Smartwatches for non-invasive hypoglycaemia detection during cognitive and psychomotor stress

Maritsch, Martin ; Föll, Simon ; Lehmann, Vera ; Styger, Naïma ; Bérubé, Caterina ; Kraus, Mathias ; Feuerriegel, Stefan ; Kowatsch, Tobias ; Züger, Thomas ; Fleisch, Elgar ; Wortmann, Felix ; Stettler, Christoph

DOI: https://doi.org/10.1111/dom.15402

Received: 3 October 2023 | Revised: 16 November 2023 | Accepted: 21 November 2023

DOI: 10.1111/dom.15402

RESEARCH LETTER

WILEY

Smartwatches for non-invasive hypoglycaemia detection during cognitive and psychomotor stress

Martin Maritsch PhD[1] | Simon Föll PhD[1] | Vera Lehmann MD[2] | Naïma Styger BSc[2] | Caterina Bérubé PhD[1] | Mathias Kraus PhD[3] | Stefan Feuerriegel PhD[4,5] | Tobias Kowatsch PhD[1,6,7] | Thomas Züger MD[2,8] | Elgar Fleisch PhD[1,9] | Felix Wortmann PhD[9] | Christoph Stettler MD[2]

[1]Department of Management, Technology and Economics, ETH Zurich, Zurich, Switzerland

[2]Department of Diabetes, Endocrinology, Nutritional Medicine and Metabolism, Inselspital, Bern, University Hospital, University of Bern, Bern, Switzerland

[3]School of Business, Economics and Society, Friedrich-Alexander University Erlangen-Nürnberg, Nürnberg, Germany

[4]Institute of AI in Management, LMU Munich, Munich, Germany

[5]Munich Center for Machine Learning (MCML), Munich, Germany

[6]Institute for Implementation Science in Health Care, University of Zurich, Zurich, Switzerland

[7]School of Medicine, University of St Gallen, St Gallen, Switzerland

[8]Department of Endocrinology and Metabolic Diseases, Kantonsspital Olten, Olten, Switzerland

[9]Institute of Technology Management, University of St Gallen, St Gallen, Switzerland

Correspondence
Christoph Stettler, Department of Diabetes, Endocrinology, Nutritional Medicine and Metabolism, Inselspital, Bern, University Hospital, University of Bern, Bern, Switzerland.
Email: christoph.stettler@insel.ch

Funding information
Clinical Trials Unit Berne; Department of Research of the University Hospital Berne; Diabetes Center Berne; Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: SNF CRSII5_183569; Swiss Diabetes Foundation

1 | BACKGROUND

Hypoglycaemia is one of the most relevant complications of diabetes[1] and induces alterations in physiological parameters[2,3] that can be measured with smartwatches and detected using machine learning (ML).[4] The performance of these algorithms when applied to different hypoglycaemic ranges or in situations involving cognitive and psychomotor stress remains unclear. Demanding tasks can significantly affect the physiological responses on which the wearable-based hypoglycaemia detection relies.[5] The present analysis aimed to investigate ML-based hypoglycaemia detection using wearable data at different levels of hypoglycaemia during a complex task involving cognitive and psychomotor challenges (driving).

2 | METHODS

This was a pre-specified analysis of two consecutive studies conducted at the University Hospital Bern (Study A: 10/2019-07/2020; Study B: 11/2021-03/2022). The studies followed the Declaration of Helsinki, guidelines of good clinical practice, and legally applicable requirements. We included individuals with type 1 diabetes, aged 21-50 years (up to 60 years for Study B). Key exclusion criteria included pregnancy, seizure disorders, severe organ dysfunction, or alcohol/drug abuse. Following approval by the local Ethics Committee

Martin Maritsch, Simon Föll, and Vera Lehmann share first authorship.

Felix Wortmann and Christoph Stettler share last authorship.

Bern (2019-00579, 2021-002018), the studies were registered on ClinicalTrials.gov (NCT04035993, NCT05183191). Participants provided written informed consent. The detailed study procedures have been described previously.[6] After the screening visit, participants were admitted to our clinical research facility for the main experiment after an overnight fast. Participants received a smartwatch (Garmin vivoactive 3) to collect heart rate variability and acceleration sensor data and a second wrist-worn wearable (Empatica E4) to collect electrodermal activity data. During a controlled hypoglycaemia procedure using intravenous insulin and glucose (Figure S1A), wearable data were collected in euglycaemia [blood glucose (BG) 5.0-8.0 mmol/L] followed by pronounced hypoglycaemia in Study A (BG 2.0-2.5 mmol/L) or mild hypoglycaemia in Study B (BG 3.0-3.5 mmol/L), while individuals were driving in a simulator. The hypoglycaemic episode lasted ~40 min. Venous BG was repeatedly measured using the Biosen C-Line analyser and BG was used as the ground truth for ML modelling. Hypoglycaemia symptoms were rated by the participants using a standardized questionnaire.[7]

The main outcome of the analysis was the diagnostic accuracy of the ML approach to detect hypoglycaemia using wearable data quantified as the area under the receiver operating characteristic curve (AUROC). The sample size was calculated based on the primary outcome of the two interventional studies (for details see Lehmann et al.[6]).

