scientific data

OPEN

DATA DESCRIPTOR

DiaTrend: A dataset from advanced diabetes technology to enable development of novel analytic solutions

Temiloluwa Prioleau[1,2]✉, Abigail Bartolome[1], Richard Comi[3] & Catherine Stanger[2]

Objective digital data is scarce yet needed in many domains to enable research that can transform the standard of healthcare. While data from consumer-grade wearables and smartphones is more accessible, there is critical need for similar data from clinical-grade devices used by patients with a diagnosed condition. The prevalence of wearable medical devices in the diabetes domain sets the stage for unique research and development within this field and beyond. However, the scarcity of open-source datasets presents a major barrier to progress. To facilitate broader research on diabetes-relevant problems and accelerate development of robust computational solutions, we provide the DiaTrend dataset. The DiaTrend dataset is composed of intensive longitudinal data from wearable medical devices, including a total of 27,561 days of continuous glucose monitor data and 8,220 days of insulin pump data from 54 patients with diabetes. This dataset is useful for developing novel analytic solutions that can reduce the disease burden for people living with diabetes and increase knowledge on chronic condition management in outpatient settings.

Background & Summary

Advanced technologies like continuous glucose monitors (CGMs) and insulin pumps are transforming the standard of care for diabetes management[1–3]. The ubiquitous nature of these devices enables real-time monitoring and treatment in daily living; this is a huge advantage over single point-in-time alternatives like glucose meters and insulin pens. Research shows that many patients with diabetes achieve better outcomes with CGMs and insulin pumps[4,5]. However, research also shows that digital data from these devices is significantly underutilized to optimize outcomes[6,7]. Meanwhile, the next generation of solutions needed to advance diabetes care, such as the hybrid and fully closed-loop artificial pancreas[8,9], depend substantially on continuous data from CGMs and insulin pumps. A major barrier to progress in this field centers around access to rich datasets that facilitate the development of novel analytic solutions. In addition, there is a large amount of related but disconnected data streams that is not often reviewed or analyzed together, which further limits our understanding of diabetes management and even prevention[10,11]. To advance research and development of robust analytic solutions for the growing population of people with diabetes, there is a critical need for open datasets to understand outpatient management, develop interventions, and build clinically-relevant decision-support solutions.

Despite the recognized need for open datasets to enable research[12], there are limited datasets for data-driven research in the diabetes domain. One is the OhioT1DM dataset[13], which consists of eight weeks of CGM, insulin pump, physiological sensor, and self-reported events from 12 people with type 1 diabetes, while another is an N-of-1 dataset, which consists of two weeks of blood glucose, insulin, and carbohydrate intake logs[14]. To broaden the scope of research on diabetes and chronic conditions in general, and accelerate development of robust computational solutions, we provide the DiaTrend dataset. The DiaTrend dataset includes CGM and insulin pump data from 54 patients with type 1 diabetes. This dataset is created from a subset of two larger studies focused on: (1) developing computational tools for self-management of diabetes[6], and (2) evaluating a digital intervention for young adults with type 1 diabetes[15]. The provided dataset includes time-aligned blood

[1]Dartmouth College, Department of Computer Science, Hanover, 03755, USA. [2]Dartmouth College, Center for Technology and Behavioral Health, Lebanon, 03766, USA. [3]Dartmouth Health, Geisel School of Medicine, Lebanon, 03766, USA. ✉e-mail: tprioleau@dartmouth.edu

