GluPredKit: Development and User Evaluation of a Standardization Software for Blood Glucose Prediction

Miriam K. Wolff[1][0009−0000−6919−8164], Sam Royston[2][0009−0003−8623−4042], Anders Lyngvi Fougner[3][0000−0001−7583−5027], Hans Georg Schaathun[1], Martin Steinert[4][0000−0002−8366−0201], and Rune Volden[1][0009−0005−0671−325X]

[1] Department of ICT and Natural Sciences, NTNU, 6009 Ålesund, Norway
[2] Replica Health, https://www.replica.health/
[3] Department of Engineering Cybernetics, NTNU, 7034 Trondheim, Norway
[4] Department of Mechanical and Industrial Engineering, NTNU, 7034 Trondheim, Norway

Abstract. Blood glucose prediction is an important component of biomedical technology for managing diabetes with automated insulin delivery systems. Machine learning algorithms hold the potential to advance this technology. However, the lack of standardized methodologies impedes direct comparisons of emerging algorithms. The purpose of this study is to address this challenge by developing a software platform designed to standardize the training, testing and comparison of blood glucose prediction algorithms. First, we design and implement the software guided by the current literature. To ensure the platform's user-friendliness, we conducted preliminary testing and a user study. In this study, four participants interacted with the software and provided feedback through the System Usability Scale (SUS) and open-ended questions. The result of the study was the software GluPredKit, which features a modular, open-source architecture, complemented by a command-line interface, comprehensive documentation, and a video tutorial to enhance usability. The user study indicates that GluPredKit offers high usability, facilitating comparisons between different algorithms. Future directions include continuously enhancing the software based on user feedback. We also invite community contributions to further expand GluPredKit with state-of-the-art components and foster a collaborative effort in standardizing blood glucose prediction research.

Keywords: Software · Benchmarking · Blood glucose prediction · Machine learning · Physiological modelling.

1 Introduction

1.1 Background

Diabetes Mellitus is a global health challenge. In 2021, there was approximately 10.5%, or 537 million, of adults between 20-79 years living with diabetes. The International Diabetes Federation predicts that the number will increase to 12.2%

by 2045 [15]. About 8 to 9 million have type 1 diabetes (T1D) [9]. Inadequate glycaemic control over time can damage the heart, blood vessels, eyes, kidneys, and nerves, increasing the probability of several health complications. Strengthening the glycaemic regulation for these patients could improve their quality of life, cut healthcare costs, and boost diabetics' workforce participation, highlighting the need for ongoing innovation in this field.

Recent advancements in technologies like Continuous Glucose Monitoring (CGM) devices and insulin pumps have improved glucose control, yet they require manual inputs like meal announcements and carbohydrate counting [31]. The challenge lies in the slow onset of subcutaneous insulin and its high variability among individuals [8]. Insulin and blood glucose (BG) dynamics are influenced by factors like meals, sleep, exercise, stress, and the menstrual cycle [4][13]. Non-diabetics may also exhibit significant variability in BG levels and insulin responses, implying that stabilizing glucose fluctuations could benefit them as well [10].

Despite technological advancements, managing diabetes remains complex. Current systems, while advanced, fall short of providing fully automated and personalized diabetes management solutions. More sophisticated BG prediction algorithms that can adapt to individual variations and real-life conditions could be used to develop enhanced artificial pancreas or decision support systems, which could alleviate the burden of diabetes management and improve patient outcomes [7].

There is a need for standardization in the development and evaluation of these predictive models [16]. The current landscape is marked by diverse approaches and methodologies, leading to challenges in comparing and validating different models. Although numerous studies claim robust algorithms for BG prediction [2][32][14], there is a recurring limitation of validation on identical datasets [11]. Our study aims to address the need for a standardized evaluation framework. By focusing on the implementation of shared code, we seek to enhance the comparability of BG prediction algorithms in future research.

1.2   Related Work

Historically, BG prediction has relied on physiological models. These models take into account various factors, including insulin absorption [27] and meal models [22]. Additionally, some models for accounting for physical activity are proposed, using inputs such as heart rate or oxygen consumption [17]. The advent of wearable sensors and smart devices has recently shifted the focus towards machine learning approaches [24].