Figure S1B displays the ML pipeline. To assess performance and decision-making for different hypoglycaemic ranges, we trained and evaluated three ML models: (a) to detect pronounced hypoglycaemia (BG 2.0-2.5 mmol/L), using data from Study A; (b) to detect mild hypoglycaemia (BG 3.0-3.5 mmol/L), using data from Study B; and (c) to detect overall hypoglycaemia, combining data from both studies. To handle electrodermal activity data, we separated the signal into a slow-changing (tonic) and a fast-changing (phasic) component. We follow conventional feature engineering for time-series classification: Each signal is cut into overlapping sequences of 60 s and shifted by 1 s. Using aggregation functions, we generated a set of 14 generic and domain-specific features (Table S1). BG values were linearly interpolated between measurements, while the binary output variable for ML models was set to 1 for hypoglycaemia, otherwise 0. We implemented a logistic regression with ridge regularization to ensure interpretable decision-making while feature importance was derived from the magnitude of the coefficients grouped by the modality.[8] Furthermore, we assessed generalization to unseen subjects by using leave-one-subject-out cross-validation to compute performance metrics' mean and standard deviation. The source code has been published,[9] and is publicly available on GitHub (https://github.com/im-ethz/smartwatches-hypo-detection).

3 | RESULTS

For Study A, 17 of 18 individuals were included in the analysis (one participant excluded because of wearable data loss). For Study B, all nine enrolled individuals were included in the analysis. Baseline characteristics are provided in Table S2. Venous blood glucose in euglycaemia was 5.9 ± 0.6 mmol/L (Study A) and 6.4 ± 0.8 mmol/L (Study B). Hypoglycaemia was at 2.4 ± 0.2 mmol/L (Study A) and 3.3 ± 0.2 mmol/L (Study B). Self-rated hypoglycaemia symptoms are reported in Table S3.

Pronounced hypoglycaemia was detected with an AUROC of 0.80 ± 0.20 and mild hypoglycaemia with an AUROC of 0.66 ± 0.12. AUROC of overall hypoglycaemia detection was 0.72 ± 0.20. Additional performance metrics are provided in Table 1.

Figure 1 displays the feature importance (i.e. the sum of the coefficients per modality divided by the sum of all coefficients) for the different input categories across the levels of hypoglycaemia. While detection of pronounced hypoglycaemia shows a higher reliance on electrodermal activity, the feature importance for mild hypoglycaemia is balanced across input modalities. For the detection of overall hypoglycaemia, the features of electrodermal activity dominate.

4 | DISCUSSION

Previously, we have shown the general concept of hypoglycaemia detection using smartwatch data.[4] The present analysis suggests that this concept extends to detecting pronounced hypoglycaemia even when people are involved in tasks related to cognitive and psychomotor stress, such as driving. The more modest performance in detecting mild hypoglycaemia may be attributed to (a) the smaller sample size of individuals undergoing mild hypoglycaemia, (b) the potential interference of the physiological response to the driving task,[5] and/or (c) a more subtle physiological response to mild hypoglycaemia when compared with lower hypoglycaemic levels.[10] Additional data may improve ML performance in mild hypoglycaemia by learning generalizable patterns.

While electrodermal activity, heart rate variability and accelerometer features constitute relevant features for hypoglycaemia detection,[2,4] the accuracy of ML decision-making across different levels of hypoglycaemia is unknown. Electrodermal activity proved to be a decisive feature modality for detecting pronounced hypoglycaemia, but appeared to be less informative in detecting mild hypoglycaemia. Electrodermal activity represents the electrical response of sweat glands to sympathetic innervation.[11] Increased

TABLE 1 Machine learning performance metrics.

