Towards Wearable-based Hypoglycemia Detection and Warning in Diabetes

Martin Maritsch
ETH Zurich
Zurich, Switzerland

Vera Lehmann
Inselspital, Bern University Hospital
Bern, Switzerland

Mathias Kraus
ETH Zurich
Zurich, Switzerland

Tobias Kowatsch
University of St. Gallen
ETH Zurich
St. Gallen & Zurich, Switzerland

Christoph Stettler
Inselspital, Bern University Hospital
Bern, Switzerland

Felix Wortmann
University of St. Gallen
St. Gallen, Switzerland

Simon Föll
Karlsruhe Institute of Technology
Karlsruhe, Germany

Caterina Bérubé
ETH Zurich
Zurich, Switzerland

Stefan Feuerriegel
ETH Zurich
Zurich, Switzerland

Thomas Züger
Inselspital, Bern University Hospital
Bern, Switzerland

Elgar Fleisch
ETH Zurich
University of St. Gallen
Zurich & St. Gallen, Switzerland

CHI '20 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6819-3/20/04.
http://dx.doi.org/10.1145/3334480.3382808

Abstract
Rigorous blood glucose management is vital for individuals with diabetes to prevent states of too low blood glucose (hypoglycemia). While there are continuous glucose monitors available, they are expensive and not available for many patients. Related work suggests a correlation between the blood glucose level and physiological measures, such as heart rate variability. We therefore propose a machine learning model to detect hypoglycemia on basis of data from smartwatch sensors gathered in a proof-of-concept study. In further work, we want to integrate our model in wearables and warn individuals with diabetes of possible hypoglycemia. However, presenting just the detection output alone might be confusing to a patient especially if it is a false positive result. We thus use SHAP (SHapley Additive exPlanations) values for feature attribution and a method for subsequently explaining the model decision in a comprehensible way on smartwatches.

Author Keywords
diabetes; hypoglycemia detection; wearables; machine learning; explainable artificial intelligence; SHAP values.

CCS Concepts
•Applied computing → **Consumer health; •Human-centered computing** → *Ubiquitous and mobile computing; Visualization;*

Figure 1: A person with an invasive, continuous glucose monitor sensor in their upper arm and its corresponding reader.

Introduction

Diabetes mellitus is a common metabolic disorder, which is characterized by elevated blood glucose (BG) levels (hyperglycemia). It affects approximately 460 million people worldwide and a near doubling of these cases is expected within the next 10 to 25 years. Roughly 10% (USD 760 billion) of global health expenditure is spent on diabetes worldwide every year [3].

Type 1 diabetes mellitus (T1DM) is a chronic autoimmune disease and accounts for approximately 10% of all diabetes mellitus cases. T1DM is characterized by an absolute insulin deficiency and, therefore, requires insulin replacement to keep BG in a normal range. Consensus Guidelines specify a target BG range of 3.9–10mmol/L (70–180 mg/dL) [1]. Depending on the medical situation and the patient preference, insulin therapy is either carried out with an insulin pump or with multiple daily injections. However, exogenous insulin therapy does not exactly replicate the normal physiological insulin secretion and, therefore, bears the risk of insulin over- and underdosing leading to states of dysglycemia (too low or too high BG).

Despite ongoing developments in the treatment of diabetes mellitus, low BG (hypoglycemia) remains one of the most relevant, acute, and disabling complications in those patients. While severe hypoglycemia may lead to heart arrhythmia, seizures, coma, or even death [17], its mild form impairs a variety of physical and psychical functions [5].

Since dysglycemia is associated with severe short- and long-term complications, a rigorous BG self-measurement is inevitable for individuals with T1DM. Conventionally, BG is measured with a fingerstick test, which requires piercing the skin to retrieve a sample of capillary blood. Besides, glucose can also be monitored using a continuous glucose monitor, which estimates the actual BG via a small sensor in the subcutaneous tissue under the skin. Such an exemplary device is depicted in Figure 1. However, these monitoring devices can impose a medical and financial burden, ranging from the inability of proper handling to the lack of reimbursement as well as individual factors (e.g., allergic reaction to the adhesive of the device).

