Machine learning to infer a health state using biomedical signals – detection of hypoglycemia in people with diabetes while driving real cars

Authors and affiliations

Vera Lehmann[1]* MD PhD, Thomas Zueger[1,2]* MD, Martin Maritsch[3]* PhD, Michael Notter[1] BSc, Simon Schallmoser[4,5] MSc, Caterina Bérubé[3] PhD, Caroline Albrecht[1] MD, Mathias Kraus[3,6] PhD, Stefan Feuerriegel[4,5] PhD, Elgar Fleisch[3,7] PhD, Tobias Kowatsch[3,8,9] PhD, Sophie Lagger[1] MSc, Markus Laimer[1] MD, Felix Wortmann[7#] PhD, Christoph Stettler[1#] MD

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

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

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

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

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

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

[7] Institute of Technology Management, University of St. Gallen, St. Gallen, Switzerland

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

[9] School of Medicine, University of St. Gallen, St. Gallen, Switzerland

** equal first authorship, # equal last authorship*

Corresponding author

Prof Christoph Stettler, MD

Department of Diabetes, Endocrinology, Nutritional Medicine and Metabolism, Inselspital, Bern

Freiburgstrasse 15, CH-3010 Bern, Switzerland

Phone: +41 31 632 40 70

E-mail: christoph.stettler@insel.ch

Abstract

Background

Hypoglycemia, one of the most dangerous acute complications of diabetes, poses a substantial risk for vehicle accidents. To date, reliable detection and warning of hypoglycemia while driving remains an unmet need, as current sensing approaches are restricted by diagnostic delay, invasiveness, availability, and costs. This work aimed to develop and evaluate a machine learning (ML) approach detecting hypoglycemia during driving through driving and gaze/head motion data.

Methods

We collected driving and gaze/head motion data (47,998 observations) during controlled eu- and hypoglycemia from 30 individuals with type 1 diabetes (24 males, age 40.1±10.3y, HbA_{1c} 6.9±0.7%, 51.9±8.0 mmol/mol) while participants drove with a real car. Machine learning (ML) models were built and evaluated to detect hypoglycemia based solely on driving and gaze/head motion data.

Results

The ML approach detected hypoglycemia with high accuracy (area under the receiver operating characteristic curve [AUROC] 0.80±0.11). When restricted to either driving or gaze/head motion data only, the detection performance remained high (AUROC of 0.73±0.07 and 0.70±0.16, respectively).

Conclusions

Hypoglycemia can be detected non-invasively during real car driving using an ML approach purely based on driving and gaze/head motion data, improving driving safety and self-management for people with diabetes. Interpretable ML also provided novel insights into behavioral changes when driving in hypoglycemia.

Study registration

ClinicalTrials.gov (NCT04569630, NCT05308095)

Introduction

Despite significant advances in diabetes care, hypoglycemia remains one of the most relevant challenges.[1] Hypoglycemia significantly affects cognitive, executive, and psychomotor abilities,[2] posing a significant risk to the safe performance of everyday tasks, such as driving. Unrecognized hypoglycemia is likely to cause a substantial number of road accidents in individuals with diabetes.[3-5]

Existing methods for hypoglycemia detection encompass self-monitoring of blood glucose (SMBG) and continuous glucose monitoring (CGM). While SMBG lacks proactive warnings, CGM provides permanent glucose readings but is restricted by invasiveness, accessibility, costs and time delay.[6] The latter is particularly relevant in driving, where rapid action must be taken. The considerable and growing number of vehicular accidents in people with diabetes,[7] emphasizes the need for novel hypoglycemia detection approaches. As shown previously, machine learning (ML) can be leveraged to infer relevant health states through biomedical signals.[8, 9]

Cars produce real-time, high-resolution information on various driving features (e.g., velocity, braking, steering, etc.), which are transferred through the Controller Area Network (CAN) bus. Driver monitoring cameras (DMCs) capturing gaze and head movements are increasingly installed to track driver behavior, also in (semi-) autonomous vehicles. A hypoglycemia detection system based exclusively on CAN and DMC data (no glucose measurement) may offer a novel, non-invasive and readily accessible solution to improve road safety for individuals with diabetes.

In a proof-of-principle driving simulator study, we recently demonstrated that an ML-based system using driving and eye tracker data might enable the detection of hypoglycemia while driving.[10] However, simulator studies have fundamental limitations compared to real car studies. Firstly, driving simulators cannot fully replicate the

physical and perceptual experiences of real car driving, lacking in haptic feedback, tactile sensations, and genuine sensory input. Secondly, simulators cannot recreate the comprehensive environmental variables of real-world driving, such as weather and road conditions, which significantly influence vehicle dynamics and behavior. Thirdly, simulators do not provide real-world consequences for dangerous situations, leading to altered participant behavior. Lastly, the type, resolution, and quality of the respective sensors vary substantially.

The present study aimed to develop and evaluate an ML-based system for hypoglycemia detection in real car driving. Specifically, our aim was to implement an ML approach based on CAN and DMC data alone to detect hypoglycemia while driving. Thereby we address the well-known limitations of simulators[11] and go beyond previous research that has never induced hypoglycemia during real car driving.

Methods

Overview

We performed two studies in people with type 1 diabetes driving a car, collecting data from differing hypoglycemic levels (