Model	AUROC	AUPRC	BACC	F1-score	MCC	Sensitivity	Specificity
Pronounced	0.80 ± 0.20	0.80 ± 0.21	0.80 ± 0.17	0.79 ± 0.18	0.61 ± 0.33	0.81 ± 0.21	0.79 ± 0.28
Mild	0.66 ± 0.12	0.67 ± 0.12	0.67 ± 0.08	0.69 ± 0.07	0.36 ± 0.16	0.75 ± 0.16	0.59 ± 0.23
Overall	0.72 ± 0.20	0.72 ± 0.20	0.73 ± 0.17	0.72 ± 0.18	0.47 ± 0.32	0.75 ± 0.23	0.70 ± 0.28

Abbreviations: AUPRC, area under the precision-recall curve; AUROC, area under the receiver operating characteristic curve; BACC, balanced accuracy; MCC, Matthews correlation coefficient.

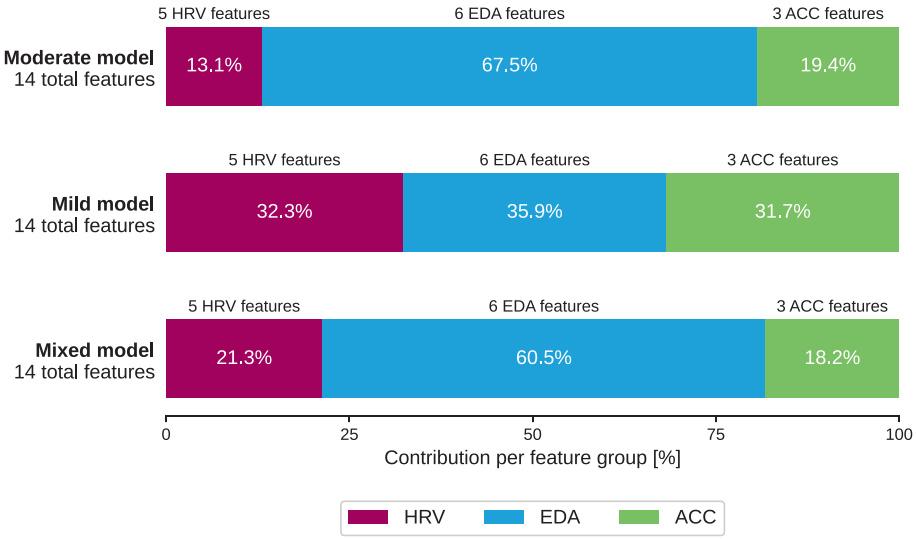

FIGURE 1 Feature importance across different hypoglycaemic levels. ACC, accelerometer; EDA, electrodermal activity; HRV, heart rate variability.

sweating enhances skin conductivity, translating into increased electrodermal activity.[12] Therefore, our findings align with the existing literature, which suggests that the threshold for neurogenic symptoms (including sweating) is about 3.0 mmol/L.[13] The dominance of electrodermal activity in the model coefficients persists for overall hypoglycaemia detection, probably because more participants were experiencing pronounced hypoglycaemia compared with mild hypoglycaemia. However, even when the model training was limited to a balanced subsample of participants, this pattern persisted.

Strengths include the study setting providing standardized hypoglycaemia using the gold standard (venous BG). The ML models are based on data recorded from unobtrusive consumer-grade devices, rendering our approach readily available. Self-reported symptom scores in hypoglycaemia were comparably low, emphasizing the need for alternative hypoglycaemia detection methods. For evaluating the ML models, we used a leave-one-subject-out approach, which allows for assessing generalizability of the models to unseen individuals. Unlike personalized models used in a previous study,[4] this approach offers scalability by eliminating the need to train individual models for each user. Interpretable ML allows assessing the decision-making across different levels of hypoglycaemia. We acknowledge that the smaller group size in Study B might have compromised the performance of mild hypoglycaemia detection. Our data were collected in standardized, stable glycaemic conditions during cognitive stress exclusively. Consequently, further research is needed to determine the exact time point of first hypoglycaemia detection and to quantify directly the impact of stress on detection performance. The glycaemic sequence (euglycaemia followed by hypoglycaemia) may have introduced bias. However, this reflects glycaemic trajectories in reality. Our dataset was derived from generally healthy individuals with well-controlled type 1 diabetes, which limits the generalizability of our results to individuals with comorbidities, hypoglycaemia unawareness, or other diabetes types.

In summary, pronounced hypoglycaemia can be detected using commercially available wearables even in situations with cognitive and psychomotor stress, such as driving. However, the proposed approach is less reliable for milder hypoglycaemia, and additional research is needed to optimize the corresponding models further.

AUTHOR CONTRIBUTIONS

The following authors contributed to the conception and design of the study: MM, SF, VL, TZ, CS and FW. Acquisition of data: MM, VL, NS, CB, MK and TZ. Analysis of data: MM, SF, VL and FW. Interpretation of data: MM, SF, VL, CS and FW. Writing the manuscript: SF, VL, CS and FW. Critical review of the manuscript; MM, NS, CB, MK, SF, TK, TZ and EF. CS is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors approved the final draft of the manuscript for submission.

