Blood Glucose Level Prediction based on Support Vector Regression using Mobile Platforms*

Maximilian P. Reymann, Eva Dorschky, Benjamin H. Groh, Christine Martindale, Peter Blank and Bjoern M. Eskofier[1], *Member IEEE, EMBS*

***Abstract*— The correct treatment of diabetes is vital to a patient's health: Staying within defined blood glucose levels prevents dangerous short- and long-term effects on the body. Mobile devices informing patients about their future blood glucose levels could enable them to take counter-measures to prevent hypo or hyper periods. Previous work addressed this challenge by predicting the blood glucose levels using regression models. However, these approaches required a physiological model, representing the human body's response to insulin and glucose intake, or are not directly applicable to mobile platforms (smart phones, tablets). In this paper, we propose an algorithm for mobile platforms to predict blood glucose levels without the need for a physiological model. Using an online software simulator program, we trained a Support Vector Regression (SVR) model and exported the parameter settings to our mobile platform. The prediction accuracy of our mobile platform was evaluated with pre-recorded data of a type 1 diabetes patient. The blood glucose level was predicted with an error of 19 % compared to the true value. Considering the permitted error of commercially used devices of 15 %, our algorithm is the basis for further development of mobile prediction algorithms.**

I. INTRODUCTION

Diabetes is one of the most prevalent diseases worldwide. The estimated number of people suffering from diabetes is currently 415 million people, but it is assumed to increase to 642 million by 2040 [1]. Patients try to avoid hypoglycemia (low blood sugar) and hyperglycemia (high blood sugar), both of which have dangerous short- and long-term effects on the body. The short-term effects can include nausea, fever and coma, whereas long-term effects include heart attack, and stroke [2]–[4]. It is vital that patients monitor their blood sugar levels closely to avoid these risks. Traditionally, patients measured their blood glucose level (BGL) with a finger stick 3-8 times a day to adjust their insulin levels according to their carbohydrate intake. In recent years, the development of Continuous Glucose Monitoring (CGM) systems has been pushed forward resulting in a variety of devices on the consumer market[2]. A CGM system usually consists of a sensor attached to the skin and an accompanying reader displaying the curve of the BGL. Because reaction times of the BGLs are delayed to the glucose intake, knowing how the BGL of the human body will develop is a key factor for preventing hypo or hyper events for diabetes patients. We define hypo and hyper events as periods in which the BGL crosses the thresholds as issued by the American Diabetes Association. Hypoglycemia in hospitalized patients has been defined as blood glucose $< 70\,\frac{mg}{dL}$, and hyperglycemia as blood glucose $> 140\,\frac{mg}{dL}$ [5]. If the BGLs were known beforehand, patients could take the required counter-measures to prevent these events. Consequently, they could achieve staying in a BGL area, which is not harmful to the body and thus reduce the immediate and long-term effects of diabetes. Bunescu et al. [6] described a method to predict the blood sugar for individual patients using physiological models. These models required up to seven specific input features to make reliable predicitions. Additionally, in their paper they did not describe the suitability of their algorithm for mobile platforms. Further research in this field led to the conclusion, that Support Vector Regression is suitable for the prediction of BGLs [6]–[9]. Chemlal et al. [8] already implemented a BGL prediction for mobile platforms, in this case as an iPhone application. However, data about glucose intake information had to be entered by the user manually for every meal the patient consumed.

Those approaches are missing an unobtrusive mobile application which does not require any active input by the user and provide a warning in case certain thresholds for BGLs are exceeded. Another approach was proposed by Breton et al. [10]. They developed a closed-loop system, where the user is not required to inject oneself with insulin anymore. Consisting of a CGM device and an insulin pump, this system is expected to regulate the glucose balance in the body autonomously over night. However, this system was aimed at type 1 patients, that require a constant injection of insulin, and is thus not usable for patients with type 2 diabetes not requiring an insulin pump. Hence, all aforementioned publications aimed for an accurate prediction of BGL but were limited to requiring a physiological model and in addition, they did not contain an unobtrusive mobile implementation.

