability of guessing that a point is hypoglycemic is varied from 0 to 1, so that recall also varies from 0 to 1. However, precision remains constant at 1/38, since 1 in every 38 points is actually hypoglycemic. The GSR threshold is a more informed method of classification. It was selected because GSR, also known as skin conductance, is a measure of sweating, which is known to happen during hypoglycemic episodes. The curves for this baseline are calculated by varying the GSR value above which a point is classified as hypoglycemic.

When an SVM with a linear kernel was trained using all 36 features shown in Table 1, performance exceeded both baselines. The precision-recall curve is shown in Figure 1. Note that, because precision is quite low, the y-axis, which nor-

mally runs from 0 to 1, was truncated at 0.2, to remove unnecessary white space and accentuate the region of interest.

To alleviate overfitting, the number of features was further narrowed down in the next experiment using feature selection inside the 25-fold cross validation loop. Features were greedily selected from the original set of 36 discriminative features to maximize the area under the precision-recall curve (AUC). Pseudocode is provided in Figure 2.

```
for i = 1 to 25
{
  // Fold i is the test fold.  Folds
  // (i+1 modulo 25) and (i+2 modulo 25)
  // are development folds for tuning
  // and feature selection.  The other
  // 22 folds are training folds.

  Tune the C parameter with grid search

  BestAUC = 0
  NumFeatures = 36
  FeaturePool = {all features}
  ChosenFeatures = {}
  Repeat {
    for j = 1 to NumFeatures
    {
      Use ChosenFeatures + Feature j
      Train and test, record AUC[j]
    }
    Choose Feature j with max AUC
    if AUC[j] > BestAUC
    {
      Add Feature j to ChosenFeatures
      Remove Feature j from FeaturePool
      NumFeatures = NumFeatures-1
      BestAUC = AUC[j]
    }
  } Until AUC[j] <= BestAUC

  Train and test with tuned C and
  ChosenFeatures
}
```

Figure 2: Pseudocode for 25-fold Cross Validation with Greedy Forward Feature Selection.

Of the original 36 features, 21 were chosen for inclusion in at least one of the 25 folds. The most commonly selected feature was chosen for inclusion in 14 folds. Table 2 shows the features in order of how frequently they were selected. The performance of the SVM with selected features is shown in comparison to that of the SVM with all 36 features in Figure 3.

5 Discussion and Future Work

The results indicate that physiological sensor data could potentially improve blood glucose level prediction, in general, and hypoglycemia detection, in particular. However, in our experiments, system performance was severely limited, due, in large part, to the following factors:

Table 2: Features Chosen for at Least One Fold by Greedy Forward Feature Selection

Time Interval	Physiological Feature	Number of Folds
[04:00, 08:00)	GSR SD-24hr	14
[12:00, 16:00)	CScore	13
[16:00, 20:00)	GSR SD-24hr	12
[20:00, 00:00)	ST-AT SD-1hr	11
[00:00, 04:00)	GSR SD-24hr	10
[08:00, 12:00)	GSR Mean-1hr	9
[16:00, 20:00)	ST-AT Mean-1hr	6
[20:00, 00:00)	ST-AT Diff-1hr	4
[20:00, 00:00)	CScore	3
[16:00, 20:00)	GSR Mean-24hr	3
[16:00, 20:00)	HR DT-1hr	3
[20:00, 00:00)	ST-AT SD-24hr	2
[16:00, 20:00)	ST-AT SD-24hr	2
[20:00, 00:00)	GSR Mean-24hr	2
[20:00, 00:00)	HR DT-24hr	2
[04:00, 08:00)	GSR Mean-24hr	1
[04:00, 08:00)	GSR DT-24hr	1
[08:00, 12:00)	GSR DT-1hr	1
[16:00, 20:00)	HR SD-24hr	1
[16:00, 20:00)	ST-AT Mean-24hr	1
[16:00, 20:00)	HR DT-24hr	1

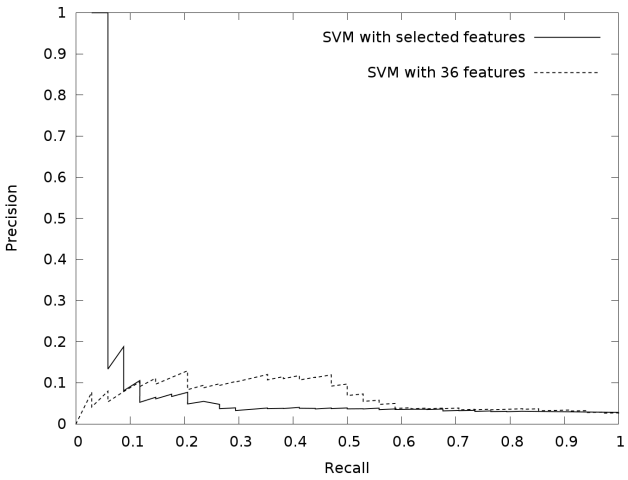

Figure 3: Precision-Recall Curve for the SVM with Selected Features vs. that of the SVM with All 36 Features