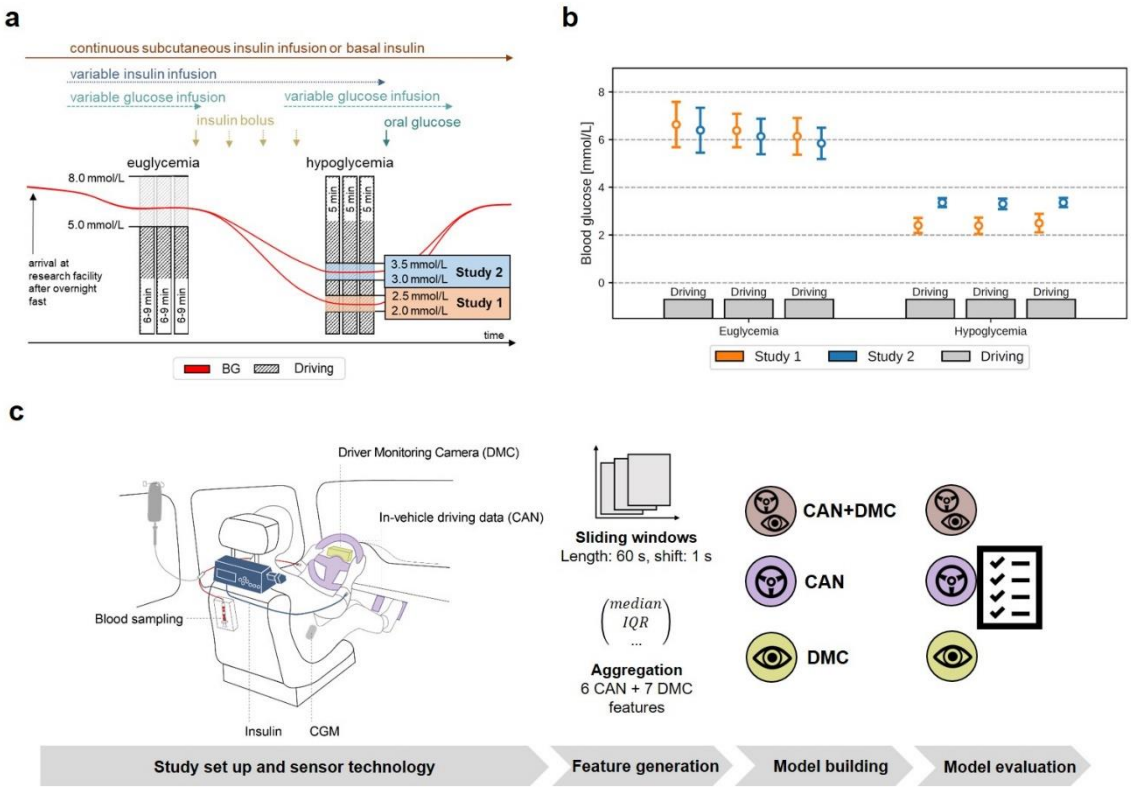

Figure 1). In study 1 (10/2020–05/2021, NCT04569630), we collected CAN and DMC data while participants drove during euglycemia and pronounced hypoglycemia (blood glucose [BG] 2.0–2.5mmol/L). In study 2 (04/2022–06/2022, NCT05308095), we collected CAN and DMC data in euglycemia and mild hypoglycemia (BG 3.0–3.5mmol/L). We used CAN and DMC data from both studies to develop and evaluate ML models predicting whether a driver was driving in eu- vs. hypoglycemia. To reflect different vehicle generations, we assessed hypoglycemia detection using three distinct

ML models: (i) the *CAN+DMC model*, which integrates driving and gaze/head motion data, representing cutting-edge technology in modern cars, (ii) the *CAN model* exclusively used driving data, considering that not all contemporary cars have DMC, and (iii) the *DMC model* used only gaze/head motion data, anticipating that in the future (semi-)autonomous driving[12] may limit the impact of CAN data.

Design and population

Both studies were conducted at the University Hospital Bern and a nearby test track, following the guidelines of good clinical practice, the Declaration of Helsinki, and the local legally applicable requirements. The Ethics Committee Bern approved both studies (2020-00685, 2021-02381). The participants provided written informed consent. We included active drivers with type 1 diabetes, aged 21–60y, with HbA_{1c}≤9.0%. Key exclusion criteria included: pregnancy; severe organ dysfunction; cardiovascular or cardiac disease; seizure disorders; drug or alcohol abuse; and medication interfering with driving performance (for details see Supplementary Methods, p2).

Procedure

Figure S1 displays the visit schedule. After screening, participants were equipped with the factory-calibrated Dexcom G6 CGM, and instructed on how to avoid hypoglycemia during participation. For the main visit (3 to 7 days after the screening), participants were admitted to our clinical research unit after an overnight fast. After insertion of the intravenous cannulas, participants were transferred to the nearby test track. Before starting the experiment, the participants completed a test driving session. Subsequently, participants drove with the study car (Volkswagen Touran, automatic gearshift) during a controlled hypoglycemia procedure on the secured test track supervised by a driving instructor (Figure S2). Driving was performed in euglycemia

(5.0–8.0mmol/L) and hypoglycemia (study 1: 2.0–2.5mmol/L, study 2: 3.0–3.5mmol/L). In each glycemic state, the participants completed ≈20min of driving, involving a randomized sequence of three driving scenarios. Each scenario lasted between 6–9min, with intermittent 1–2-min pauses for venous BG measurement with the Biosen C-Line glucose analyzer. To capture driving data, we used a CAN-bus interface and recorded vehicle signals that were resampled to 50Hz. For gaze/head motion tracking (DMC), we used a pre-series near-infrared camera system (Robert Bosch GmbH, Stuttgart, Germany),[13] which was mounted behind the steering wheel on the steering column and captured imagery with a framerate of up to 50Hz. The DMC was calibrated before each experiment.

Participants were instructed that hypoglycemia targeting a BG level of 2.0–2.5mmol/L (study 1) or 3.0–3.5mmol/L (study 2) would be induced. However, throughout the experiment they remained blinded to their actual BG measurements. Participants rated (i) eight hypoglycemia symptoms, (ii) their perceived need for immediate treatment, (iii) and the difficulty level experienced while driving on a seven-point Likert scale (6=extreme; 0=none).[14] Moreover, participants estimated their BG level in mmol/L.

One to three days afterwards, the safety assessment was performed. Further details on the study procedure can be found in the Supplementary Methods (pp 2–3).

Outcome

Our study outcome was the diagnostic accuracy of our ML approach in detecting hypoglycemia quantified as the area under the receiver operating characteristic curve (AUROC). For the sample size calculation see Supplementary Methods (p 3).