Despite these advancements, the field faces critical model comparability and validation challenges. Jacobs et al. highlight the diversity in methodologies and datasets, which complicates direct comparisons between different studies [16]. Initiatives such as the BG level prediction challenges in International Workshops on Knowledge Discovery in Healthcare Data (KDH) in 2018 and 2020 have addressed these challenges by providing standardized datasets and evaluation criteria [20][19].

Standardization efforts, while valuable, often overlook the viability in real-life scenarios. For instance, the commonly used 30- and 60-minute prediction horizons may not capture the longer-term effects of meals and insulin. Moreover, more open code and detailed implementation information are needed to ensure progress in model validation and adaptability. A participant in the KDH workshop pointed out the restrictive nature of excluding what-if scenarios in model training and evaluations [18], which are vital for developing decision-support algorithms [5]. In addition, what-if events are essential for model predictive control (MPC) because MPC relies on evaluating counterfactual actions in the future [26].

1.3   Study Objectives

This study aims to enhance the reproducibility, training, and comparative analysis of BG prediction algorithms.

This study aims to develop and evaluate a software for facilitating future research and comparison of BG prediction algorithms. We implement an open-source platform that integrates various preprocessing techniques, BG prediction methodologies, evaluation metrics, and plots. Figure 1 offers a high-level illustration of the GluPredKit ecosystem. The platform offers a standardized approach to facilitate data processing, algorithm training, evaluation, and comparison. It supports fetching data from prevalent diabetes data sources, ensuring simplicity when applying new data to existing model implementations. We assess GluPredKit's usability with a user study employing the System Usability Scale (SUS) and open-ended questions involving a cohort of four participants.

BG prediction models that are trained using GluPredKit can be integrated into practical applications in diabetes management, like software applications for real-time predictions [28], applications providing decision support, or in MPC or as an additional component for other control strategies like PID or reinforcement learning [12].

GluPredKit addresses the limitations of existing BG prediction packages, which often focus on single models with varying implementations. Our software provides a unified platform for comparing and evaluating BG prediction algorithms, streamlining the assessment of model improvements. Designed for versatility, GluPredKit accommodates diverse dataset features, hypothetical scenarios, and provides detailed visualizations, such as plots and predicted trajectories, alongside evaluation metrics. Additionally, it functions both as a standalone package and as an integrable dependency.

1.4   Paper Structure

The "Methods" section outlines GluPredKit's system evaluation criteria, including the design and definition of requirements. This is followed by a description of the user study design and participant profiles. In the "Results" section, we present the technical compliance of GluPredKit with established requirements and the findings from the user study. The "Discussion" section interprets these
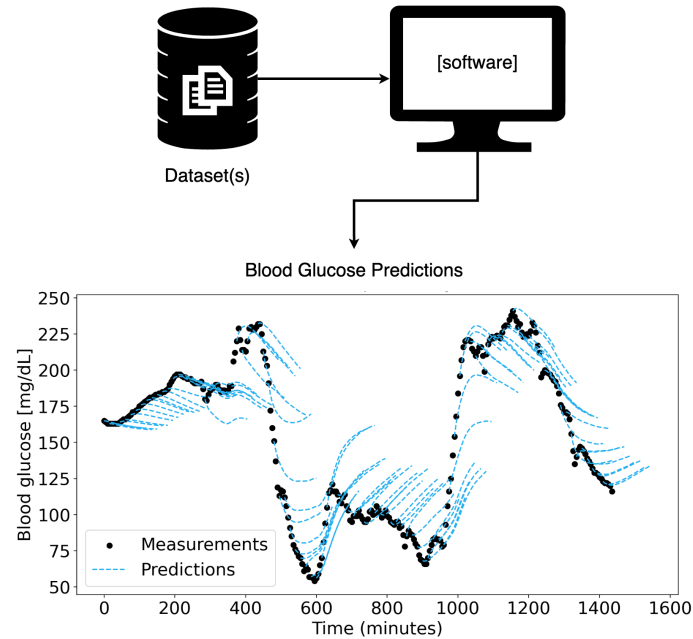
Fig. 1. High-level visualization of the GluPredKit ecosystem. The upper image symbolize data storage and how GluPredKit acquires its data, while the bottom graph exemplifies the potential output of BG predicted trajectories in GluPredKit. The trajectories represent predictions for 120 minutes.

results, considering study limitations and suggesting directions for future research. The paper concludes with the "Conclusions" section, which summarizes key findings and discusses their significance in the field of BG prediction.