In this work, we present an algorithm for mobile platforms to predict the BGL of a patient with a CGM device. It is based on the fact that patients already using a CGM system have all the data necessary to make a prediction accessible on their device. The algorithm can easily be integrated into existing CGM readers or process the data transmitted by a CGM sensor. Consequently, the predicted BGL is analyzed and the patient is warned if he is about to enter a hypo or

*This work was not supported by any organization

[1]Digital Sports Group, Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany (e-mail: maximilian.reymann@fau.de)

[2]DexCom™G4 PLATINUM®, Abbott FreeStyle Navigator, GlucoDay® S, Medtronic MGuardian® REAL-Time System and MiniMed Paradigm®

978-1-4577-0220-4/16/$31.00 ©2016 IEEE

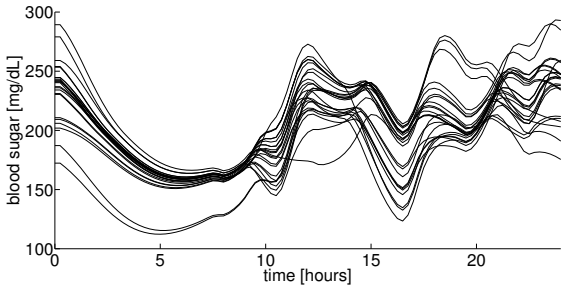

Fig. 1. Example BGL data for one subject over 25 days, simulated with the online simulation tool 2aida. Every black line represents one day.

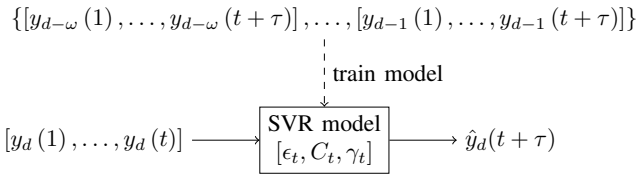

Fig. 2. Structure of the training and testing set. For every estimation, the SVR model uses trainings data $y_{d-\omega},\ldots,y_{d-1}$, testing data y_d and a specific set of regression parameters $\epsilon_t, C_t, \gamma_t$ to generate a individual model and predict the BGL $\hat{y}_d$.

hyper period. The algorithm was designed with data of an online simulation tool, followed by an evaluation with real patient data on a mobile platform. We compared the results to the outcomes of other papers and discuss them regarding their reliability.

II. Methods

A. Data acquisition

Data were acquired from different sources: an online freeware diabetes software simulator program (www.2aida.net) and a CGM system incorporated in this study. For developing the prediction algorithm, the online simulation program was used. Data of five different subjects were employed to evaluate the algorithm. For each patient, 25 days of BGLs were simulated (Fig. 1). The simulation produced BGL curves according to carbohydrate intakes and insulin shots defined over the day. Here, values for typical glucose intake (45 to 60 grams per meal [11]) have been entered to obtain 25 different curves, representing different days. The sampling rate was given by one reading every 15 minutes.

The second source used for the mobile evaluation was anonymized readings from a FreeStyle Libre® CGM system of a diabetes type 1 patient which were handed over in retrospect and on a voluntary basis. Hereby, all experimental procedures followed the principles outlined in the Declaration of Helsinki. The blood glucose sensor was attached to the disinfected skin on the upper arm of the patient, being able to store up to eight hours data of the subcutaneous glucose level. When placing the reader near the sensor, the data from the past 8 hours were transmitted to the reader. The sensor did not require any further calibration and could stay on the patient up to 14 days. In total, the CGM data amounted to 4635 readings over a period of roughly 35 days.

B. Blood glucose level prediction

The algortihm was designed with the simulated patient data in Matlab® R2013a. The open source implementation of Support Vector Machines libSVM was used [12]. Data from each patient were split into testing and training data sets. The BGL on day d at the time t was given by $y_d(t)$. The window size was chosen to as ω, and τ represented the prediction time. Training data were all the BGLs in the training window ω up to the day d of the prediction. Subsequently, testing data were the BGLs on day d until the time t. For a given window size, the testing and training set was structured as depicted in Fig. 2. The feature vector for the training and testing sets consisted of data of one day up to the time t $[y_d(1),\ldots,y_d(t)]$ and the label vector was given by $y_d(t+\tau)$. Respectively, we trained one SVR-model for every prediction time $t+\tau$ during one day. We considered the prediction times τ: 30 min and 60 min for later evaluation.

The data sets were scaled to a range from zero to one to avoid attributes in greater numeric ranges dominating those in smaller ranges [13]. The parameter selection resulted in the following parameters:

- Regression type epsilon SVR
- Radial basis function (RBF) as kernel
- ϵ set to default value of 0.001
- (Kernel coefficient) γ set to a range from $[2^{-15}, 2^{-14}, \ldots, 2^{3}]$
- (Penalty parameter) C set to a range from $[2^{-5}, 2^{-4}, \ldots, 2^{15}]$

Following a five-fold cross validation on the training data set, we performed a grid search using the parameters C and γ to find the smallest error. Those parameters were then applied for training the model used for the prediction of $\hat{y}_d$ based on the testing set.