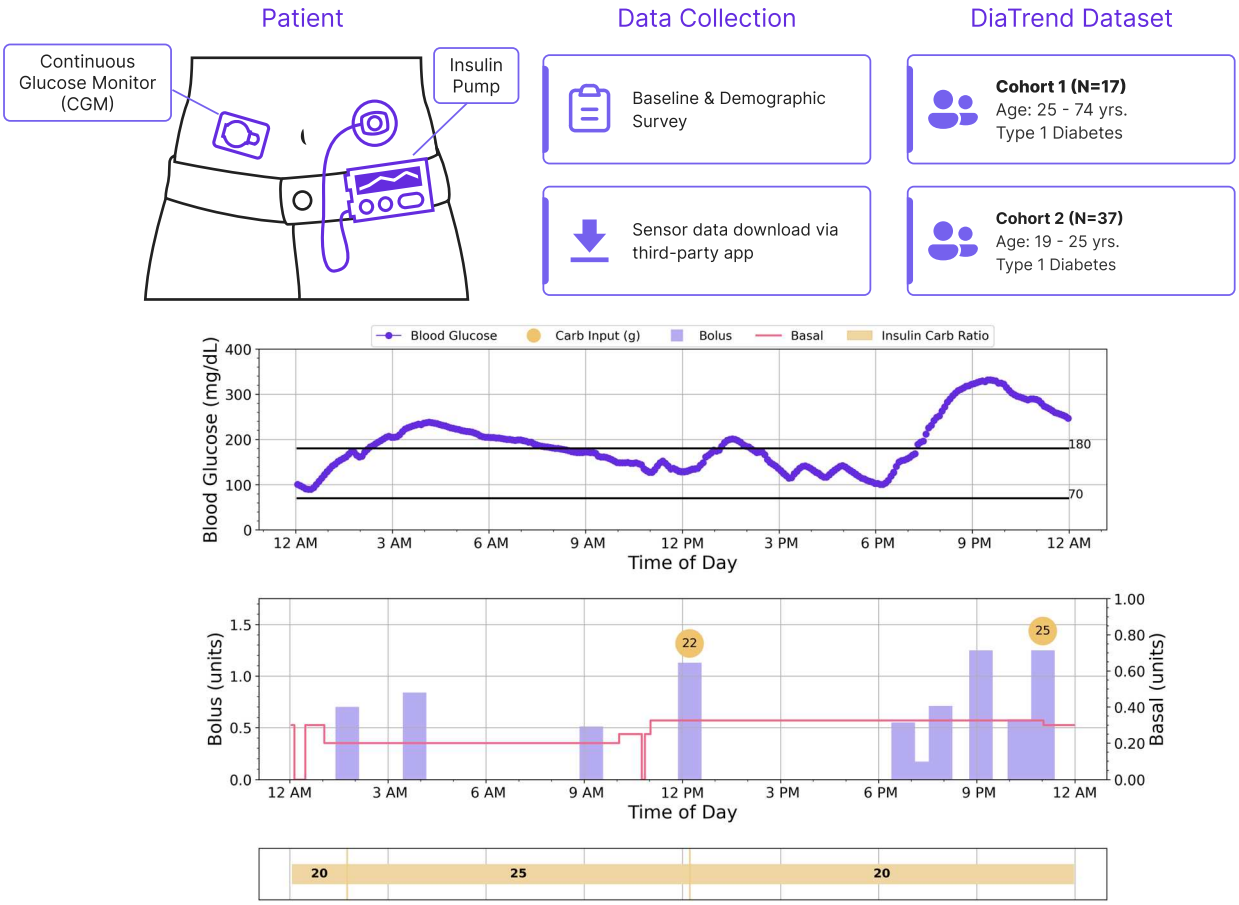

Fig. 1 Overview of the data collection process and data provided in the DiaTrend dataset.

glucose samples recorded on average every 5 minutes with FDA-approved CGMs by Dexcom[16], Abbott[17], and Medtronic[18], and insulin pump data comprising basal and bolus insulin doses, carbohydrate intake logs, and other pump settings such as insulin-carb ratio and more. Figure 1 presents an overview of the data collection process and data provided.

The DiaTrend dataset is useful for several research directions including more common tasks like blood glucose prediction[19–26], prediction of adverse glycemic events (i.e., hypoglycemia and hyperglycemia)[27–30], detection of unannounced meals[31–35], and algorithm development for insulin delivery[36,37]. However, this dataset is also useful to support further research on less studied topics like discovering digital biomarkers of glycemic control[7], mining patterns/trends in diabetes management[6,38,39], understanding adherence to wearable medical devices and patterns of missing data[40,41], developing novel visual analytic and data visualization solutions[42], and designing decision-support tools through user-centered studies[43–46]. Additionally, prospective researchers can find more opportunities for artificial intelligence in the diabetes domain through recent reviews in literature[47–49].

Methods

Participants. The DiaTrend dataset includes CGM and insulin pump data from a total of 54 patients with type 1 diabetes (age: 19–74 years, gender: 17 males, 37 females). Table 1 provides an overview of the demographic and clinical characteristics of patients in this dataset, including the distribution across age groups, gender, race, diabetes type, and hemoglobin A1C. Participants were recruited through two independent studies. Study 1 (also known as Digital SMD) recruited patients from Dartmouth Health in 2019, while study 2 (also known as SweetGoals[15]) is an ongoing randomized control trial that recruits patients through social media and online platforms. Both studies were approved by the Committee for Protection of Human Subjects at Dartmouth College (STUDY00031632 and STUDY00023559, respectively) and all participants provided verbal and written consent prior to joining either study. In addition, participants provided consent to share their data openly to the broader research community. To protect the privacy of study participants and minimize the risk of patient re-identification, the DiaTrend dataset is provided via a controlled access mechanism[50], similar to related datasets in the field[13].

Cohort 1 (n = 17), from the Digital SMD study[6], includes persons with type 1 diabetes between the ages of 25 to 74 years old who use a CGM and insulin pump for daily management of their condition and consented to share their retrospective device data for research. Meanwhile, cohort 2 (n = 37), from the SweetGoals study[15], includes persons with type 1 diabetes for longer than 18 months between the ages of 19 to 29 years old who use a Glooko compatible glucometer or CGM, reported a clinical visit within the previous 6 months from the