Heart Rate Variability and Diabetes

Heart rate variability (HRV) reflects the adaptions of the human heart to abrupt physiological changes. It gives insight into involuntary and subconscious functions as it measures the interactions between the sympathetic and parasympathetic divisions of the autonomic nervous system. HRV shows individual responses to physical exercise, psychophysiological stress, and heart diseases [16].

For diabetes, changes in HRV measures have been associated with hypoglycemic episodes [14]. Also, short-time measurements (i.e., 5 minutes) of HRV have been shown to correlate with hypoglycemia [14]. To reach a wider audience in everyday life, we aim to use broadly available consumer wearable devices such as smartwatches instead of professional electrocardiographies (ECGs) to measure HRV. The study in [11] shows that data from smartwatches was highly accurate when compared to professional ECGs in long-term measurements.

Wearable Health Computing

Wearables are often equipped with multiple sensors to collect physiological data. For example, there are wrist-based consumer smartwatches which provide data such as interbeat intervals recorded by an optical sensor (see Figure 2), three-axis accelerometer data, step count, burned calories, or proprietary stress values [6].

Additionally, such wearables are – compared to professional medical equipment – more easily accessible to a broad au-

Figure 2: Back view of a consumer smartwatch with its optical pulse sensor active (green LEDs).

dience and also suitable for ubiquitous use. Using such common and low-priced alternatives to medical equipment could mean an improvement of existing diabetes therapies and self-management applications [12].

Objective

This work demonstrates our research-in-progress on the potential of applying methods of machine learning to detect hypoglycemic states among individuals with T1DM based on smartwatch sensor data. Furthermore, we aim to use approaches of explainable artificial intelligence (AI) to make the classification output comprehensible to individuals with diabetes in everyday life. For this purpose, we propose a smartwatch-based hypoglycemia warning interface which explains the underlying model decision-making in a comprehensible way to individuals with diabetes.

Methods

Heart Rate Variability

HRV features are calculated from an inter-beat interval sequence, which is a series of time distances between normal-to-normal heartbeat intervals. HRV can be measured in both the time and frequency domain [16]. In the time domain, measures such as root mean square of the successive differences (RMSSD), standard deviation of normal-to-normal interval (SDNN), number of successive normal-to-normal interval differences exceeding 50ms (NN50), percentage of successive normal-to-normal interval differences exceeding 50ms (pNN50), median normal-to-normal interval (medianNNI), coefficient of variation of normal-to-normal intervals (CVNNI), and coefficient of variation of successive differences (CVSD) are prevailing. In the frequency domain measures of high frequency (HF), low frequency (LF), very-low frequency (VLF), ultra-low frequency (ULF), and the ratio of LF to HF (LF/HF-ratio) are commonly used (see [15] for an overview).

According to the Task Force of the European Society of Cardiology [16], short-term recordings of 2 to 5 minutes should be assessed with SDNN and RMSSD in the time domain and HF, LF, and VLF in the frequency domain. RMSSD has also been found to be reliable when calculated with a sample of 60 seconds, which was not always the case for SDNN or frequency-based measures [9]. Ultra-short-term measures of HRV such as RMSSD would thus allow monitoring streams of physiological changes with relatively high resolution and short-term adaption capability.

Classification and Feature Attribution

yExplainability is one of the major urges of machine learning applications in healthcare. Thus, we particularly focused our research on models and algorithms, of which the outputs are comprehensible and we refrained from using complex deep neural network architectures. We applied different kinds of machine learning models ranging from logistic regression to decision trees for binary classification. Finally, we focused on building a model based on a gradient boosting decision tree (GBDT) [4]. An undoubted advantage of comparatively simple models such as decision trees is, besides their easier comprehensibility, that they can also be employed in resource-constrained computing environment such as smartwatches.

With increasing model complexity, the correct interpretation of the model output gets inherently difficult. However, we want to address the question of why a model makes certain classifications and which features these decisions are based upon. Therefore we leverage SHAP (SHapley Additive exPlanations) values, which have already successfully been applied in a medical research context [7]. SHAP values assign a feature attribution value to each observation and class. They thus explain the impact of each feature on the model output. These values are relevant for deriving

Domain	Feature
time	RMSSD
	SDNN
	NN50
	pNN50
	medianNNI
	CVNNI
	CVSD
frequency	VLF
	LF
	HF
	LF/HF-ratio

Table 1: Generated HRV features.