Analysis and ML approach

Figure 1c outlines the ML pipeline of the three hypoglycemia detection models (CAN+DMC, CAN, and DMC). Input data consisted of in-vehicle (CAN) and/or driver monitoring camera (DMC) data. Existing approaches for drowsiness and intoxicated driving detection[15-17] informed our choice of features (Supplementary Methods, p 4). The fundamental CAN data signals were 'steering wheel angle', 'steering wheel velocity', 'brake pedal position', 'gas pedal position', 'vehicle velocity', and 'vehicle acceleration'. From the DMC video data, we extracted gaze and head rotation signals using a standard proprietary algorithm that is commercially available in the automotive industry. The algorithm uses a four step procedure (face detection, eye region localization, pupil detection, gaze vector calculation) to determine the gaze and the head rotation (face detection, facial landmark detection, head pose calculation, head rotation calculation). Ultimately, we extracted two gaze ('gaze velocity' and 'gaze acceleration') and five head rotation signals ('head rotation velocity, 'head rotation acceleration', 'head acceleration roll', 'head acceleration pitch', and 'head acceleration yaw'). For feature engineering, we followed the conventions for time-series classification and used a sliding window approach.[18] We cut all signals into sequences (windows) of 60sec with a shift of 1sec between adjacent windows; then we applied statistical aggregation functions on each sequence, generating a set of interpretable features for each sequence (e.g., median gaze velocity). This resulted in six features for CAN and seven features for DMC (Table S1). For each window, the binary output variables were set based on the venous BG (to 1 for BG <3.9mmol/L, and to 0 otherwise) measured with the gold-standard device (Biosen).

The ML models were implemented as logistic regression with ridge regularization (see Supplementary Methods, p 4 for further specifications). Following best practice[19], the

performance for hypoglycemia detection was measured in a leave-one-subject-out cross-validation (n=30), i.e., the ML models were trained on n=29 subjects and are then evaluated on the left-out subject. This procedure was repeated until each subject has been left out for evaluation once. Both studies had no overlap in study participants. Results are presented as the out-of-sample detection performance, averaged across participants (i.e., macro-average). To measure the performance variation across participants, we provide the standard deviation (SD) at the participant level. To ensure generalizability, hyperparameters were fixed and thus the same across models and participants. Additionally, we ran robustness checks to confirm our results, including the evaluation of other (non-)linear ML models (e.g., gradient boosting decision tree, etc.), a sensitivity analysis across different window lengths, and different training and evaluation procedures (Supplementary Methods, p 5 and Tables S2–S4). We analyzed the regression coefficients and odds ratios (OR) to assess the influence of each feature on the decision-making of the CAN+DMC model.

Statistical reporting

Results are presented as mean±SD (unless otherwise specified). For paired BG and CGM measurements we used the Shapiro-Wilk test to assess normality, followed by a comparison through paired t-tests or Wilcoxon signed rank tests as appropriate. For the self-rated symptoms, we computed neurogenic and neuroglycopenic scores by averaging the four respective symptoms within each category[20]. The overall symptom score was determined by averaging all individual symptoms. Normality checks were applied to symptom scores, individual symptoms, and self-estimated BG values. Comparisons between euglycemia and hypoglycemia were made using paired t-tests or Wilcoxon signed rank tests. Statistical significance was set at p<0.05.

Software used for data analysis

Descriptive statistical analyses were conducted with STATA 16.0 (StataCorp LLC, Texas, USA). The ML models were implemented using Python 3.7.13 using the scikit-learn package (version 1.0.2). For robustness checks, the package XGBoost (version 1.3.3) was employed. Input features to the ML models were computed using NumPy (version 1.20.3) and SciPy (version 1.7.3). Evaluation metrics were computed using scikit-learn (version 1.0.2). Details regarding the software used for data management can be found in the Supplementary Methods (p 5).

Results

Study population and glucose

Thirty individuals with type 1 diabetes generating 47,998 observations were included in the final analysis (age 40.1±10.3y, 24 males, HbA_{1c} 6.9±0.7%, Table 1). The study flows are depicted in Figure S1. In study 1 (20 participants), mean venous BG in eu- and hypoglycemia was 6.38±0.84mmol/L, and 2.43±0.35mmol/L. Corresponding CGM values were 6.81±0.85mmol/L and 3.12±0.63mmol/L (p<0.01 compared to venous BG). In study 2 (10 participants), mean venous BG in eu- and hypoglycemia was 6.18±0.83mmol/L, and 3.34±0.20mmol/L, respectively. Corresponding CGM readings were 6.67±1.26mmol/L and 3.85±0.50mmol/L, respectively (p<0.01 compared to venous BG). Individual BG measurements during driving are depicted in Figure S3.

ML-based detection of hypoglycemia

The feature engineering approach described in the Methods led to 32,537 (15,461) observations for study 1 (study 2), out of which 16,196 (7,728) observations come from driving in euglycemia (class 0, BG≥3.9mmol/L) and 16,341 (7,733) from driving in hypoglycemia (class 1, BG<3.9mmol/L).