Characteristics	Count (n = 54)	%Dist
Age		
19–24 yrs	24	44.4%
25–34 yrs	14	25.9%
35–44 yrs	4	7.4%
45–54 yrs	3	5.6%
55–74 yrs	9	16.7%
Gender		
Female	37	68.5%
Male	17	31.5%
Race		
White/Caucasian	48	88.9%
Asian or Pacific Islander	2	3.7%
Black/African American	1	1.9%
Black/African American & White	1	1.9%
Other	1	1.9%
Not Reported	1	1.9%
Diagnosis		
Type 1 Diabetes	54	100%
Hemoglobin A1C		
6.0–6.9	8	14.8%
7.0–7.9	23	42.6%
8.0–8.9	17	31.5%
9.0–11.0	3	5.6%
Not Reported	3	5.6%

Table 1. DiaTrend dataset: Demographic and clinical characteristics of patients with diabetes.

recruitment date, and self-reported their most recent Hemoglobin A1C (HA1C) value as >7.5%. It is important to note that all device data included in the DiaTrend dataset was collected at baseline (i.e., prior to any intervention). Additionally, each individual's dataset spans varying time periods based on the available retrospective data at the time of recruitment. Given our focus on advanced diabetes technology for novel analytic solutions, only participants who use CGMs (with <30% missing data) and insulin pumps for daily management are included in the provided dataset.

Dataset description. The DiaTrend dataset includes a total of 27,561 days of CGM data and 8,220 days of insulin pump data from 54 patients with type 1 diabetes. In addition, the DiaTrend dataset includes demographic and clinical characteristics for each subject, including metrics such as age, gender, race, diabetes type and HA1C - see Table 1. There is an average of 510 days (range: 31–1885 days) of CGM data per subject, and an average of 152 days (range: 31–780 days) of insulin pump data per subject - see Fig. 2. Within the insulin pump data, there is an average of 993 total bolus doses per subject (range: 132–4939 doses) and an average of 438 total carb inputs per subject (range: 1–2310 input) - see Fig. 3. These data were collected as part of the Digital SMD[6] and SweetGoals[15] studies during which each patient's retrospective CGM and insulin pump data was downloaded through a third-party application (i.e., Tidepool[51] or Glooko[52]). It is important to note that since the SweetGoals study is a randomized control trial, only retrospective baseline data collected during the initial screening is included as part of the DiaTrend dataset (i.e., the provided data does not include sensor data from the intervention period of that study). In addition, HA1C - the primary clinically-validated metric for accessing glycemic control - was collected via the patient's electronic health record (i.e., the most recent HA1C) in the Digital SMD study and via a mail-in home test in the SweetGoals study at the time of the baseline assessment (approximately the endpoint of the device data).

Data Records

All data records in the DiaTrend dataset are stored and accessible via the Synapse platform[50]. The deposited data consists of 54 Excel files–one file for each subject. Each file has a CGM sheet that provides blood glucose data that was collected by the CGM. The CGM sheet includes 2 columns, namely, date and mg/dL. In addition, each subjects' file also has a Bolus sheet, which describes bolus insulin doses and meal announcements (i.e., user-entered estimates of carbohydrate content in meals logged to calculate bolus insulin needed to metabolize glucose from the meal consumed[53]). The Bolus sheet includes the following 7 columns: date, normal, carbInput, insulinCarbRatio, bgInput, recommended.carb, and recommended.net. It is important to note that only 17 subject files that have a Basal sheet, which describes the subject's basal infusions in 3 columns, namely, date, duration, and rate. The subject files that have basal data are as follows: S29-S31, S36-S39, S42, S45-S47, S49-S54. In addition, 37 (out of 54) Bolus sheets include 4 more columns, namely, recommended.correction, insulinSensitivityFactor, targetBloodGlucose, and insulinOnBoard. The subject files that have the 4 additional columns in