2   Methods

In designing our study, we followed a model for the research process inspired by Oates et al. [23, p. 34]. Our study started with establishing the objectives and design of the GluPredKit platform, drawing from current literature, experience, and logical reasoning, as detailed in 2.1. Subsequently, we implemented the software according to these specifications and conducted preliminary tests to ensure its functionality. For validation purposes, we engaged users in testing the platform, enabling them to explore its different features. The method for data collection was a questionnaire incorporating the SUS and open-ended questions. Finally, we analyzed the questionnaire responses to evaluate the platform's effectiveness.

2.1   Design and Implementation

The complete list of our implemented components in each module in GluPredKit is maintained in the repository's documentation and will always be kept up to date with the latest changes. In deciding which data sources to integrate into the parsers module, we prioritized sources containing several relevant data types so that each source could provide a complete dataset. We identified various sources, including web APIs and data exports that integrate with a broad range of diabetes equipment, as well as the benchmark dataset Ohio T1DM.

We used the identified prediction models from a comprehensive benchmark study by Xie et al. [30]. They identified commonly used models based on linear- and nonlinear machine-learning regression techniques and deep-learning algorithms. The implemented components in GluPredKit can provide as a benchmark for new models presented in the future.

Recently, Jacobs et al. did a consensus study about evaluating BG prediction algorithms, which formed the basis for our decided implemented evaluation metrics [16]. RMSE, often a primary metric in conjunction with relative error measures like MAE, is a standard evaluation approach in machine learning. However, these metrics may fall short in capturing the clinical significance of BG predictions, as they uniformly penalize errors across the glucose range. To address this limitation, the glucose-specific RMSE has been proposed [6] [16]. Moreover, Clarke or Parkes error grid analysis, initially designed for assessing BG measurements [1], has evolved into widely adopted metrics for evaluating the clinical applicability of BG prediction algorithms. These error grid zones are defined based on actual measurements rather than predictions in future time but serve as a valuable supplementary tool for extending the assessment using existing metrics.

Our scatter plot, aligning with prior studies [30], compares predicted and measured values. In addition, we introduce two novel plots: The graph in the bottom of Figure 1 is an example of a plot of predicted BG trajectories generated using GluPredKit. The black dots indicate CGM measurements, while the dashed lines from each measurement point demonstrate the model's predictions over time. Forecasted trajectories over time, in contrast to single-horizon predictions, are crucial for MPC [26]. MPC is a common control approach in artificial pancreas systems [21]. The last plot focuses on a single predicted trajectory, illustrating real-time responses to events like meals or insulin injections, offering insights into the model's adaptability and clinical potential.

2.2   User Feedback

Pilot Testing To ensure the robustness and efficacy of our software and documentation, we undertook multiple rounds of pilot testing. In an iterative process, this involved the authors and colleagues until we ascertained a seamless system operation devoid of errors or ambiguities. Participants engaged in a user study which required them to explore the software's features and complete a survey. Feedback garnered during these sessions, especially initial impressions, proved
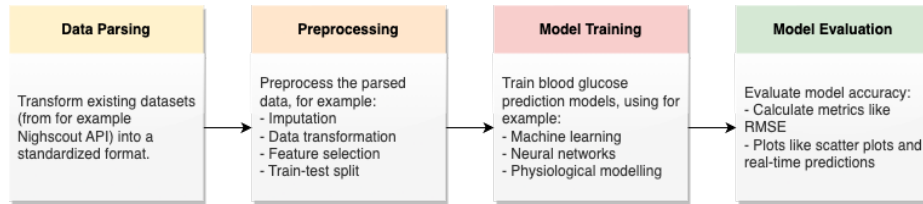
Fig. 2. Overview of the GluPredKit pipeline. This diagram delineates the sequential stages involved in the BG prediction process, starting from data parsing from existing datasets, through preprocessing and model training, culminating in model evaluation and real-time predictions.

instrumental in implementing enhancements to both the software and its supporting documents.

User Testing The user testing phase assessed the platform's usability, architectural integrity, and overall efficacy. The test was designed to ascertain whether GluPredKit meets the needs of new researchers in the field by balancing user-friendliness with essential functionality. Participants underwent a three-step process: (1) Engage with GluPredKit's functionalities, (2) Complete the SUS questionnaire, and (3) Respond to open-ended questions for additional feedback.