Metric	Mean	SD
AUC	0.938	0.068
accuracy	82.7%	13.9%
sensitivity	76.7%	25.4%
specificity	84.2%	19.7%

SD = standard deviation

Table 2: Results of the 10-fold cross validation of a simple GBDT classifier for hypoglycemia.

physiological conditions from the model output.

Data collection

We conducted a proof-of-concept study with one otherwise healthy individual with T1DM. Data was collected in a naturalistic setting over a period of one week. Physiological data by means of inter-beat intervals was recorded with an *Empatica E4* smartwatch. Glucose data was obtained via a continuous glucose monitor, the *FreeStyle Libre*, which provides a measurement roughly every 15 minutes. The subject continued with their BG management as usual during the study period. Thus, the recorded hypoglycemic phases represent those, which were not recognized or not proactively prevented by the subject.

During pre-processing, we derived the HRV features listed in Table 1 from the inter-beat interval sequence. Further, statistical heart rate features (minimum, maximum, mean, standard deviation) were calculated. For feature calculation, we used a sliding window approach with a window size of 180 seconds and a step size of 1 second. BG measurements were re-sampled by piece-wise linear interpolation to match the frequency of physiological data. The data cleaning and pre-processing resulted in a total of 74,552 observations of which 15,168 (20.4%) belong to the positive class (hypoglycemia).

Results

Hypoglycemia Detection Model

The best performing model was based on a GBDT. The model input are the heart rate and HRV features and the classification task is defined as a binary decision between normal BG levels (negative) and hypoglycemia (positive). For our analysis, we define hypoglycemia as observations with a BG level of <3.9mmol/L according to [1].

We ran a stratified 10-fold cross-validation on the dataset. The trained GBDT model was able to classify the observations in the test set with a mean accuracy of 82.7% and a mean area under curve (AUC) of 0.938. See Table 2 for further result metrics.

Feature Attribution

The SHAP explainer was run on the model from one of the cross validation folds. Figure 3 shows an exemplary SHAP dependency plot for the features *RMSSD* (along x-axis) and *LF/HF-ratio* (dot coloring), which are known to have a relationship with hypoglycemia [10]. A higher SHAP value (y-axis) in the graph corresponds to a higher probability of classification as hypoglycemia. The figure thus shows an inverse relationship between *RMSSD* and the probability of classifying hypoglycemia. The plot further shows the interaction with the frequency-domain based *LF/HF-ratio* feature and reveals that an increased feature value corresponds to an increased probability of classification as hypoglycemia.

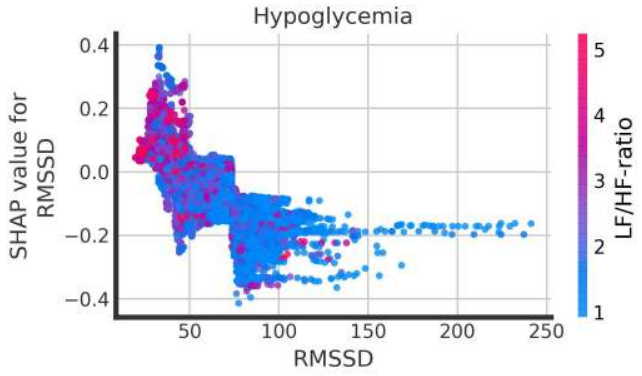

Figure 3: SHAP dependency plot visualizing the impact of *RMSSD* and *min HR* feature values on the hypoglycemia probability output.

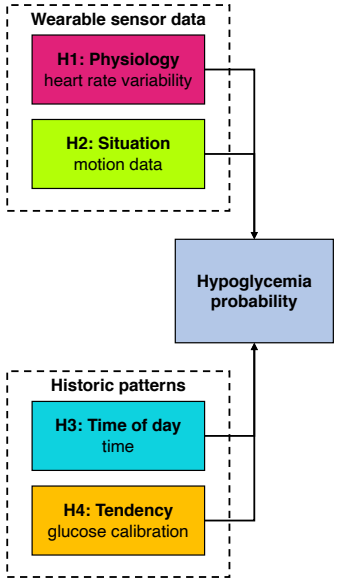

Figure 4: Four hypothesized cause-effect-relationships for hypoglycemic BG levels in T1DM and the data which is used to assess them.