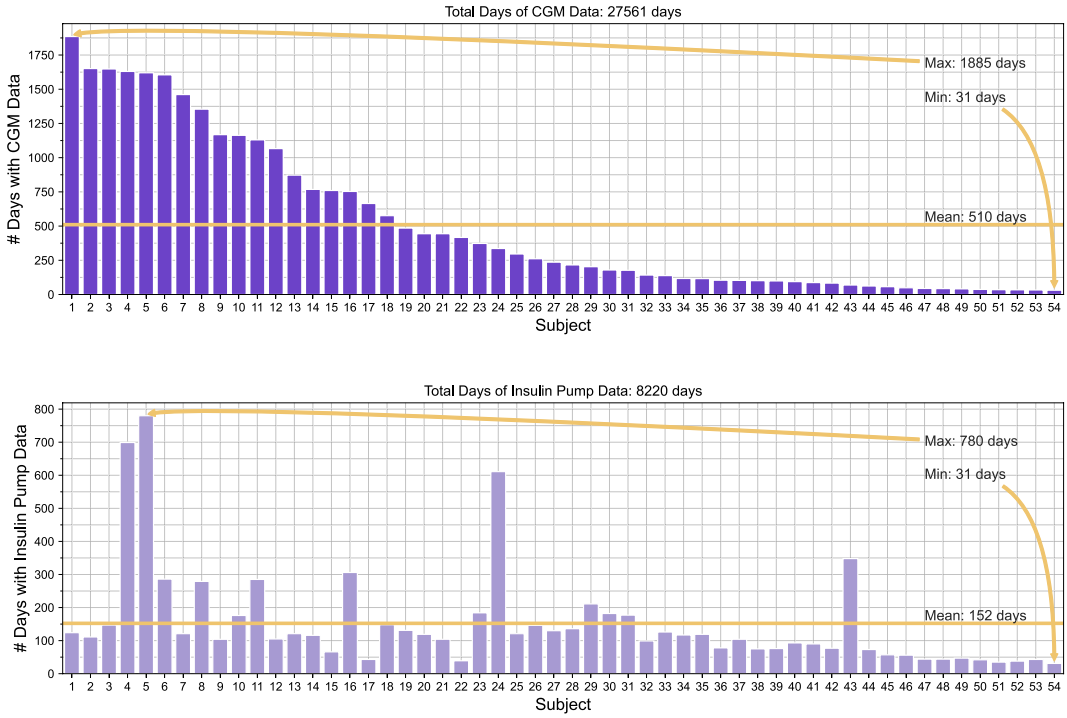

Fig. 2 Overview of the number of days of sensor data per patient in the DiaTrend dataset.

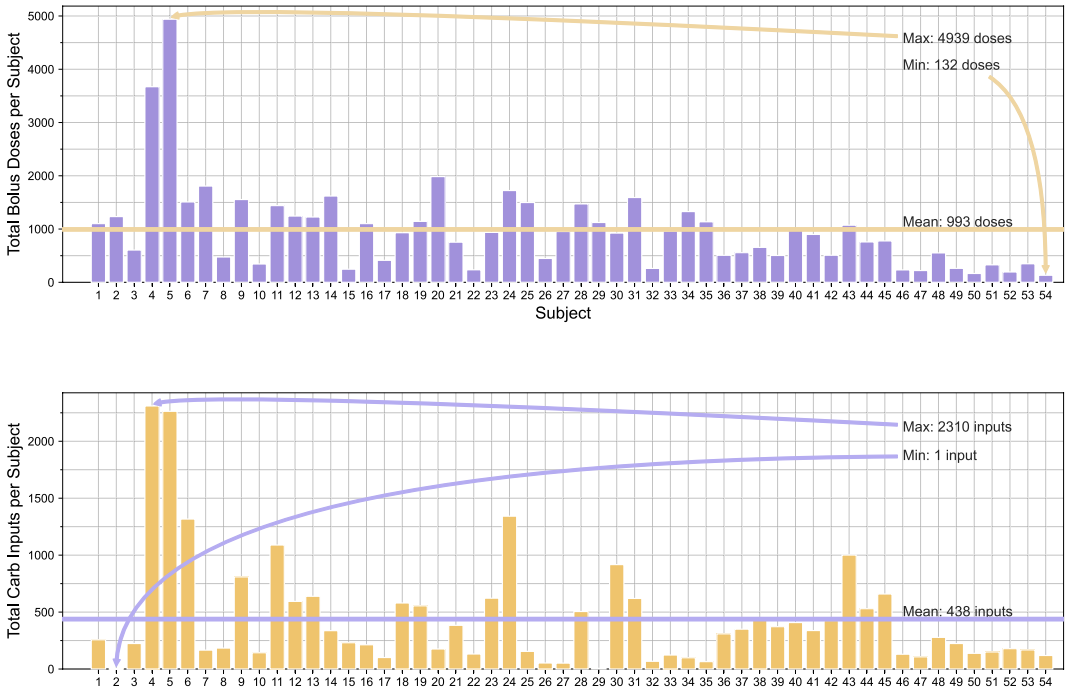

Fig. 3 Overview of the total number of bolus and carb input data per patient in the DiaTrend dataset.

the bolus sheets are as follows: S1-S28, S32-S35, S40, S41, S43, S44, and S48. Each row in all three of the Excel sheets refers to one record collected at a given timestamp in the column titled 'date'. All data records in each subject file are time-ordered according to the device log. More specifically, CGMs record a blood glucose sample approximately every 5 minutes, meanwhile insulin pumps have irregularly sampled data records because they depend on user triggers for bolus insulin doses and user settings for basal insulin doses. Excluding the date column, the rest of the data can be read as floating point numbers. Table 2 provides a detailed breakdown of each data record, the format, and a description.

Sheet Name	Column Name	Format	Description
CGM	date	datetime (yyyy-mm-dd HH:MM:SS)	Date and time that glucose reading was recorded
	mg/dL	Float64	Blood glucose reading in mg/dL
Bolus	date	datetime (yyyy-mm-dd HH:MM:SS)	Date and time that bolus was administered
	normal	Float64	Amount of bolus insulin delivered (units)
	carbInput	Float64	Total carbs announced for bolus (grams)
	insulinCarbRatio	Float64	Patient setting for grams of carbs covered per one unit of insulin
	bgInput	Float64	Blood glucose reading at time of bolus (mg/dL)
	recommended.carb	Float64	Amount of insulin recommended to cover carb intake for normal bolus
	recommended.net	Float64	Amount of insulin recommended for bolus delivery
	recommended.correction	Float64	Amount of insulin recommended for correction component of normal bolus
	insulinSensitivityFactor	Float64	Patient setting for how one unit of insulin lowers blood glucose level
	targetBloodGlucose	Float64	Target blood glucose value for after bolus delivery
	insulinOnBoard	Float64	Amount of active insulin remaining from prior insulin doses
Basal	date	datetime (yyyy-mm-dd HH:MM:SS)	Date and time of basal infusion
	duration	Float64	Duration of basal infusion (ms)
	rate	Float64	Rate of basal infusion (units/hr)

Table 2. Overview of the data records, format, and description in the DiaTrend dataset.