Participants. The study involved four participants, selected based on their interest in BG prediction and proficiency in coding. None had prior exposure to GluPredKit. Although the sample size may appear limited for the quantitative evaluation of SUS scores, the specificity of our inclusion criteria naturally limits the participant pool. Furthermore, the open-ended questions were expected to yield some insights into the software's effectiveness despite the smaller sample size.

Part 1 - Testing GluPredKit Functionalities Participants received instructions to explore various features of GluPredKit, utilizing a sample dataset in .csv format. The usability test included tasks such as data preprocessing, algorithm training, testing, and comparison, all performed using the platform's command-line interface.

Part 2 - SUS The SUS, a widely recognized industry standard for usability assessment, consists of ten questions [3]. Participants rated each question on a scale of 1 to 5.

Part 3 - Open-Ended Questions This section aimed to gather insights into participants' background knowledge and perceptions of the platform's flexibility, particularly regarding the code architecture. Questions also focused on the adequacy of documentation, usability, and the most valuable features encountered. Participants were asked to hypothetically estimate time savings compared to

other software, acknowledging the absence of similar existing software for direct comparison. Finally, open-ended questions solicited feedback on potential improvements and the anticipated impact of GluPredKit, informing future development and gauging general interest in the software.

3   Results

3.1   GluPredKit - System Description

GluPredKit consists of an open-source Python repository that facilitates the whole pipeline of collecting new data, preprocessing the data, training a model, and evaluating and comparing prediction models with statistical and graphical representations. The pipeline is illustrated in Figure 2. A CLI is implemented for users to interact with the system. The code is uploaded to PyPi to be installed with only one command.

Detailed installation guidelines can be accessed from the PyPi documentation (https://pypi.org/project/GluPredKit/) or the project's GitHub repository (https://github.com/miriamkw/GluPredKit).

The software is accompanied by a video tutorial to enhance its usability (https://www.youtube.com/watch?v=GMu_Om1gTsk).

System Architecture GluPredKit's system architecture is a collection of modules specifically tailored to a facet of the BG prediction pipeline, as visualized in Figure 2. The architecture is divided into distinct modules: parsers, preprocessors, models, metrics, and plots. Each of these has a foundational base model from which specific classes derive. These foundational classes define function requirements and expected input and output formats.

Command Line Interface As illustrated in Figure 3, GluPredKit provides a CLI that includes a suite of commands for managing the end-to-end data processing and model evaluation workflow. The figure also depicts the file structure in GluPredKit and the file flow for organizing the processed data, model outputs, and evaluation results.

The GluPredKit's functionality is initiated by the `setup_directories` command, which constructs a data folder with necessary subfolders. Then, the parse command can be called, producing a raw dataset and storing it within the `data/raw` directory, as shown in the figure. This dataset forms the foundation for subsequent commands. For instance, the `generate_config` command relies on the raw dataset to create a preprocessing and training configuration. Each step in the workflow corresponds with specific directories and files within the system's file structure, ensuring an organized and efficient data management environment.
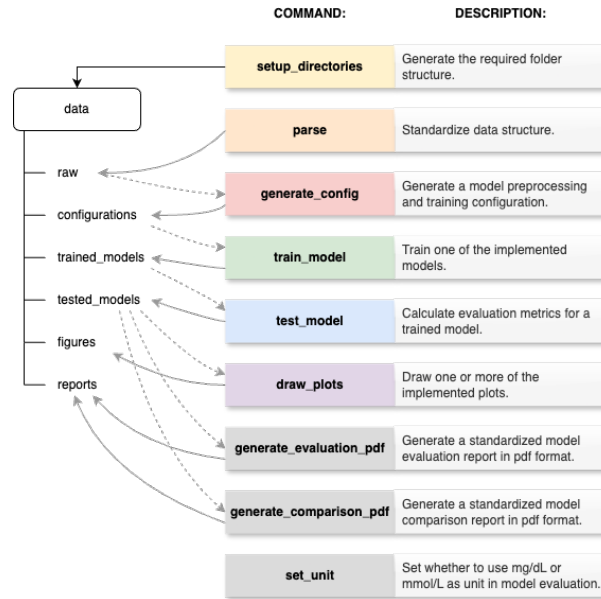
Fig. 3. Overview of the CLI commands and how they interact with the file structure in GluPredKit.