C. Mobile implementation

To demonstrate the feasibility of the mobile application, the Support Vector Regression was evaluated as real time application on a mobile platform. To account for the irregular distributed patient data of the CGM reader, cubic spline interpolation, resulting in a constant sample rate of 1 reading every 5 minutes, was applied [14]. The computation of the prediction was equal to the implementation described above. We implemented the app with Android Studio for Android 4.4.. Patient data were available as a time series with one BGL value for each specific point in time. Reducing the amount of computation was achieved by using the preselected parameters of the simulated study. Each point of a day had a constant parameter set $[\epsilon_t, C_t, \gamma_t]$, which was exported to the app and used for the SVR model. A graphical output was used to display the last 12 hours of BGL data and the predicted value $\hat{y}_d(t+\tau)$. The model was updated in 15 minute intervals, corresponding to the sampling rate produced by the reader. A notification was displayed, when the predicted value exceeded a certain threshold. For hypo

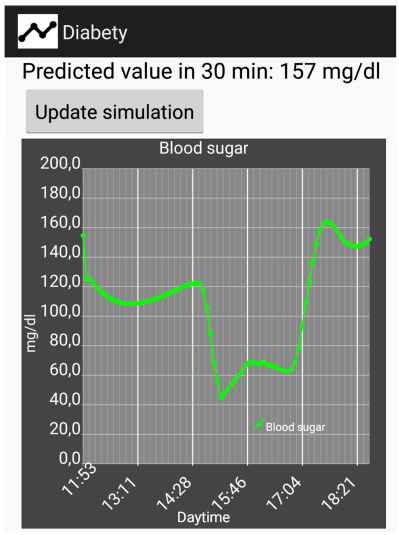

Fig. 3. Screen shot of the mobile application "Diabety".

events, the user received a notification "Your blood sugar is getting low ($< 90 \frac{mg}{dL}$) in the next 30 minutes. Please eat or drink something now!". Values for this notification have to be set individually in coordination with a physician. For the purpose of demonstration we used the recommedation for adolescents and young adults of the American Diabetes Association [3]. In addition to the notification function, we incorporated an alarm, which was planned to be activated if the patient had a hypo event during the night.

D. Evaluation

We visually checked the blood glucose data from the online simulation tool for a realistic trend. Unrealistic outcomes, for instance values under $< 30 \frac{mg}{dL}$ and over $> 400 \frac{mg}{dL}$ of the online simulation tool were not used and different days were simulated. The accuracy of our algorithm was evaluated for the online data base and data of the diabetes patient. The prediction of the BGL for $\hat{y}$ resulted in a difference to the actual value at y. As evaluation measure, we used the arithmetic mean of the relative error over all samples of one day T:

$$\delta(d) = \frac{1}{T} \sum_{t=1}^{T} \frac{|\hat{y}_d(t+\tau) - y_d(t+\tau)|}{y_d(t+\tau)}. \quad (1)$$

We evaluated different window sizes to find a reasonable training data size for our mobile implementation. The accuracy of the algorithm on the mobile platform was tested on a OnePlus One smart phone running Android 5.1.1. Here, we simulated different points over one day and considered the mean of the computation time as result.

III. Results

Simulation results showed, that the error for the prediction decreased when increasing the window size ω (see Fig. 4). Window sizes were gradually increased from the minimum required input data of 5 days to all 21 available training days.

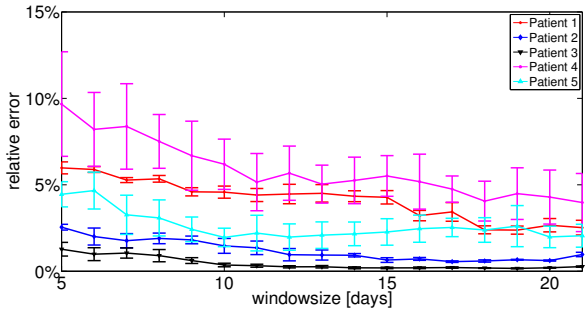

Fig. 4. Results of the support vector regression for a 30 minute ahead prediction. Different colors represent the different patients. The different curves share the property of converging to a relative error δ of 0.2-4 %.