For hypoglycemia detection, the CAN+DMC model yielded an AUROC of 0.80±0.11 (

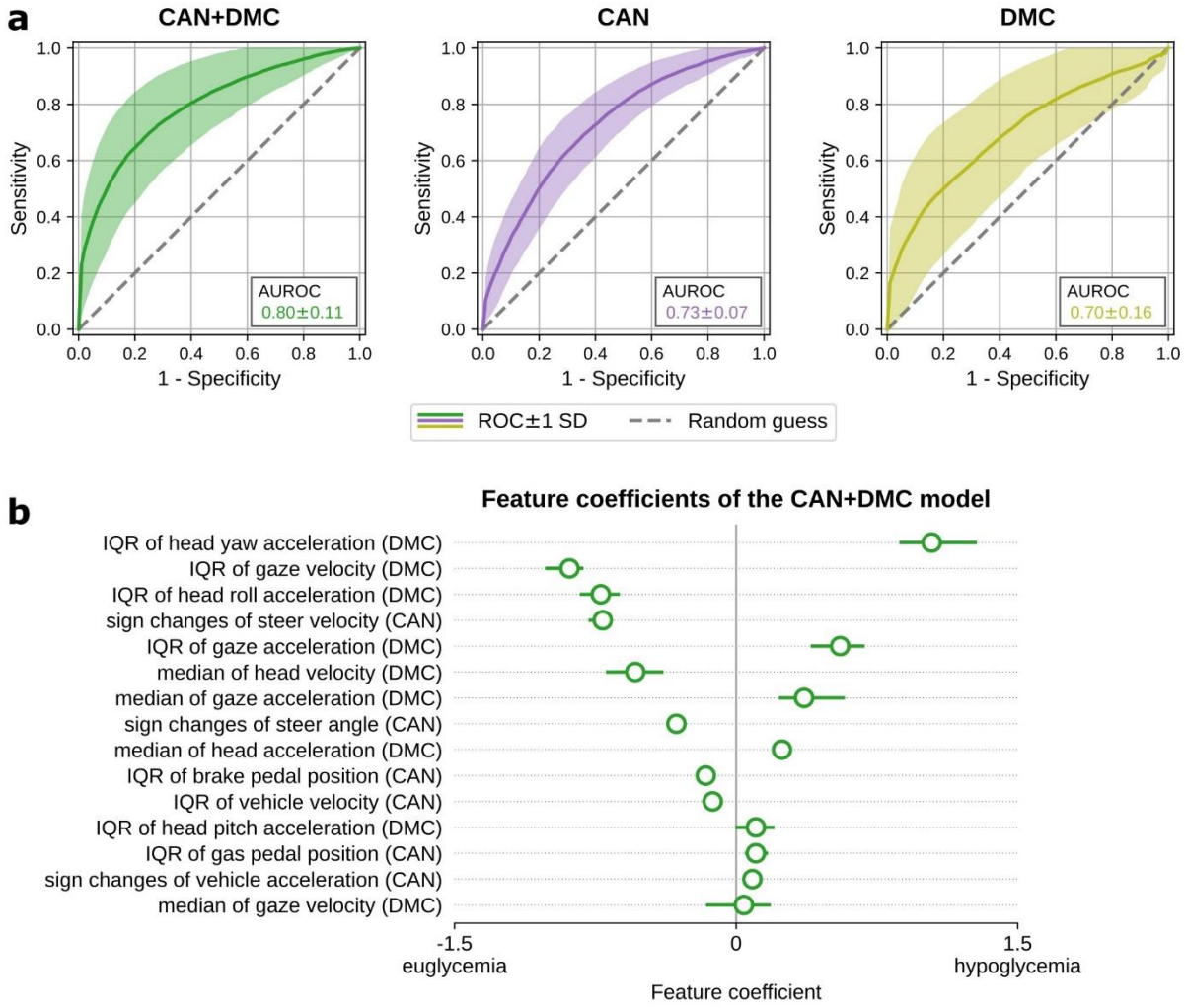

Figure 2a). The CAN model showed an AUROC of 0.73±0.07, and the DMC model achieved an AUROC of 0.70±0.16. Further performance metrics are shown in Table 2. To explain the decision-making of the CAN+DMC model, we provide the coefficients and odds ratios (ORs) of the input features in Figure 2b. Figure S4 illustrates the plots for the area under the precision-recall curve (AUPRC).

To ensure the robustness of our results, we conducted several checks including the evaluation of other (non-)linear ML models (e.g., gradient boosting decision tree, etc.), a sensitivity analysis for different window lengths, and applying various training and evaluation procedures (Tables S2–S4). Overall, the checks confirmed the choice of our ML model and the window length of 60sec used for the final analysis. Using

different training and evaluation procedures, the performance of the models remained stable across all combinations with minor decreases for training procedures involving less data (i.e., only mild hypoglycemia).

Hypoglycemic symptoms

Participants underestimated BG both in pronounced and in mild hypoglycemia. While they reported marked symptoms in pronounced hypoglycemia, differing significantly from baseline, there was no difference in symptom scores between baseline and mild hypoglycemia. Notably, 40% of the participants would keep driving in mild hypoglycemia. Table S5 reports the self-estimated BG values and self-rated hypoglycemic symptoms.

Discussion

In this paper, we present a non-invasive ML approach to detect hypoglycemia in people with diabetes during real car driving. The approach exclusively uses driving and gaze/head motion data (CAN+DMC model). Limiting the input to driving data (CAN model) or gaze/head motion data (DMC model) still resulted in considerable performance. Furthermore, interpretable ML provided insights into behavioral changes when driving in hypoglycemia.

Driving requires swift reactions to changing traffic and road conditions, relying on cognitive, executive, and motor abilities, all of which are impaired by hypoglycemia. SMBG, recommended as safety measure before driving, cannot proactively warn during a drive. Alternatively, CGM provides permanent glucose measurements but faces limitations due to invasiveness, accessibility, and accuracy, especially during hypoglycemia.[21] In the present study CGM significantly overestimated glucose values during hypoglycemia. Manual calibration might alleviate this constraint yet would not eliminate CGM's inherent time delay[6]. Raising CGM alarm thresholds could provide earlier warnings but is likely to have a negative impact on overall glycemic control.[22] Conversely, the present approach may offer a readily available, scalable, and non-invasive warning system that can potentially substitute or complement existing detection methods during driving, thereby improving traffic safety of people with diabetes. In the future, the proposed approach could be used to trigger warning messages or conversational turns delivered by in-vehicle voice assistants. Such assistants are already implemented by various car manufacturers and implementation of voice-based alert systems seems feasible. Still, future studies have to test the acceptance and effectiveness of these systems.

Consistent with earlier reports,[14] our participants substantially underestimated the degree of hypoglycemia, a majority planning to continue driving without countermeasures. These observations in well-controlled individuals and intact hypoglycemia awareness underscore the need for novel approaches to detect hypoglycemia.

The relevance of the proposed ML approach goes beyond diagnostic properties. Evaluating the coefficients of the input features provides valuable insights into behavioral alterations during hypoglycemic driving. For the driving features (CAN), we observed fewer micro-corrections of the vehicle trajectory and speed in hypoglycemia, translating into a less proactive driving style with reduced fine control. These changes were first and foremost reflected in sign changes of steer velocity (directional changes of steering wheel), sign changes of steer angle (number of times zero position of steering wheel is crossed), and interquartile range of brake pedal position (Figure 2b). Analysis of DMC data revealed more monotonous gaze/head motion behavior. However, when drivers shifted their gaze or head, they did it more abruptly, as indicated by increased acceleration features. These behavioral changes had a particular impact on interquartile range of head yaw (left-right movement) acceleration, gaze velocity, and head roll (left-right head tilt) acceleration (Figure 2b). In summary, our findings indicate that hypoglycemia is associated with a less attentive, less fine-controlled, and more hectic behavior.