Implementation Details GluPredKit has primarily been developed using Python. We have utilized GitHub for version control and for hosting the code in an open-source manner. GluPredKit is available on PyPi, facilitating the use of the CLI directly or integrating GluPredKit as a sub-component within another system. The CLI is implemented using the click library.

In our endeavor to implement BG prediction models, we have used multiple libraries. For instance:

– Vanilla Long Short-Term Memory
– Keras, built atop TensorFlow, facilitates training our Deep Neural Networks (DNN), specifically Long Short-Term Memory and Temporal Convolutional Network models.
– PyTorch, also a DNN library.
– XGBoost is employed for gradient-boosting trees.
– For the remaining machine learning models, we have utilized scikit-learn.

3.2   User Test Results

SUS Results Figure 4 visualizes the SUS scores for each participant, revealing an impressive average of 86 (SD = 9). Interpreting these scores, as per [25], our system garners an 'A' grade in usability on average, though one participant falls into the 'B' range. Descriptively, these scores align with 'Best Imaginable' to
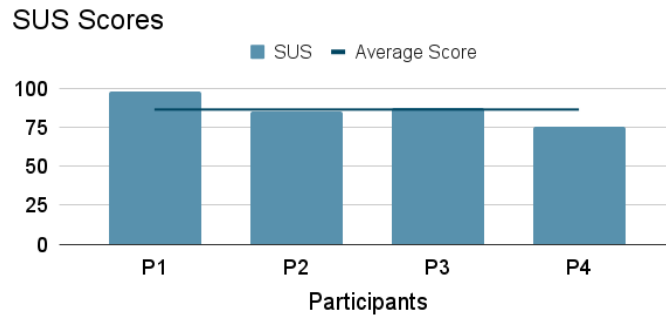
Fig. 4. Bar chart of individual SUS scores alongside a line indicating the average SUS score.

'Good' user experiences, indicating high usability satisfaction among the participants.

Open-Ended Questionnaire Insights

Feedback on Software and Documentation Participants generally found the documentation well detailed, although some noted challenges in data parsing and error feedback. Suggestions for improvement included more intuitive data parsing guides and more apparent error indications.

Recommendations for System Enhancement Respondents recommended further documentation enhancements, flexibility in DNN configuration, expanding data source compatibility, and adding automated hyperparameter tuning. Some suggested developing a user interface to broaden accessibility beyond technically proficient users.

Impact of GluPredKit Participants anticipated significant time savings using GluPredKit, with estimations like "10x longer without it" for comparable tasks. They emphasized GluPredKit's potential in advancing diabetes management through its standardized approach and the possibility of community-driven improvements and educational benefits. Highlights include:

> "Its standardized approach and ease of integrating and comparing new models can significantly enhance prediction accuracy and efficiency in diabetes management."

and

> "The most impressive part is that it encourages code contributions and the establishment of a community around the software for continuous improvement. [...] I also think it has an educational value. The software, through its documentation and examples, can serve as an educational resource for individuals entering the field of BG prediction [...]."

4   Discussion

In this study, we developed GluPredKit, a specialized software for BG prediction algorithm comparison and evaluation. The platform has integrated a wide range of predictive methodologies and algorithms. Community-contributions are welcomed, and instructions for contributions are outlined in the GluPredKit documentation (https://github.com/miriamkw/GluPredKit?tab=readme-ov-file# contributing-with-code).

We evaluated the usability of GluPredKit through user feedback. Four participants engaged with all features of GluPredKit and then provided feedback through a questionnaire. The SUS was chosen for its simplicity and widespread use in assessing software usability across diverse fields. Given the small sample size of four participants, the SUS was supplemented with open-ended questions to gather more in-depth qualitative feedback. The software's usability was quantitatively affirmed by an average SUS score of 86 (SD = 9), indicating high user satisfaction. The study's qualitative feedback from open-ended questions underscored GluPredKit's effectiveness in addressing standardization challenges and its educational benefits, enriched by comprehensive documentation. Participants also offered constructive suggestions for further enhancements, such as developing a graphical user interface, refining data parsing documentation, and expanding the modular framework to include more components.