ACKNOWLEDGMENTS

The authors thank all the volunteers for their participation and enthusiasm. Thanks go to Constance Bischoff and Joana Filipa Rodrigues Cunha Freitas (study nurses), Mona Spoerri (medical student) and Vincent Vorburger (trainee) for their valuable support during the clinical visits.

FUNDING INFORMATION

The studies were funded by the Swiss National Science Foundation (Sinergia grant, SNF CRSII5_183569), with additional grants provided by the Clinical Trials Unit Berne, the Swiss Diabetes Foundation, the Diabetes Center Berne and the Department of Research of the University Hospital Berne.

CONFLICT OF INTEREST STATEMENT

CB, TK and EF are affiliated with the Centre for Digital Health Interventions (CDHI), a joint initiative of the Institute for Implementation Science in Health Care, University of Zurich, the Department of Management, Technology, and Economics at ETH Zurich, and the Institute of Technology Management and School of Medicine at the University of St Gallen. CSS, a Swiss health insurer, partly funds CDHI. TK and

EF are also cofounders of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. However, neither CSS nor Pathmate Technologies was involved in this research. All other authors have no conflicts of interest to declare.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/dom.15402.

DATA AVAILABILITY STATEMENT

The data sets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

ORCID

Martin Maritsch https://orcid.org/0000-0001-9920-0587
Vera Lehmann https://orcid.org/0000-0002-6038-809X
Christoph Stettler https://orcid.org/0000-0003-1691-6059

REFERENCES

1. Seaquist ER, Anderson J, Childs B, et al. Hypoglycemia and diabetes: a report of a workgroup of the American Diabetes Association and the Endocrine Society. *Diabetes Care*. 2013;36:1384-1395.
2. Olde Bekkink M, Koeneman M, de Galan BE, Bredie SJ. Early detection of hypoglycemia in type 1 diabetes using heart rate variability measured by a wearable device. *Diabetes Care*. 2019;42:689-692.
3. Koeneman M, Olde Bekkink M, Meijel LV, Bredie S, de Galan B. Effect of hypoglycemia on heart rate variability in people with type 1 diabetes and impaired awareness of hypoglycemia. *J Diabetes Sci Technol*. 2022;16:1144-1149.
4. Lehmann V, Föll S, Maritsch M, et al. Noninvasive hypoglycemia detection in people with diabetes using smartwatch data. *Diabetes Care*. 2023;46:993-997.
5. Mehler B, Reimer B. How demanding is" just driving?" a cognitive workload-psychophysiological reference evaluation. *Driving Assesment Conference*. Vol 2019. University of Iowa; 2019.
6. Lehmann V, Zueger T, Maritsch M, et al. Machine learning for non-invasive sensing of hypoglycaemia while driving in people with diabetes. *Diabetes Obes Metab*. 2023;25:1668-1676.
7. Cox DJ, Gonder-Frederick LA, Kovatchev BP, Julian DM, Clarke WL. Progressive hypoglycemia's impact on driving simulation performance. Occurrence, awareness and correction. *Diabetes Care*. 2000;23:163-170.
8. Park JH, Cho HE, Kim JH, et al. Machine learning prediction of incidence of Alzheimer's disease using large-scale administrative health data. *NPJ Digit Med*. 2020;3:46.
9. Föll S, Maritsch M, Spinola F, et al. FLIRT: a feature generation toolkit for wearable data. *Comput Methods Programs Biomed*. 2021;212:106461.
10. Verhulst CEM, Fabricius TW, Teerenstra S, et al. Glycaemic thresholds for counterregulatory hormone and symptom responses to hypoglycaemia in people with and without type 1 diabetes: a systematic review. *Diabetologia*. 2022;65:1601-1612.
11. Boucsein W, Fowles DC, Grimnes S, et al. Publication recommendations for electrodermal measurements. *Psychophysiology*. 2012;49:1017-1034.
12. Posada-Quintero HF, Chon KH. Innovations in electrodermal activity data collection and signal processing: a systematic review. *Sensors*. 2020;20:479.
13. Hepburn DA, Deary IJ, Frier BM, Patrick AW, Quinn JD, Fisher BM. Symptoms of acute insulin-induced hypoglycemia in humans with and without IDDM. Factor-analysis approach. *Diabetes Care*. 1991;14:949-957.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Maritsch M, Föll S, Lehmann V, et al. Smartwatches for non-invasive hypoglycaemia detection during cognitive and psychomotor stress. *Diabetes Obes Metab*. 2023;1-4. doi:10.1111/dom.15402