In an earlier driving simulator study, we have shown that an ML-based approach using driving and eye tracker data might enable detecting hypoglycemia while driving.[10] However, simulator studies are limited regarding perceptual experience environmental conditions, perception of danger, and sensor quality. In the present study, we now developed and evaluated a non-invasive hypoglycemia detection system using signals

from commercially available automotive systems in real cars. Remarkably, detection accuracy in the simulator and the car are comparable, despite the increased complexity and external influences in real car driving. In addition, lab eye tracking hardware is superior to standard DMC systems so that we could not include gaze features such as saccades or fixations due to the limited framerate of the in-vehicle DMC. Nevertheless, these finding corroborate the robustness of the proposed approach and the potential for integrating such a detection system into automotive systems.

Earlier studies in individuals with diabetes driving in a simulator have shown that participants in hypoglycemia tend to drive off-road and across the midline.[14, 23] However, using position-based information to trigger alarms is not suitable for a preventive system in reality, as it would only detect mishaps after they have occurred. Conversely, the proposed ML approach relies on subtle alterations in driving behavior, enabling early detection and aiming to prevent such incidents. The need for novel approaches in this field is underlined by the substantial and increasing number of driving accidents related to hypoglycemia,[3, 4] with recent studies emphasizing that people with diabetes drive a significant amount of time in undetected hypoglycemia.[5]

We report consistent performance for the three presented ML models (CAN+DMC, CAN, and DMC). This offers the potential to implement our ML-based hypoglycemia warning system in different vehicle generations. The CAN+DMC model is suitable for modern cars with integrated driver monitoring cameras, while the CAN model can be implemented in current cars without DMC. Finally, the performance based on gaze/head motion data alone (DMC) supports applicability in future (semi-)automated driving.[12] There are currently no specific guidelines for a hypoglycemia detection system while driving. However, standards for drowsiness detection systems are

understood to place sensitivity for regulatory approval at 40%,[24] which our approach exceeds considerably.

The strength of our work is the interventional and prospective design providing data from real car driving and covering different hypoglycemic ranges. To our knowledge, this is the first study to report standardized hypoglycemia interventions while driving. Venous blood glucose served as the gold standard and was regularly measured confirming the consistent adherence to targeted glycemic ranges during data collection. Additionally, our approach is implementable without the need for additional sensors in the vehicle or on the body. We intentionally employed interpretable ML models following current ML principles.[25] While our ML approach was developed using data solely from individuals with type 1 diabetes, as hypoglycemia induction was ethically justifiable in this population, the concept may be applicable to other groups (e.g., type 2 diabetes), other driver states (e.g., intoxication), and/or different medical conditions, although further validation is required. The resource-intensive and complex procedures limited participant numbers, but the analysis was based on a large number of observations (47,998) due to high-resolution in-vehicle parameters and BG values, providing a robust basis for ML modeling. The male predominance in our studies limits generalization to female individuals with type 1 diabetes. The model was built on data of well-controlled and generally healthy individuals with type 1 diabetes, since hypoglycemia induction was ethically justifiable in this population. Further studies are needed to assess the impact of sex, age[7], comorbidities (e.g., severe neuropathy), and hypoglycemia unawareness on detection performance. Hypoglycemia was induced using intravenous insulin to precisely and safely regulate BG levels during the experiments, and validation in the context of naturally occurring hypoglycemia is needed. Legal restrictions in Switzerland prohibited hypoglycemia induction during

normal road traffic, necessitating the study to be conducted on a closed-off test track. The choice of driving sequence (euglycemia followed by hypoglycemia) aimed to minimize carry-over effects, as driving performance may still be impacted for up to 75 minutes after restoring euglycemia.[26] Learning bias was minimized through a prior test-driving session, and the amount of euglycemic and hypoglycemic values in the study was balanced, though not fully mirroring real-world clinical settings. This could result in a slightly higher rate of false positive warnings, which may be acceptable at this conceptual stage.

To conclude, we introduce an ML-based approach solely based on driving and gaze/head motion date to detect hypoglycemia non-invasively during car driving. The approach applies to current vehicles and anticipates future advancements in automotive technology. In addition to hypoglycemia detection, interpretable ML improves our understanding of behavioral changes during hypoglycemia.

Funding

Swiss National Science Foundation (Sinergia grant, SNF CRSII5_183569), Clinical Trials Unit Berne, Swiss Diabetes Foundation, Diabetes Center Berne, Department of Research of the University Hospital Berne, Federal Department of Defence, Civil Protection and Sport of Switzerland, and Automobile Club Switzerland.

Author contributions

VL, TZ, and MM contributed equally as first authors. FW and CS contributed equally as last authors. The following authors contributed to the conception and design of the study: VL, TZ, MM, SF, FW, and CS; acquisition of data: VL, TZ, MM, CB, MN, CA and SL; analysis of data: VL, MM, SS, MK, SF; interpretation of data: VL, TZ, MM, CB, MK, SF, TK, FW and CS; writing the manuscript: VL, TZ, MM, MK, SF, FW, and CS; all authors critically reviewed the manuscript. 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.

Data sharing

Any requests for raw data will be reviewed by the scientific study board comprising the principal investigator and senior researchers leading the involved research groups. Only applications for non-commercial use will be considered and should be sent to the corresponding author. Applications should outline the purpose of the data transfer. Any data that can be shared will need approval from the scientific study board and a Material Transfer Agreement in place. All data shared will be de-identified. The code for independent replication is available on GitHub (*link will be inserted upon acceptance*).

Declaration of interests

CB, TK, and EF are affiliated with the Centre for Digital Health Interventions, 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. Centre for Digital Health Interventions is partly funded by CSS, a Swiss health insurer. 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 declare no competing interests.