- The dataset was small and skewed toward non-hypoglycemic events
- All of the data was acquired from a single patient
- Only features acquired from the physiological sensor band were used
- Data was acquired while the patient led his normal, everyday life

Consider, first, that our small dataset contained only 34 hypoglycemic events. Clearly, more hypoglycemic events would translate into more positive examples, which could improve the performance of the machine learning model. However, each bout of hypoglycemia is a negative experience for the person with diabetes, who therefore tries hard to avoid such events.

Second, all of the data was acquired from a single subject, although there is great variability among individuals with diabetes. Our subject was a middle-aged male who adhered to best practices for diabetes management and had excellent blood glucose control. One individual strategy he employed was to exercise, by walking briskly, when his blood glucose levels were high (hyperglycemia). He would do this in lieu of, or in addition to, the more common strategy of taking extra insulin to correct for hyperglycemia. This may have confounded the ability to use GSR as an indicator of hypoglycemia. While GSR rises with hypoglycemia, it also rises with exercise.

Third, we intentionally used only features based on physiological sensor data in order to test the usefulness of these sensors. We also collect CGM, insulin, and patient-entered life-event data for use in blood glucose prediction models. A patient may enter life events that impact blood glucose via their smart phone. For example, they may enter that they are beginning to exercise, feeling stressed, going to sleep, or eating. While this paints a broader picture of the world around the patient, it can be burdensome for the patient and is subject to inaccuracies and omissions. In future work, we plan to explore how wearable, unobtrusive sensors can augment, validate, or even eliminate patient life-event data entry as indicators of the physiological state of the patient.

Schechter et al. [2012] used only physiological sensors to detect nocturnal hypoglycemia in hospitalized adolescents with diabetes. Their goal was to replace the expensive, invasive CGM sensors currently in use with inexpensive, noninvasive sensors. They reported a sensitivity of 100% with a specificity of 85.7%. One of the physiological parameters they measured was tremor, for which we did not have a sensor. Shaking, like sweating, is a symptom of hypoglycemia. Their sensors, however, were not all mobile, and their patients were all in bed in a controlled hospital environment throughout the experiment. Results have not yet been extended to the outpatient environment, but near-term feasibility increases as fitness bands continue to improve the number, type and accuracy of included sensors. In the meantime, our patient led his normal, everyday life as we collected data. He worked a demanding job, enjoyed an active sex life, flew across time zones on airplanes, overate at family celebrations, and missed or delayed meals when pressed at work.

Attempts to model blood glucose levels date back to the 1960s [Boutayeb and Chetouani, 2006]. While no definitive model exists yet, it should be noted that early efforts were hindered by the lack of CGM data, which first became available in 1999. Most blood glucose models developed to date are mathematical formalisms of physiological processes. AIDA is an early, freely available model [Lehmann and Deutsch, 1992]. Another influential, but proprietary, model was developed at the University of Virginia [Kovatchev *et al.*, 2009]. These models are commonly used to simulate diabetes patients when actual patient data is unavailable. Work more in line with our own approach was reported by Duke [2009], who combined a physiological model with Gaussian process regression. The clinical importance of solving this problem, combined with the high level of technical challenge, has led to an uptick in recent research [Jensen *et al.*, 2013; Zecchin *et al.*, 2013; Wang *et al.*, 2014].

A major impediment to progress in solving this problem is the lack of actual patient data upon which to build models and experiment. Patient privacy concerns and regulations, including The Health Insurance Portability and Accountability Act of 1996 (HIPAA), make it difficult for researchers to share data. We are currently collecting data from two additional subjects, and we have Institutional Review Board (IRB) approval to collect data from up to 30 more for our blood glucose prediction research. We are asking subjects, during the informed consent process, for permission to de-identify and share the data we collect. As future work, we plan to organize one or more workshops for researchers interested in this problem to come together and explore different approaches to blood glucose prediction using this common dataset.

Feature engineering is an essential part of our machine learning approach to hypoglycemia detection. While the results reported in this paper indicate the potential utility of noninvasive physiological sensors for modeling blood glucose behavior, it is unclear whether the current set of manually designed features is optimal. Given the small number of hypo events in our dataset, trying more features raises the likelihood of finding spurious correlations with the label. As we collect more data, we plan to explore the use of other information available from the fitness band, including sleep state and step count, as well as other formulations of HR, GSR, ST and AT. In future work, we plan to leverage recent advances in unsupervised feature learning and deep learning in order to automatically learn the complex dependencies between physiological parameters and blood glucose.