Discussion

Our limited proof of-concept study already showed promising results that it is feasible to detect hypoglycemic conditions based on smartwatch-based sensor data. Furthermore, we are convinced that we can enhance the detection on the basis of additional data and well-known cause-effect relationships.

Potential Cause-Effect Relationships in Hypoglycemia Detection

Our ongoing research activities are based on a comprehensive framework for hypoglycemia detection which is summarized in Figure 4. The presented work leads us to the conclusion, that physiological data can indeed be used to infer hypoglycemic phases (**H1**). The SHAP dependency plot in Figure 3 shows that lower RMSSD values correspond to a higher probability of the model classifying an observation as hypoglycemia. This observation is in line with a prior study, stating that the RMSSD decreases during hypoglycemic phases [10]. Interestingly, the GBDT model was able to find and learn the relationship between RMSSD and hypoglycemia autonomously.

Furthermore, we will investigate three additional, potential cause-effect relationships for hypoglycemia detection in individuals with T1DM (**H2**, **H3**, **H4**). The underlying fundamental principles of all four cause-effect relationships (**H1**–**H4**) have been investigated individually by related work. These subsequently briefly discussed relationships are assessable with the capabilities of today's consumer smartwatches.

H1: Physiology As discussed earlier, altered HRV values such as a lower RMSSD can give an indication of hypoglycemic states [10]. The HRV features relevant for physiological detection can be computed from the inter-beat interval sequence recorded by a smartwatch. Furthermore, stress assessment in the form of a discrete value is already today a common feature in consumer smartwatches.

H2: Situation The current posture and activity situation of a person is a further profound hypoglycemia indicator. Lying positions like resting on a couch watching TV or sleeping over a prolonged period bear the risk of reduced hypoglycemia awareness and therefore may increase the risk of severe and prolonged hypoglycemia [13]. Moreover, being physically active over an extended period of time, leads to a higher likeliness of hypoglycemia as well [18]. Therefore we aim to detect the posture and activity situation based on accelerometer data of a smartwatch to infer the risk of hypoglycemia.

H3: Time of day The time of the day gives a general indication of hypoglycemia probability [2]. The BG level of individuals with T1DM tends to adhere to a personal, daily BG pattern for any given day. This pattern can be learned over time and personalize the influence of the time of day on hypoglycemia warnings.

H4: Tendency There is a general tendency of the BG level behavior over the day which can be inferred from the fasting BG level in the morning [8]. We thus propose, analogous to some continuous glucose monitors which need to be calibrated regularly, to provide the detection system with the actual BG level in the morning. The relevant data for determining the tendency is a calibration of the algorithm given by a measurement of fasting glucose in the morning. This potential relationship seems particularly promising for hypoglycemia detection together with the time of day.

Explaining Model Outputs to Individuals with Diabetes

Once the model is implemented on a smartwatch and yields a positive classification result (hypoglycemia), we want to warn the user. However, the model will not be completely

Figure 5: A mock-up of the smartwatch user interface for the decision explanation.

accurate in its detection. It would aim for a trade-off with high sensitivity (recall) and prefer to warn the user too often instead of too seldom. Consequently, the model output might sometimes be confusing to the user, especially if it is a false positive result. We thus aim to explain the model output to the user and let them know, how the model came to its conclusion in case of a warning.

Furthermore, too frequent warnings may as well have the negative impact of the user ignoring them. Thus, we will distinguish between detected states of hypoglycemia and severe hypoglycemia (BG level of <3.0mmol/L [1]). Since severe hypoglycemia is accompanied by more perceptible symptoms, we expect future models to recognize those critical phases even better. While for mild hypoglycemia the warning should be adjustable to the user preference, the warning for severe hypoglycemia should be presented in a more assertive way.

We aim to employ SHAP values for real-time feature attribution of the model output. In the case of a positive classification (hypoglycemia), we will transfer these feature attributions to estimate the influence of each of our four hypothesized cause-effect-relationships on the hypoglycemic state. Along with a hypoglycemia warning, the estimated influences will then be visualized to the user in a comprehensible way on the smartwatch display.