Acknowledgments

The authors thank all the participants for their time and enthusiasm. Thanks go to Joana Filipa Rodrigues Cunha Freitas (study nurse), and Vincent Vorburger (trainee) for their valuable assistance during the clinical visits. We are grateful to Mirjam Zueger for providing support in the graphics design. We thank Laura Goetschi for her administrative support. Thanks go to Daniel Haudenschild from The Federal Department of Defence, Civil Protection and Sport of Switzerland for supporting us in logistical procedures. We want to thank Anita Brechtbuehl and Stefanie Gilgen from the Automobile Club Switzerland for their support in developing the driving scenarios. Finally yet importantly, we thank the driving instructor Markus Brand for his flexibility and for keeping a good atmosphere in the car with his funny stories.

References

1. Cryer PE and Gerich JE. Glucose counterregulation, hypoglycemia, and intensive insulin therapy in diabetes mellitus. *N Engl J Med* 1985; 313: 232-241. 1985/07/25. DOI: 10.1056/nejm198507253130405.
2. Graveling AJ, Deary IJ and Frier BM. Acute hypoglycemia impairs executive cognitive function in adults with and without type 1 diabetes. *Diabetes Care* 2013; 36: 3240-3246. 2013/06/20. DOI: 10.2337/dc13-0194.
3. Cox DJ, Ford D, Gonder-Frederick L, et al. Driving mishaps among individuals with type 1 diabetes: a prospective study. *Diabetes Care* 2009; 32: 2177-2180. 2009/11/27. DOI: 10.2337/dc08-1510.
4. ECRI. Diabetes and Commercial Motor Vehicle Safety (Federal Motor Carrier Safety Administration). *June 2011 Update*. Plymouth Meeting, Pennsylvania, ECRI, 2011.
5. Merickel J, High R, Smith L, et al. Driving Safety and Real-Time Glucose Monitoring in Insulin-Dependent Diabetes. *Int J Automot Eng* 2019; 10: 34-40. 2019/01/01. DOI: 10.20485/jsaeijae.10.1_34.
6. Schmelzeisen-Redeker G, Schoemaker M, Kirchsteiger H, et al. Time Delay of CGM Sensors: Relevance, Causes, and Countermeasures. *J Diabetes Sci Technol* 2015; 9: 1006-1015. DOI: 10.1177/1932296815590154.
7. Saunders AL, Bodine C, Snell-Bergeon J, et al. Higher Prevalence of Hypoglycemia and Unsafe Driving Practices in Adults With Type 1 Diabetes. *Diabetes Care* 2023. DOI: 10.2337/dc22-2035.
8. Yu K-H, Beam AL and Kohane IS. Artificial intelligence in healthcare. *Nature Biomedical Engineering* 2018; 2: 719-731. DOI: 10.1038/s41551-018-0305-z.
9. Lehmann V, Föll S, Maritsch M, et al. Noninvasive Hypoglycemia Detection in People With Diabetes Using Smartwatch Data. *Diabetes Care* 2023. DOI: 10.2337/dc22-2290.
10. Lehmann V, Zueger T, Maritsch M, et al. Machine learning for non-invasive sensing of hypoglycemia while driving in people with diabetes. *Diabetes, Obesity and Metabolism*; n/a. DOI: https://doi.org/10.1111/dom.15021.
11. de Winter J, Leeuwen PM and Happee R. *Advantages and Disadvantages of Driving Simulators: A Discussion*. 2012.
12. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles.
13. Interior monitoring systems - How monitoring the vehicle interior increases safety, comfort, and convenience, https://www.bosch-mobility-solutions.com/en/solutions/interior/interior-monitoring-systems/ (2022, accessed 13.12.2022 2022).
14. Cox DJ, Gonder-Frederick LA, Kovatchev BP, et al. Progressive hypoglycemia's impact on driving simulation performance. Occurrence, awareness and correction. *Diabetes Care* 2000; 23: 163-170. 2000/06/27.
15. Shahidi Zandi A, Comeau FJE, Mann RE, et al. Preliminary Eye-Tracking Data as a Nonintrusive Marker for Blood Δ-9-Tetrahydrocannabinol Concentration and Drugged Driving. *Cannabis Cannabinoid Res* 2021; 6: 537-547. 2021/08/26. DOI: 10.1089/can.2020.0141.
16. Zandi AS, Quddus A, Prest L, et al. Non-Intrusive Detection of Drowsy Driving Based on Eye Tracking Data. *Transportation Research Record* 2019; 2673: 247-257. DOI: 10.1177/0361198119847985.

17. Koch K, Maritsch M, van Weenen E, et al. Leveraging driver vehicle and environment interaction: Machine learning using driver monitoring cameras to detect drunk driving. *arXiv preprint arXiv:230108978* 2023.
18. Hyland SL, Faltys M, Hüser M, et al. Early prediction of circulatory failure in the intensive care unit using machine learning. *Nature medicine* 2020; 26: 364-373. DOI: 10.1038/s41591-020-0789-4.
19. Vabalas A, Gowen E, Poliakoff E, et al. Machine learning algorithm validation with a limited sample size. *PloS one* 2019; 14: e0224365.
20. Potashner D, Brown RE, Li A, et al. Paradoxical Rise in Hypoglycemia Symptoms With Development of Hyperglycemia During High-Intensity Interval Training in Type 1 Diabetes. *Diabetes Care* 2019; 42: 2011-2014. DOI: 10.2337/dc19-0609.
21. Farrell CM, McNeilly AD, Hapca SM, et al. Real-time Continuous Glucose Monitoring During a Hyperinsulinemic-Hypoglycemic Clamp Significantly Underestimates the Degree of Hypoglycemia. *Diabetes Care* 2020; 43: e142-e143. 2020/08/21. DOI: 10.2337/dc20-0882.
22. Lin YK, Groat D, Chan O, et al. Alarm Settings of Continuous Glucose Monitoring Systems and Associations to Glucose Outcomes in Type 1 Diabetes. *J Endocr Soc* 2020; 4: bvz005. 2020/01/30. DOI: 10.1210/jendso/bvz005.
23. Cox DJ, Gonder-Frederick L and Clarke W. Driving decrements in type I diabetes during moderate hypoglycemia. *Diabetes* 1993; 42: 239-243. 1993/02/01.
24. Fredriksson R, Lenné MG, van Montfort S, et al. European NCAP Program Developments to Address Driver Distraction, Drowsiness and Sudden Sickness. *Frontiers in Neuroergonomics* 2021; 2. Perspective. DOI: 10.3389/fnrgo.2021.786674.
25. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* 2019; 1: 206-215. DOI: 10.1038/s42256-019-0048-x.
26. Zammitt NN, Warren RE, Deary IJ, et al. Delayed recovery of cognitive function following hypoglycemia in adults with type 1 diabetes: effect of impaired awareness of hypoglycemia. *Diabetes* 2008; 57: 732-736. 2007/11/28. DOI: 10.2337/db07-0695.