The field of BG prediction, as noted by Jacobs et al. [16], faces a challenge due to the need for uniformity in methodologies, evaluation metrics, and data usage. This inconsistency hampers research progression, as new models are often compared against varying standards. GluPredKit addresses these issues by promoting open-source code, standardizing critical stages like data parsing and preprocessing, and establishing consistent evaluation metrics.

There are currently no established requirements or standards for measuring the accuracy of blood glucose predictions in terms of clinical safety. Our software includes a CLI command for generating detailed evaluation reports on various aspects of the models, adhering to the recommendations outlined in the consensus paper by Jacobst et al. [16]. In an earlier study [29], we outlined several recommendations for improving model development and evaluation. However, the lack of standardized requirements highlights an important area for future research.

If widely adopted, GluPredKit has the potential to streamline the benchmarking of new algorithms against established ones. Its modular architecture, designed around a "plug-and-play" framework, ensures compatibility with various datasets, models and evaluation metrics. GluPredKit is intended for collaborative efforts, where researchers can employ GluPredKit for iterative refinement of their predictive models. This collaborative effort involves conducting evaluations across diverse datasets, and making incremental improvements. If a researcher notices that their model consistently outperforms others compared to existing benchmarks, they can submit a pull request, contributing their enhanced model to the academic community. This iterative and collaborative approach could el-

evate the quality of predictive models, where GluPredKit serves as a hub for benchmarking within the academic landscape.

However, our study has limitations, notably the small number of participants in the user study, which may only partially capture the broader range of potential users. Furthermore, replicating existing studies poses inherent challenges, although our results closely align with established findings. A key design goal for GluPredKit was to balance flexibility with simplicity, avoiding unnecessary complexity.

The use of sensitive health data in our research involves important ethical considerations regarding data privacy. To address these concerns, we strictly adhere to the stipulations outlined in the Data User Agreements associated with the dataset. In addition, we have developed and implemented an internal protocol for safe data handling, in collaboration with our institution's research data support service. This protocol includes minimizing data copies, avoiding automatic synchronization servers, employing secure VPN connections, enforcing strong password protection, and ensuring the prompt deletion of data following analysis.

5   Conclusion

This study presents the design, development, and user evaluation of GluPredKit, a software platform aimed at standardizing BG prediction and facilitating comparative analysis of both existing and future algorithms. Our contributions include the implementation of a standardized pipeline for data acquisition, preprocessing, model training, and evaluation, along with the integration of recent evaluation guidelines. We also developed a CLI and created a video tutorial to support the community in adopting and extending the platform. Additionally, we conducted pilot testing and a user study, using SUS score analysis and open-ended questions to gain qualitative insights into the user experience.

While the small sample size of four participants limits the generalizability of the user study, this testing provided valuable feedback for future enhancements. Key opportunities for future work include addressing the lack of established standards for evaluating the clinical safety of BG predictions and implementing the usability improvements suggested by the participants. Moving forward, our goal is to integrate data-driven BG prediction into clinical practice, ultimately enhancing diabetes care and outcomes.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Benesch, C.: Error grid analysis (2017), https://blog.profil.com/blog/error-grid-analysis

2. Bevan, R., Coenen, F.: Experiments in non-personalized future blood glucose level prediction. The University of Liverpool Repository (2020), https://livrepository.liverpool.ac.uk/3100649/

3. Brooke, J.: Sus: A quick and dirty usability scale (1996)

4. Cinar, A., Turksoy, K.: Advances in Artificial Pancreas Systems Adaptive and Multivariable Predictive Control. Springer (2018). https://doi.org/https://doi.org/10.1007/978-3-319-72245-0

5. Contador, S., Colmenar, J.M., Garnica, O., Velasco, J.M., Hidalgo, J.I.: Blood glucose prediction using multi-objective grammatical evolution: analysis of the "agnostic" and "what-if" scenarios. Genetic Programming and Evolvable Machines **23**(2), 161–192 (Jun 2022). https://doi.org/10.1007/s10710-021-09424-6

6. Del Favero, S., Facchinetti, A., Cobelli, C.: A Glucose-Specific Metric to Assess Predictors and Identify Models. IEEE Transactions on Biomedical Engineering **59**(5), 1281–1290 (May 2012). https://doi.org/10.1109/TBME.2012.2185234, http://ieeexplore.ieee.org/document/6135492/