Figure 5 shows a mock-up for a smartwatch user interface explaining the cause-effect relationships to an individual with diabetes. In this particular case, the model warns of a low BG level based on multiple factors, which could for example be based on the person lying on a couch watching TV. First, indicated by the almost closed violet ring, the physiological state of the person is an important contribution to detecting hypoglycemia (**H1**). This could indicate altered HRV values such as a lowered RMSSD. Furthermore, indicated by the greenish ring, the situation and activity context are an important factor for the model decision (**H2**). As mentioned above, during rest periods hypoglycemia is unlikely to be detected by individuals with diabetes due to limited perception capabilities. Thus, the features for **H1** and **H2** are the major reason for the hypoglycemia warning in this case.

We believe, that such explanations to individuals with diabetes have an educational effect and help them to better understand and cope with their disease.

Conclusion

This work shows the results of our ongoing research in hypoglycemia detection from physiological data. We show the feasibility of building a machine learning model for detecting hypoglycemia. Additionally, we give an outlook on how we plan to explain model classifications of hypoglycemia to individuals with diabetes with easy understandable visualizations on a smartwatch.

In further work, we will integrate additional sensors as input features to the model. This includes movement information from the accelerometer as well as a sensor for electrodermal activity, which is today already available in professional smartwatches such as the *Empatica E4*. In further work, we plan to implement and evaluate our concept for the hypoglycemia warning system on a smartwatch in a longitudinal field-study among multiple individuals with continuous glucose monitors for reference.

While wearables might not yield perfect accuracy in detecting hypoglycemia, we believe that our visually explained hypoglycemia warning can contribute to the everyday life of individuals with diabetes.

REFERENCES

[1] Tadej Battelino, Thomas Danne, Richard M. Bergenstal, Stephanie A. Amiel, Roy Beck, Torben Biester, Emanuele Bosi, Bruce A. Buckingham, William T. Cefalu, Kelly L. Close, Claudio Cobelli, Eyal Dassau, J. Hans DeVries, Kim C. Donaghue, Klemen Dovc, Francis J. Doyle, Satish Garg, George Grunberger, Simon Heller, Lutz Heinemann, Irl B. Hirsch, Roman Hovorka, Weiping Jia, Olga Kordonouri, Boris Kovatchev, Aaron Kowalski, Lori Laffel, Brian Levine, Alexander Mayorov, Chantal Mathieu, Helen R. Murphy, Revital Nimri, Kirsten Nørgaard, Christopher G. Parkin, Eric Renard, David Rodbard, Banshi Saboo, Desmond Schatz, Keaton Stoner, Tatsuiko Urakami, Stuart A. Weinzimer, and Moshe Phillip. 2019. Clinical Targets for Continuous Glucose Monitoring Data Interpretation: Recommendations From the International Consensus on Time in Range. *Diabetes Care* (2019). DOI: http://dx.doi.org/10.2337/dci19-0028

[2] Enzo Bonora, Francesco Calcaterra, Simonetta Lombardi, Nicoletta Bonfante, Gianni Formentini, Riccardo C. Bonadonna, and Michele Muggeo. 2001. Plasma Glucose Levels Throughout the Day and HbA1c Interrelationships in Type 2 Diabetes. *Diabetes Care* 24, 12 (2001), 2023–2029. DOI: http://dx.doi.org/10.2337/diacare.24.12.2023

[3] International Diabetes Federation. 2019. *IDF Diabetes Atlas* (9 ed.). International Diabetes Federation. http://diabetesatlas.org

[4] Jerome H. Friedman. 2001. Greedy function approximation: A gradient boosting machine. *Annals of Statistics* 29, 5 (2001), 1189–1232. DOI: http://dx.doi.org/10.2307/2699986

[5] Alex J. Graveling, Ian J. Deary, and Brian M. Frier. 2013. Acute hypoglycemia impairs executive cognitive function in adults with and without type 1 diabetes. *Diabetes Care* 36, 10 (2013), 3240–3246. DOI: http://dx.doi.org/10.2337/dc13-0194

[6] Katrin Hänsel, Romina Poguntke, Hamed Haddadi, Akram Alomainy, and Albrecht Schmidt. 2018. What to Put on the User: Sensing Technologies for Studies and Physiology Aware Systems. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, Article Paper 145, 14 pages. DOI: http://dx.doi.org/10.1145/3173574.3173719