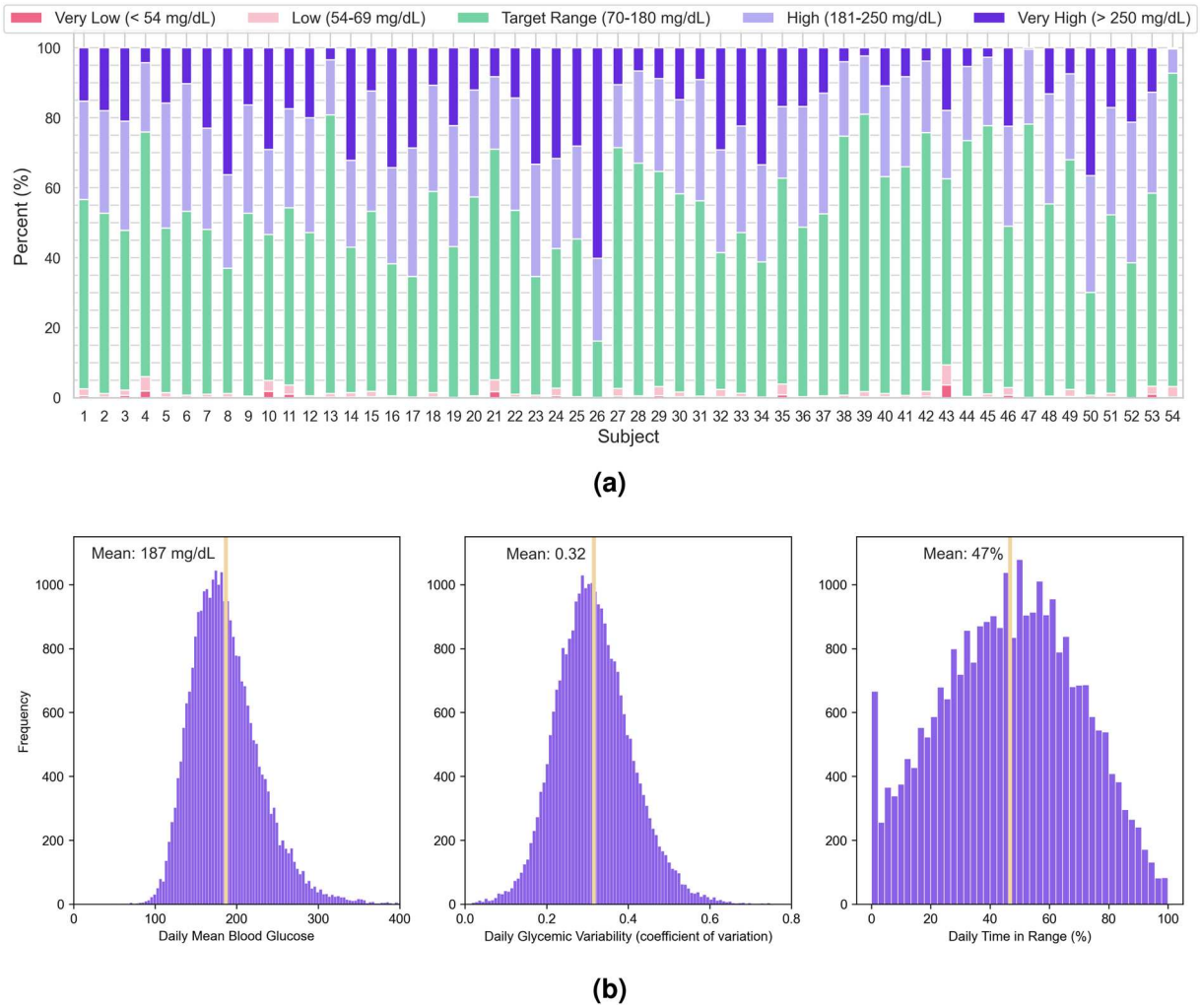

Fig. 4 Descriptive summary of CGM data in the DiaTrend dataset. (**a**) The percent of blood glucose samples in 5 clinically-relevant categories. (**b**) The distributions of daily mean blood glucose, daily glycemic variability, and daily time in [target] range.