Figure Legends

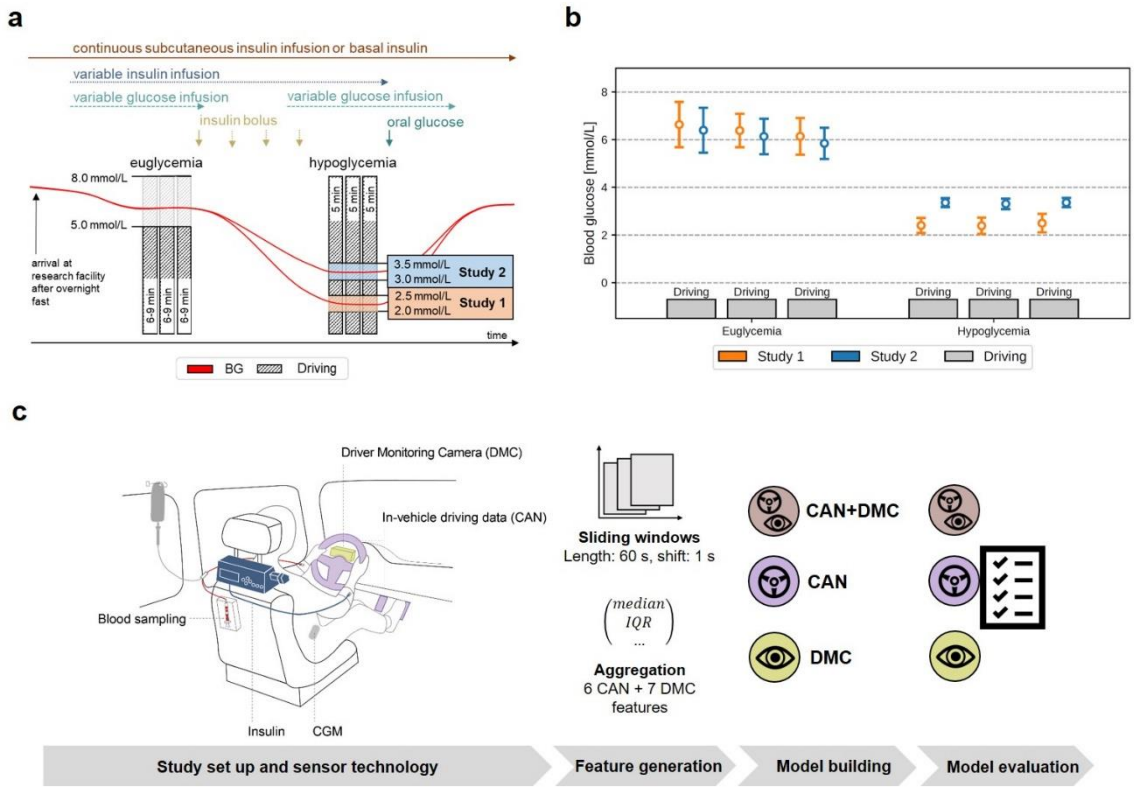

Figure 1: Overview. (a) Controlled hypoglycemia procedures for studies 1 and 2 using variable insulin and glucose administration with corresponding driving sessions in eu- and hypoglycemia. The intended range for blood glucose (BG) in hypoglycemia was 2.0–2.5 mmol/L in study 1, and 3.0–3.5 mmol/L in study 2. Driving sessions consisted of three 6 to 9-minute drives with three different driving scenarios while in-vehicle driving (CAN) and gaze/head motion (DMC) data was collected. **(b)** Venous BG in eu- and hypoglycemia for studies 1 and 2 are shown as mean (circles) with the standard deviation (whiskers). **(c)** The car, driver monitoring camera and glucose management setup in both studies is shown on the left part of the panel, while the procedure for building and evaluating our machine learning (ML) models is displayed in the right part of the panel. IQR, interquartile range.