[7] Scott M. Lundberg, Bala Nair, Monica S. Vavilala, Mayumi Horibe, Michael J. Eisses, Trevor Adams, David E. Liston, Daniel King-Wai Low, Shu-Fang Newman, Jerry Kim, and Su-In Lee. 2018. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering* 2, 10 (oct 2018), 749–760. DOI: http://dx.doi.org/10.1038/s41551-018-0304-0

[8] Sumie Mitsuishi, Rimei Nishimura, Kiyotaka Ando, Daisuke Tsujino, and Kazunori Utsunomiya. 2015. Can fasting glucose levels or post-breakfast glucose fluctuations predict the occurrence of nocturnal asymptomatic hypoglycemia in type 1 diabetic patients receiving basal-bolus insulin therapy with long-Acting insulin? *PLoS ONE* 10, 12 (2015), 1–9. DOI: http://dx.doi.org/10.1371/journal.pone.0144041

[9] Loretto Munoz, Arie Van Roon, Harriëtte Riese, Chris Thio, Emma Oostenbroek, Iris Westrik, Eco J.C. De Geus, Ron Gansevoort, Joop Lefrandt, Ilja M. Nolte, and Harold Snieder. 2015. Validity of (Ultra-)Short recordings for heart rate variability measurements. *PLoS ONE* 10, 9 (2015), e0138921. DOI: http://dx.doi.org/10.1371/journal.pone.0138921

[10] Marleen Olde Bekkink, Mats Koeneman, Bastiaan E. de Galan, and Sebastian J. Bredie. 2019. Early Detection of Hypoglycemia in Type 1 Diabetes Using Heart Rate Variability Measured by a Wearable Device. *Diabetes Care* 42, 4 (Apr 2019), 689 LP – 692. DOI:http://dx.doi.org/10.2337/dc18-1843

[11] Dung Phan, Lee Yee Siong, Pubudu N. Pathirana, and Aruna Seneviratne. 2015. Smartwatch: Performance evaluation for long-term heart rate monitoring. In *4th International Symposium on Bioelectronics and Bioinformatics, ISBB 2015*. 144–147. DOI: http://dx.doi.org/10.1109/ISBB.2015.7344944

[12] Shriti Raj, Kelsey Toporski, Ashley Garrity, Joyce M. Lee, and Mark W. Newman. 2019. "My Blood Sugar is Higher on the Weekends": Finding a Role for Context and Context-Awareness in the Design of Health Self-Management Technology. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, Article Paper 119, 13 pages. DOI: http://dx.doi.org/10.1145/3290605.3300349

[13] A. M. Robinson, Heather M. Parkin, Ian A. Macdonald, and Robert B. Tattersall. 1994. Physiological response to postural change during mild hypoglycaemia in patients with IDDM. *Diabetologia* 37, 12 (Dec 1994), 1241–1250. DOI: http://dx.doi.org/10.1007/BF00399798

[14] Hartmut Schächinger, Johannes Port, Stuart Brody, Lilly Linder, Frank H. Wilhelm, Peter R. Huber, Daniel Cox, and Ulrich Keller. 2004. Increased high-frequency heart rate variability during insulin-induced hypoglycaemia in healthy humans. *Clinical Science* 106, 6 (2004), 583–588. DOI: http://dx.doi.org/10.1042/cs20030337

[15] Fred Shaffer and J. P. Ginsberg. 2017. An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health* 5 (2017), 258. DOI: http://dx.doi.org/10.3389/fpubh.2017.00258

[16] Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology. 1996. Heart Rate Variability. *Circulation* 93, 5 (1996), 1043–1065. DOI: http://dx.doi.org/10.1161/01.CIR.93.5.1043

[17] Edith W. Ter Braak, Alexander M. Appelman, Mariëlle F. Van Laak, Ronald P. Stolk, Timon W. Van Haeften, and Dirk W. Erkelens. 2000. Clinical characteristics of type 1 diabetic patients with and without severe hypoglycemia. *Diabetes Care* 23, 10 (2000), 1467–1471. DOI: http://dx.doi.org/10.2337/diacare.23.10.1467

[18] Lisa M. Younk, Maia Mikeladze, Donna Tate, and Stephen N. Davis. 2011. Exercise-related hypoglycemia in diabetes mellitus. *Expert Review of Endocrinology & Metabolism* 6, 1 (2011), 93–108. DOI:http://dx.doi.org/10.1586/eem.10.78