6 Summary and Conclusion

Hypoglycemia is a major health and safety concern for people with type 1 diabetes. Continuous glucose monitoring systems aid in hypoglycemia detection, but have drawbacks, employing invasive sensors with inherent noise. This paper reports on preliminary machine learning experiments that aim to improve hypoglycemia detection by leveraging data from noninvasive physiological sensors found in commercially available fitness bands. A middle-aged subject with T1D provided CGM, fitness band, insulin, and life-event data, under normal daily living conditions, for two months. The fitness band outputs sensor data for continuous heart rate, galvanic skin response, skin temperature, and air temperature.

Statistical tests identified 36 features derived from this sensor data that could differentiate hypoglycemic from non-hypoglycemic states. Using these features, support vector machines were trained to classify instances as hypoglycemic (positive) or non-hypoglycemic (negative). In an initial experiment, an SVM with a linear kernel using all 36 features was able to outperform two simple baselines. In an experi-

ment designed to alleviate overfitting, the number of features was narrowed down using greedy feature selection. Results show proof of concept that physiological sensor data from fitness bands can provide discriminative features for hypoglycemia detection. However, system performance was limited by the size and nature of the dataset.

The results are being utilized in ongoing work to improve the performance of overall blood glucose prediction models that also use CGM, insulin, and life-event data. In addition to incorporating new features derived from noninvasive physiological sensors in these models, we also plan to explore the automatic discovery of physiological dependencies by leveraging unsupervised feature learning and deep learning algorithms. Finally, to help alleviate the lack of available patient data, we plan to collect and de-identify blood glucose, insulin, life-event and physiological sensor data from up to 30 additional patients. We look forward to sharing this data with other researchers interested in solving this difficult and clinically important problem.

7 Acknowledgements

This work was supported by the National Science Foundation under Grant # 1117489. The authors would like to acknowledge the contributions and support of SmartHealth Lab members Babak Bokaie, Quintin Fettes, Charlie Murphy, Kevin Plis, Hannah Quillin, and Yi Yu.

References

[Boutayeb and Chetouani, 2006] A. Boutayeb and A. Chetouani. A critical review of mathematical models and data used in diabetology. *BioMedical Engineering OnLine*, 5(43), 2006.

[Bunescu *et al.*, 2013] R. Bunescu, N. Struble, C. Marling, J. Shubrook, and F. Schwartz. Blood glucose level prediction using physiological models and support vector regression. In *Proceedings of the Twelfth International Conference on Machine Learning and Applications (ICMLA)*, pages 135–140. IEEE Press, 2013.

[Chang and Lin, 2001] C.-C. Chang and C.-J. Lin. *LIBSVM: A library for support vector machines*, 2001. Software available at http://www.csie.ntu.edu.tw/ ˜cjlin/libsvm, accessed April, 2016.

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

[Jensen *et al.*, 2013] 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 & Therapeutics*, 15(7), 2013.

[Kovatchev *et al.*, 2009] B. P. Kovatchev, M. Breton, C. DallaMan, and Claudio C. Cobelli. In silico preclinical trials: A proof of concept in closed-loop control of type 1 diabetes. *Journal of Diabetes Science and Technology*, 3(1):44–55, 2009.

[Lehmann and Deutsch, 1992] E. D. Lehmann and T. Deutsch. A physiological model of glucose-insulin interaction in type 1 diabetes mellitus. *Journal of Biomedical Engineering*, 14(3):235–242, 1992.

[Marling *et al.*, 2011] C. R. Marling, J. H. Shubrook, S. J. Vernier, M. T. Wiley, and F. L. Schwartz. Characterizing blood glucose variability using new metrics with continuous glucose monitoring data. *Journal of Diabetes Science and Technology*, 5(4):871–878, 2011.

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

[Plis *et al.*, 2014] K. Plis, R. Bunescu, C. Marling, J. Shubrook, and F. Schwartz. A machine learning approach to predicting blood glucose levels for diabetes management. In *Modern Artificial Intelligence for Health Analytics: Papers Presented at the Twenty-Eighth AAAI Conference on Artificial Intelligence*, pages 35–39. AAAI Press, 2014.

[Schechter *et al.*, 2012] A. Schechter, O. Eyal, N. Zuckerman-Levin, V. Amihai-Ben-Yaacov, N. Weintrob, and N. Shehadeh. A prototype of a new noninvasive device to detect nocturnal hypoglycemia in adolescents with type 1 diabetes: A pilot study. *Diabetes Technology & Therapeutics*, 14(8):683–689, 2012.

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

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

[Wang *et al.*, 2014] Q. Wang, P. Molenaar, S. Harsh, K. Freeman, J. Xie, C. Gold, M. Rovine, and J. Ulbrecht. 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, 2014.

[Zecchin *et al.*, 2013] 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*, 15(1):66–77, 2013.