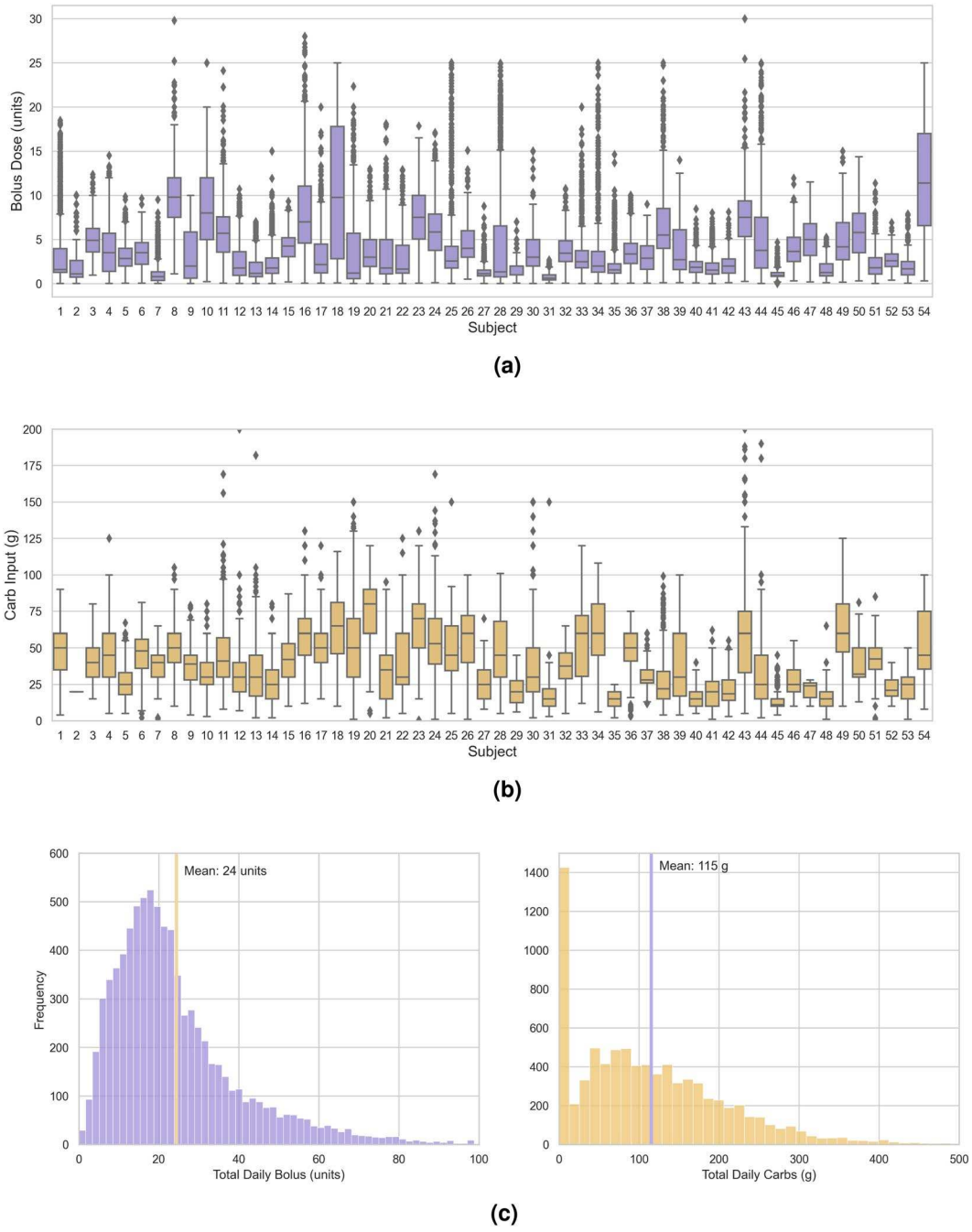

Fig. 5 Descriptive summary of insulin pump data in the DiaTrend dataset. (**a**) A box plot of all bolus insulin doses per subject. (**b**) A box plot of all carb input entries per subject. (**c**) The distributions of total daily bolus insulin and total daily carb inputs across all subjects.

Technical Validation

For each patient included in the DiaTrend dataset, we provide an overview of their blood glucose data using clinically-validated metrics for assessing glycemic control[54,55]. This includes the percentage of all blood glucose readings in 5 clinically-relevant categories, namely, very low (<54 mg/dL), low (54–69 mg/dL), target range (70–180 mg/dL), high (181–250 mg/dL), and very high (>250 mg/dL). From Fig. 4, we can observe that blood glucose is highly variable and only a minority of patients living with diabetes (less than 10% in our dataset) meet the clinical target of maintaining blood glucose within the target range of 70–180 mg/dL for more than 70% of the time[54]. Fig. 4b presents histograms for daily mean blood glucose (mean = 187 mg/dL), daily glycemic variability (mean = 0.33), and daily time in range (mean = 47%). From this figure, we can observe a normal distribution for each clinically-relevant metric in the DiaTrend dataset.