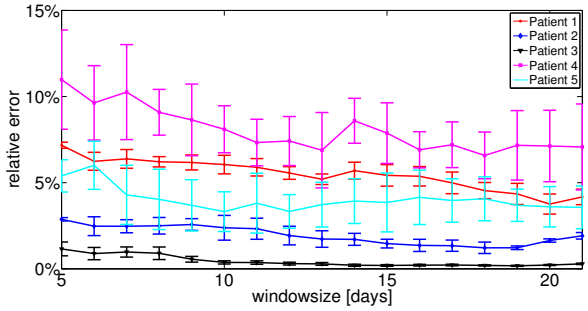

Fig. 5. Results of the support vector regression for a 60 minute ahead prediction. Different colors represent the different patients. The different curves share the property of converging to a relative error δ of 0.3-7 %.

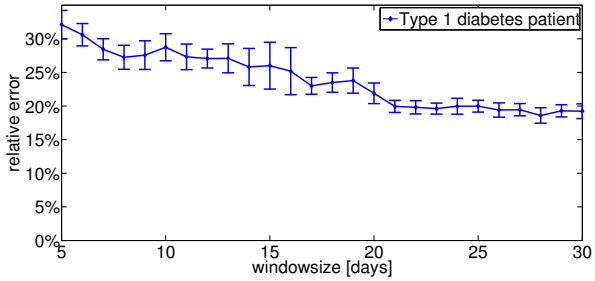

Fig. 6. Results of the support vector regression for a 30 minute ahead prediction using real patient data of a TD1 patient. After a windowsize ω of about 21, it is visible, that the relative error ω stays at about 19 %.

The error decreased for all considered prediction times τ, whereby greater values for τ showed a higher relative error δ (compare Fig. 4 and 5). For a 30 min ahead prediction, we were able to show that the error converges to 0.2-4 % after increasing the window size gradually up to 21 days. Relative errors δ were distinct for every patient, however shared the same converging properties. The relative error on the diabetes patient data for a 30 min prediction converged after 21 days to a δ of 19.8 %. Computation of the prediction on the mobile platform took between 0.3 seconds and 8 seconds for a training window size of 5 and 21 days, respectively.

IV. Discussion

Comparing the results of our algorithm performance on the simulated data to the real patient data, a difference is noticeable. This is due to the small variance of the simulated

data in contrast to the real patient data set. The authors of the 2aida tool mention, that it is meant for education / teaching / demonstration use only, because it cannot account for the complexity of the human glucoregulatory system. For the purpose of this paper the data can be seen as adequate, since the algorithm was only trained on these data. The evaluation led to an error of 19.8 %. However, it has to be considered that the permitted error of finger stick glucose sensors is 15 %, as issued by the FDA [15]. Hence, our prediction can be used to further optimize mobile prediction algorithms. Even more accurate results could be achieved by improving the accuracy on larger sets of patient data, i.e. all past data measured by a CGM device of a patient. Moreover, the algorithm can be improved by optimizing it regarding regression models for prediction on time series [16]. According to the research of Chemlal et al. [8], integrating activity data of the user can improve accuracy and detect changes in the BGL. Additionally, for future applications we suggest integrating an automated method for BGL verification.

The mobile computation was time efficient, on average taking 95 ms for computation. In the real application, the algorithm is effectively running as a background activity and has to be updated every 5-10 minutes [14]. Besides contributing to the physiological effects of diabetes, a mobile application can also contribute to the quality of life for patients. The use of CGM systems allows patients to keep track of their BGL without picking their finger. Checking their BGL, can now be done by looking at their phone. Additionally the implementation on a smart phone enables the further implementation of various functions such as:

1) Notification function
2) Alarm during the night
3) Sharing their BGLs with their parents (for children)
4) Exporting the data and sending it to their physician

An application like the one presented in this paper will be very valuable, for patients using a CGM device. Further development will make these systems cheaper and thus accessible to more patients.

V. Conclusion

Previous attempts at predicting the blood sugar required the use of a physiological model or were not designed for mobile platforms. We demonstrated, that it is possible to perform a prediction using SVR on a mobile platform. Based on simulations on the real patient data, we concluded that a window size of 21 days is a sufficient training data size to obtain a prediction value capable of describing BGL development.