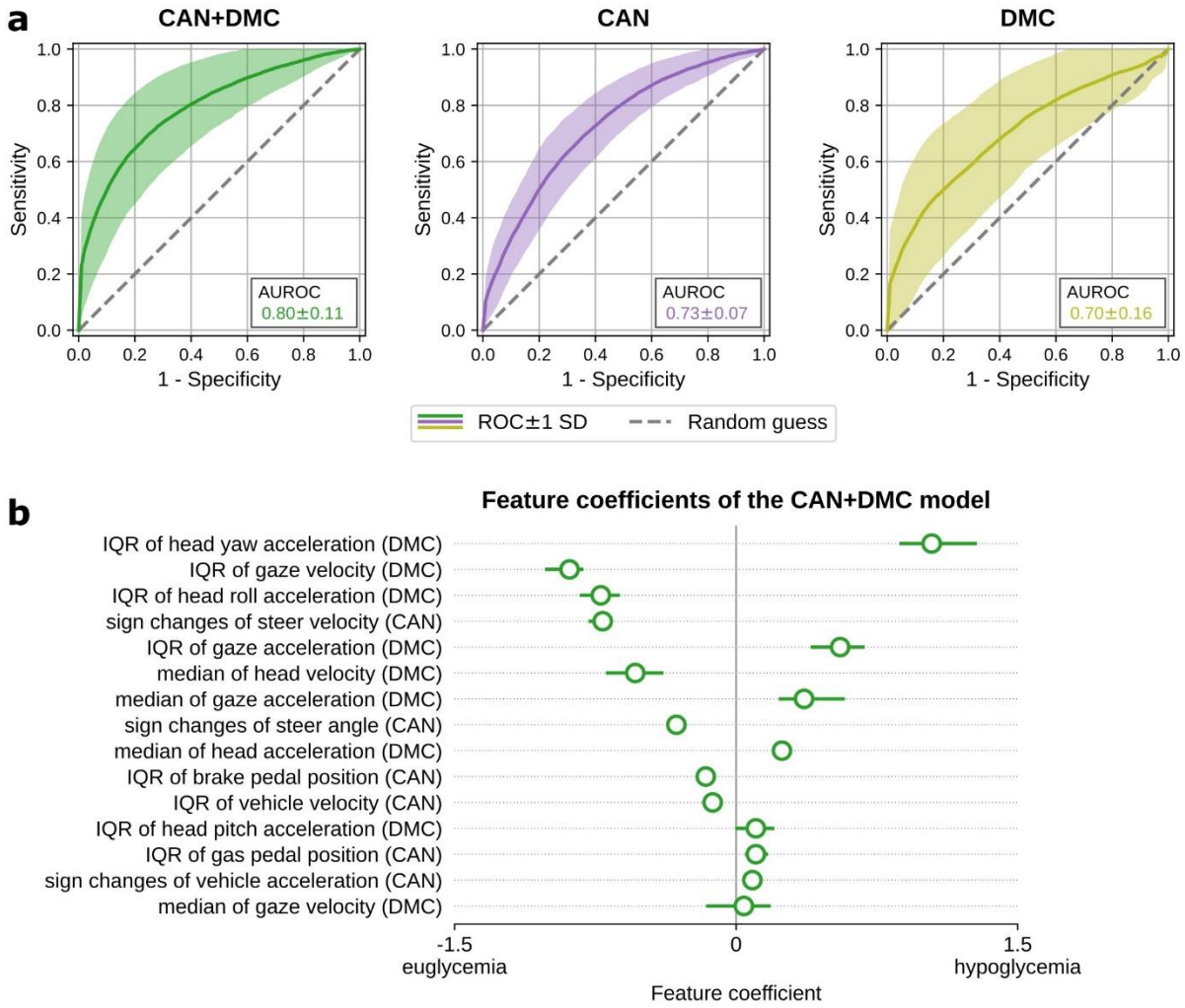

Figure 2: Machine learning (ML) detects hypoglycemia based on real car driving, and gaze/head motion data. (a) Reported is the area under the curve for the receiver operating characteristic (AUROC) to detect hypoglycemia. Here, we report the performance using combined in-vehicle driving, gaze/head motion data (CAN+DMC), driving data exclusively (CAN), and gaze/head motion data exclusively (DMC). The AUROC illustrates the mean true positive rate (=sensitivity) against the false positive rate (=1−specificity). The shaded areas illustrate the standard deviation (SD) at various thresholds across all participants. The grey dashed line is a naïve baseline corresponding to a model that has no discriminatory power and decides at random (AUROC=0.50). **(b)** For interpretability, we examine the ML model (here: logistic regression) and report the regression coefficients and odds ratios (OR) of the CAN+DMC model. The regression coefficients and ORs quantify how each feature influences the output of the model. Features with positive coefficients move the output of the ML model toward predicting hypoglycemia, whereas features with a negative coefficient move the output of the ML model toward predicting euglycemia. The feature coefficients across all cross-validation folds are reported as mean (circles) with whiskers ranging from minimum to maximum. The most important features are: interquartile range (IQR) of head yaw acceleration (coefficient: 1.04 OR 2.83), IQR of gaze velocity (coefficient: -0.89; OR 0.41), and IQR of head roll acceleration (coefficient: -0.72; OR 0.49). ROC, receiver operating characteristic.

Tables

Table 1: Baseline characteristics.

Continuous variables are shown using the mean±standard deviation. Impaired hypoglycemia awareness is indicated by a Clarke score of higher than 3 points. BMI, body mass index; CSII, continuous subcutaneous insulin infusion; HbA_{1c}, glycated hemoglobin; IU, insulin units; MDI, multiple daily injections; TDD, total daily insulin dose.

Variable	Overall (*n*=30)	Study 1 (*n*=20)	Study 2 (*n*=10)
Age [years]	40.1±10.3	41.6±9.8	37.2±11.1
Sex	24 male, 6 female	17 male, 3 female	7 male, 3 female
Insulin treatment	12 MDI, 18 CSII	9 MDI, 11 CSII	3 MDI, 7 CSII
Weight [kg]	84.1±16.4	83.1±14.0	86.0±21.0
Height [m]	1.76±0.09	1.77±0.07	1.76±0.13
BMI [kg/m]	26.8±3.8	26.5±3.7	27.3±4.0
TDD [IU/day/kg]	0.59±0.20	0.59±0.15	0.59±0.29
HbA_{1c} [%]	6.9±0.7	6.8±0.70	7.1±0.8
HbA_{1c} [mmol/mol]	51.9±8.0	50.8±7.7	54.3±8.4
Clarke score >3	0 / 30	0 / 20	0 / 10
Peripheral neuropathy	3 / 30	1 / 20	2 / 10
Diabetes duration [years]	20.8±12.1	20.8±11.1	20.6±14.5
Driving experience [years]	21.8±11.5	24.7±10.3	16.1±12.1
Kilometers driven per year [km/year]	12,353±12,420	13,645±10498	9,770±15,914

Table 2: Machine learning (ML) performance metrics.

Reported are the performance metrics of the three ML models as mean±standard deviation based on combined driving, gaze/head motion data (CAN+DMC), driving data exclusively (CAN), and gaze/head motion data exclusively (DMC) detection. AUPRC, area under the precision-recall curve; BACC, balanced accuracy; F1, F1-score; MCC, Matthews correlation coefficient.

	AUROC	AUPRC	BACC	F1	MCC	Sensitivity	Specificity
CAN+DMC	0.80±0.11	0.79±0.12	0.74±0.09	0.72±0.12	0.48±0.17	0.70±0.15	0.78±0.10
CAN	0.73±0.07	0.71±0.09	0.68±0.06	0.69±0.07	0.37±0.11	0.72±0.11	0.65±0.12
DMC	0.70±0.16	0.70±0.16	0.68±0.10	0.66±0.17	0.38±0.20	0.66±0.22	0.70±0.23