7. Dovc, K., Battelino, T.: Evolution of Diabetes Technology. Endocrinology and Metabolism Clinics **49**(1), 1–18 (Mar 2020). https://doi.org/10.1016/j.ecl.2019.10.009, https://www.endo.theclinics.com/article/S0889-8529(19)30091-X/fulltext, publisher: Elsevier

8. Gingras, V., Taleb, N., Roy-Fleming, A., Legault, L., Rabasa-Lhoret, R.: The challenges of achieving postprandial glucose control using closed-loop systems in patients with type 1 diabetes. Diabetes, Obesity and Metabolism **20**(2), 245–256 (2018). https://doi.org/10.1111/dom.13052, https://onlinelibrary.wiley.com/doi/abs/10.1111/dom.13052, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/dom.13052

9. Gregory, G.A., Robinson, T.I.G., Linklater, S.E., Wang, F., Colagiuri, S., Beaufort, C.d., Donaghue, K.C., Magliano, D.J., Maniam, J., Orchard, T.J., Rai, P., Ogle, G.D.: Global incidence, prevalence, and mortality of type 1 diabetes in 2021 with projection to 2040: a modelling study. The Lancet Diabetes & Endocrinology **10**(10), 741–760 (Oct 2022). https://doi.org/10.1016/S2213-8587(22)00218-2, https://www.thelancet.com/journals/landia/article/PIIS2213-8587(22)00218-2/fulltext, publisher: Elsevier

10. Hall, H., Perelman, D., Breschi, A., Limcaoco, P., Kellogg, R., McLaughlin, T., Snyder, M.: Glucotypes reveal new patterns of glucose dysregulation. PLOS Biology **16**(7), e2005143 (Jul 2018). https://doi.org/10.1371/journal.pbio.2005143, https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.2005143, publisher: Public Library of Science

11. Hameed, H., Kleinberg, S.: Comparing Machine Learning Techniques for Blood Glucose Forecasting Using Free-living and Patient Generated Data. Proceedings of Machine Learning Research **126**, 871–894 (Aug 2020)

12. Hernández, T., Ángel, M.: Glucose Regulation for In-Silico Type 1 Diabetes Patients Using Reinforcement Learning. Ph.D. thesis, UiT Norges arktiske universitet (May 2021), https://munin.uit.no/handle/10037/20861, accepted: 2021-04-13T13:28:17Z Publisher: UiT Norges arktiske universitet

13. Herranz, L., Saez-De-Ibarra, L., Hillman, N., Gaspar, R., Pallardo, L.: Glycemic changes during menstrual cycles in women with type 1 diabetes. Medicina Clinica **146**(7), 287–291 (2016). https://doi.org/10.1016/j.medcli.2015.11.044

14. Idriss, T.E., Idri, A., Abnane, I., Bakkoury, Z.: Predicting Blood Glucose using an LSTM Neural Network. In: 2019 Federated Conference on Com-

puter Science and Information Systems (FedCSIS). pp. 35–41 (Sep 2019). https://doi.org/10.15439/2019F159, iSSN: 2300-5963

15. International Diabetes Federation: Idf diabetes atlas. https://www.diabetesatlas.org (2021)

16. Jacobs, P.G., Herrero, P., Facchinetti, A., Vehi, J., Kovatchev, B., Breton, M., Cinar, A., Nikita, K., Doyle, F., Bondia, J., Battelino, T., Castle, J.R., Zarkogianni, K., Narayan, R., Mosquera-Lopez, C.: Artificial intelligence and machine learning for improving glycemic control in diabetes: best practices, pitfalls and opportunities. IEEE reviews in biomedical engineering **PP** (Nov 2023). https://doi.org/10.1109/RBME.2023.3331297

17. Jaloli, M., Cescon, M.: Modeling Physical Activity Impact on Glucose Dynamics in People with Type 1 Diabetes for a Fully Automated Artificial Pancreas (Jul 2023), http://arxiv.org/abs/2307.07923, arXiv:2307.07923 [cs, eess, stat]

18. Joedicke, D., Garnica, O., Kronberger, G., Colmenar, J.M., Winkler, S., Velasco, J.M., Contador, S., Hidalgo, J.I.: Analysis of the performance of Genetic Programming on the Blood Glucose Level Prediction Challenge 2020. In: CEUR Workshop Proceedings (2020)