This work can be the basis for further development. After living with diabetes for a long time, patients become dull to the effects their body displays under low blood sugar. This can lead to unforeseen change of consciousness. Our app can help to prevent hypo and hyper periods, because it could detect such situations even before the drop or increase in BGL has fully started. Due to the fact that reaction times of the BGLs are delayed to the intake of glucose, our application is useful to sustain a stable BGL. The next step is evaluating our app with live CGM patient data and making the algorithm more robust. Additionally, the ideal use case would be one app supporting all sensors on the market. Moreover, the results of this paper can also be the basis for optimizing closed-loop systems and providing information to the insulin pump.

Acknowledgment

The authors would like to thank the diabetes patient for providing his BGL data, and we thank the Innovation Research Lab team at FAU for their contribution to this paper. Furthermore, we would like to thank Dexcom for supplying us with a sensor to record data for this project.

References

[1] International Diabetes Federation, *IDF Diabetes Atlas*, vol. 7 ed. Brussels, Belgium: International Diabetes Federation, 2015.

[2] N. W. Salomon, U. S. Page, J. E. Okies, J. Stephens, A. H. Krause, and J. C. Bigelow, "Diabetes mellitus and coronary artery bypass. Short-term risk and long-term prognosis.," *The Journal of thoracic and cardiovascular surgery*, vol. 85, pp. 264–71, Feb 1983.

[3] American Diabetes Association, "Standards of Medical Care in Diabetes - 2016," *The Journal of Clinical and Applied Research and Education*, vol. 39, pp. 14–80, Jan 2016.

[4] Diabetes Digital Media Ltd, "Short Term Complications - Hypoglycemia, Ketoacidosis & HHS," 2016.

[5] E. R. Seaquist, J. Anderson, B. Childs, P. Cryer, S. Dagogo-Jack, L. Fish, S. R. Heller, H. Rodriguez, J. Rosenzweig, and R. Vigersky, "Hypoglycemia and diabetes: a report of a workgroup of the American Diabetes Association and the Endocrine Society," *Diabetes care*, vol. 36, pp. 1384–95, May 2013.

[6] R. Bunescu, N. Struble, C. Marling, J. Shubrook, and F. Schwartz, "Blood Glucose Level Prediction Using Physiological Models and Support Vector Regression," in *12th International Conference on Machine Learning and Applications*, vol. 1, pp. 135–140, 2013.

[7] E. I. Georga, V. C. Protopappas, D. Ardigò, M. Marina, I. Zavaroni, D. Polyzos, and D. I. Fotiadis, "Multivariate Prediction of Subcutaneous Glucose Concentration in Type 1 Diabetes Patients Based on Support Vector Regression," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, pp. 71–81, Jan 2013.

[8] S. Chemlal, S. Colberg, M. Satin-Smith, E. Gyuricsko, T. Hubbard, M. W. Scerbo, and F. D. McKenzie, "Blood glucose individualized prediction for type 2 diabetes using iPhone application," *IEEE 37th Annual Northeast Bioengineering Conference*, pp. 1–2, 2011.

[9] K. Plis, R. Bunescu, C. Marling, J. Shubrook, and F. Schwartz, "A Machine Learning Approach to Predicting Blood Glucose Levels for Diabetes Management," *Modern Artifical Intelligence for Health Analytics*, pp. 35–39, Mar 2014.

[10] M. Breton and A. Farret, "Fully integrated artificial pancreas in type 1 diabetes: modular closed-loop glucose control maintains near normoglycemia.," *Diabetes*, vol. 61, pp. 2230–2237, Sep 2012.

[11] American Diabetes Association, "Understanding carbohydrates," Aug 2013. URL: http://www.diabetes.org/food-and-fitness/food/what-can-i-eat/understanding-carbohydrates/sugar-and-desserts.html.

[12] C.-C. Chang and C.-J. Lin, "Libsvm : a library for support vector machines.," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.

[13] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, "A Practical Guide to Support Vector Classification," *BJU international*, vol. 101, no. 1, pp. 1396–400, 2008.

[14] D. A. Gough, K. Kreutz-Delgado, and T. M. Bremer, "Frequency characterization of blood glucose dynamics," *Annals of Biomedical Engineering*, vol. 31, no. 1, pp. 91–97, 2003.

[15] FDA, "Self-Monitoring Blood Glucose Test Systems for Over-the-Counter Use Draft Guidance for Industry and Food and Drug Administration Staff," 2014.

[16] K. Lin, Q. Lin, C. Zhou, and J. Yao, "Time Series Prediction Based on Linear Regression and SVR," vol. 1, no. 2006, pp. 7–10, 2007.