Similarly, we provide an overview of each patient's insulin pump data using box plots and histograms. Figure 5a,b show box plots with descriptive statistics associated with bolus insulin doses and carb inputs,

respectively, for each subject. Additionally, Fig. 5c shows the distributions of total daily bolus insulin doses (units) and total daily carb inputs (g), respectively. From this figure, we can observe a mean total daily bolus of 24 units and a mean total daily carb input of 115 g, both with a positively skewed distribution. In particular, we observe a high number of days (~1400 days) with very low carb inputs (~0 g); this could be indicative of missed mealtime boluses (i.e., no bolus insulin used during mealtimes)–this is a common contributor to poor glycemic outcomes[56–58].

Limitations. There are some important considerations and limitations associated with the DiaTrend dataset provided in this paper. First, there is imbalance in the representation of subjects across the dimensions of race, gender, and age. More specifically, majority of patients whose CGM and insulin pump data is provided (i.e., 48 out of 54 or 89%) are non-Hispanic White/Caucasian. Also, this dataset includes a lower representation of males (n = 17 out of 54 or 32%) compared to females, and a lower representation of older adults (e.g., for age ≥45 years old, n = 12 or 22%). The limitation with regards to race (i.e., low representation of participants from non-White/Caucasian races, including Hispanics, non-Hispanic Black/African Americans, and Asians) is partly due to the geographical location (i.e., New Hampshire) from which some participants (17 out of 54) were recruited. However, the imbalance in representation also underscores racial disparities that have been identified in prior literature relating to access and use of advanced diabetes technologies, particularly CGMs and insulin pumps[59,60]. Additionally, the limitation with regards to age (i.e., low representation of older adults and higher representation of young adults) is primarily due to the targeted focus on young adults with type 1 diabetes in the SweetGoals study[15]. A second limitation of the DiaTrend dataset is that it lacks full temporal alignment in the CGM and insulin pump data for each participant. This difference is apparent from Fig. 2 which shows more CGM data than insulin pump data for a number of subjects. While the reason for this is unknown, we suspect that it is primarily due lower data storage capacity on insulin pumps compared to CGMs, which in turn limits the amount of retrospective data available for download from insulin pumps or patients' switching insulin delivery systems (e.g., to multiple daily injections or other devices that are not compatible with the third-party platform). Third, there are various forms of missing data associated with the provided dataset. As previously mentioned, all data provided in this paper represents retrospective data collected directly from the user's devices (i.e., CGMs and insulin pumps) and downloaded through a third-party application (i.e., Tidepool[51] or Glooko[52]). Given this, missing data in the data files are due to either missing data in the user's device or technical issues with the third-party platform used for download. For example, basal insulin data is not available for subjects from cohort 2 (37 out of 54) due to technical issues with Glooko not providing basal data from the insulin pumps at the time of data collection for this study. These forms of missing data might limit some research efforts with the provided dataset. However, despite the aforementioned limitations, the DiaTrend dataset represents one of the largest open-source datasets currently available in the diabetes domain. This critical resource provides a unique opportunity to advance development of novel data-driven solutions that can improve the lives of people living with diabetes. In addition, this dataset provides a necessary benchmark to evaluate the generalizability of numerous diabetes-relevant algorithms in literature[19–36].

Usage Notes

The DiaTrend dataset is provided for research and educational purposes that support the development of novel data-driven solutions for the diabetes community and beyond. Consistent with exemplar studies[13,61,62], we have set governance structures in place to balance the need for open datasets that advance research and protect the privacy of participants.

Researchers interested in accessing the DiaTrend dataset should complete the following steps:

1. Register for a Synapse account (www.synapse.org).
2. Become a Synapse Certified User with a validated user profile.
3. Submit an Intended Data Use statement.
4. Agree to the Conditions of Use.

Code availability

Python was used for all data processing described in this paper. The Python code used to generate all figures in this paper is available on the Augmented Health Lab's Github: https://github.com/Augmented-Health-Lab/Diatrend.

Received: 20 March 2023; Accepted: 14 August 2023;
Published online: 23 August 2023

References

1. Beck, R. W., Bergenstal, R. M., Laffel, L. M. & Pickup, J. C. Advances in technology for management of type 1 diabetes. *The Lancet* **394**, 1265–1273 (2019).
2. American Diabetes Association and others. 7. Diabetes technology: standards of medical care in diabetes-2021. *Diabetes Care* **44**, S85–S99 (2021).
3. Cappon, G., Vettoretti, M., Sparacino, G. & Facchinetti, A. Continuous glucose monitoring sensors for diabetes management: a review of technologies and applications. *Diabetes & Metabolism Journal* **43**, 383–397 (2019).
4. Rodbard, D. Continuous glucose monitoring: a review of recent studies demonstrating improved glycemic outcomes. *Diabetes Technology & Therapeutics* **19**, S–25 (2017).
5. Taylor, P. J., Thompson, C. H. & Brinkworth, G. D. Effectiveness and acceptability of continuous glucose monitoring for type 2 diabetes management: a narrative review. *Journal of Diabetes Investigation* **9**, 713–725 (2018).
6. Bartolome, A., Shah, S. & Prioleau, T. Glucomine: A case for improving the use of wearable device data in diabetes management. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **5**, 1–24 (2021).