19. Knowledge Discovery in Healthcare Data: Blood glucose level prediction challenge. 5th International Workshop on Knowledge Discovery in Healthcare Data (2020), https://sites.google.com/view/kdh-2020/bglp-challenge

20. Marling, C., Bunescu, R.: The ohiot1dm dataset for blood glucose level prediction: Update 2020. CEUR Workshop Proceedings **2675**, 71–74 (Sep 2020), https://pubmed.ncbi.nlm.nih.gov/33584164/

21. Moon, S.J., Jung, I., Park, C.Y.: Current Advances of Artificial Pancreas Systems: A Comprehensive Review of the Clinical Evidence. Diabetes & Metabolism Journal **45**(6), 813–839 (Nov 2021). https://doi.org/10.4093/dmj.2021.0177, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8640161/

22. Mørch-Thoresen, E.: Review of the Effect of Meals on the Glucose-Insulin Regulatory System. Master's thesis, NTNU (2021), https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2781065, accepted: 2021-09-23T18:53:34Z

23. Oates, B.J., Griffiths, M., McLean, R.: Researching Information Systems and Computing. Sage Publications (2005), https://us.sagepub.com/en-us/nam/researching-information-systems-and-computing/book272083

24. Rodríguez-Rodríguez, I., Rodríguez, J.V., Campo-Valera, M.: Applications of the Internet of Medical Things to Type 1 Diabetes Mellitus. Electronics **12**(3), 756 (Jan 2023). https://doi.org/10.3390/electronics12030756, https://www.mdpi.com/2079-9292/12/3/756, number: 3 Publisher: Multidisciplinary Digital Publishing Institute

25. Sauro, J.: 5 ways to interpret a sus score. MeasuringU (2018), https://measuringu.com/interpret-sus-score/, accessed: 2023-09-19

26. Schwenzer, M., Ay, M., Bergs, T., Abel, D.: Model predictive control: a review of its applications in supply chain management. International Journal of Advanced Manufacturing Technology (2021). https://doi.org/10.1007/s00170-021-07682-3, https://link.springer.com/article/10.1007/s00170-021-07682-3

27. Wilinska, M., Chassin, L., Schaller, H., Schaupp, L., Pieber, T., Hovorka, R.: Insulin kinetics in type-1 diabetes: continuous and bolus delivery of rapid acting insulin. IEEE Transactions on Biomedical Engineering **52**(1), 3–12 (Jan 2005). https://doi.org/10.1109/TBME.2004.839639, https://ieeexplore.ieee.org/document/1369583, conference Name: IEEE Transactions on Biomedical Engineering

28. Wolff, M.K., Schaathun, H.G., Fougner, A.L., Steinert, M., Volden, R.: Mobile software development kit for real time multivariate blood glucose prediction. IEEE Access **12**, 5910–5919 (2024). https://doi.org/10.1109/ACCESS.2024.3349496
29. Wolff, M.K., Steinert, M., Fougner, A.L., Oh, D., Årsand, E., Volden, R.: Real-time blood glucose prediction reveals a discrepancy between performance metrics and real-world evaluations. In: 2024 IEEE International Conference on Real-time Computing and Robotics (RCAR). pp. 570–575 (2024). https://doi.org/10.1109/RCAR61438.2024.10671342
30. Xie, J., Wang, Q.: Benchmarking Machine Learning Algorithms on Blood Glucose Prediction for Type I Diabetes in Comparison With Classical Time-Series Models. IEEE Transactions on Biomedical Engineering **67**(11), 3101–3124 (Nov 2020). https://doi.org/10.1109/TBME.2020.2975959, conference Name: IEEE Transactions on Biomedical Engineering
31. Zhou, K., Isaacs, D.: Closed-Loop Artificial Pancreas Therapy for Type 1 Diabetes. Current Cardiology Reports **24**(9), 1159–1167 (Sep 2022). https://doi.org/10.1007/s11886-022-01733-1, https://doi.org/10.1007/s11886-022-01733-1
32. Zhu, T., Yao, X., Li, K., Herrero, P., Georgiou, P.: Blood glucose prediction for type 1 diabetes using generative adversarial networks. CEUR Workshop Proceedings **2675** (Jan 2020)