7. Bartolome, A. & Prioleau, T. A computational framework for discovering digital biomarkers of glycemic control. *npj Digital Medicine* **5**, 1–9 (2022).
8. Thabit, H. & Hovorka, R. Coming of age: the artificial pancreas for type 1 diabetes. *Diabetologia* **59**, 1795–1805 (2016).
9. Doyle, F. J. III, Huyett, L. M., Lee, J. B., Zisser, H. C. & Dassau, E. Closed-loop artificial pancreas systems: Engineering the algorithms. *Diabetes Care* **37**, 1191–1197 (2014).
10. Walsh, J., Roberts, R., Morris, R. & Heinemann, L. Device connectivity: The next big wave in diabetes. *Journal of Diabetes Science and Technology* **9**, 701–705 (2015).
11. Iyengar, V., Wolf, A., Brown, A. & Close, K. Challenges in diabetes care: Can digital health help address them? *Clinical Diabetes* **34**, 133–141 (2016).
12. Bietz, M. J. *et al*. Opportunities and challenges in the use of personal health data for health research. *Journal of the American Medical Informatics Association* **23**, e42–e48 (2016).
13. Marling, C. & Bunescu, R. The OhioT1DM dataset for blood glucose level prediction: update 2020. In *CEUR workshop proceedings*, vol. 2675, 71 (NIH Public Access, 2020).
14. Katz, D. & Price, B. *Two week diabetes data set*. https://doi.org/10.21954/ou.rd.5756379.v1 (2018).
15. Stanger, C. *et al*. A digital health intervention (sweetgoals) for young adults with type 1 diabetes: protocol for a factorial randomized trial. *JMIR Research Protocols* **10**, e27109 (2021).
16. Dexcom. Dexcom continuous glucose monitoring. https://www.dexcom.com/ (2023).
17. Abbott. Freestyle libre continuous glucose monitor. https://www.abbott.com/freestyle-libre-2-continuous-glucose-monitor-cgm.html (2023).
18. Medtronic. The guardian connect system. https://www.medtronicdiabetes.com/products/guardian-connect-continuous-glucose-monitoring-system (2023).
19. Gu, K., Dang, R. & Prioleau, T. Neural physiological model: A simple module for blood glucose prediction. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 5476–5481 (IEEE, 2020).
20. Li, K., Daniels, J., Liu, C., Herrero, P. & Georgiou, P. Convolutional recurrent neural networks for glucose prediction. *IEEE Journal of Biomedical and Health Informatics* **24**, 603–613 (2019).
21. Deng, Y. *et al*. Deep transfer learning and data augmentation improve glucose levels prediction in type 2 diabetes patients. *npj Digital Medicine* **4**, 1–13 (2021).
22. Zhu, T., Li, K., Herrero, P., Chen, J. & Georgiou, P. A deep learning algorithm for personalized blood glucose prediction. In *KHD@ IJCAI*, 64–78 (2018).
23. Li, K., Liu, C., Zhu, T., Herrero, P. & Georgiou, P. Glunet: A deep learning framework for accurate glucose forecasting. *IEEE Journal of Biomedical and Health Informatics* **24**, 414–423 (2019).
24. Martinsson, J. *et al*. Automatic blood glucose prediction with confidence using recurrent neural networks. In *KHD@ IJCAI* (2018).
25. Woldaregay, A. Z. *et al*. Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes. *Artificial Intelligence in Medicine* **98**, 109–134 (2019).
26. Zaidi, S. M. A. *et al*. Multi-step ahead predictive model for blood glucose concentrations of type-1 diabetic patients. *Scientific Reports* **11**, 24332 (2021).
27. Gadaleta, M., Facchinetti, A., Grisan, E. & Rossi, M. Prediction of adverse glycemic events from continuous glucose monitoring signal. *IEEE Journal of Biomedical and Health Informatics* **23**, 650–659 (2018).
28. Mosquera-Lopez, C., Dodier, R., Tyler, N., Resalat, N. & Jacobs, P. Leveraging a big dataset to develop a recurrent neural network to predict adverse glycemic events in type 1 diabetes. *IEEE Journal of Biomedical and Health Informatics* (2019).
29. Seo, W., Lee, Y.-B., Lee, S., Jin, S.-M. & Park, S.-M. A machine-learning approach to predict postprandial hypoglycemia. *BMC Medical Informatics and Decision Making* **19**, 1–13 (2019).
30. Dave, D. *et al*. Feature-based machine learning model for real-time hypoglycemia prediction. *Journal of Diabetes Science and Technology* **15**, 842–855 (2021).
31. Zheng, M., Ni, B. & Kleinberg, S. Automated meal detection from continuous glucose monitor data through simulation and explanation. *Journal of the American Medical Informatics Association* **26**, 1592–1599 (2019).
32. Ramkissoon, C. M., Herrero, P., Bondia, J. & Vehi, J. Unannounced meals in the artificial pancreas: detection using continuous glucose monitoring. *Sensors* **18**, 884 (2018).
33. Xie, J. & Wang, Q. A variable state dimension approach to meal detection and meal size estimation: in silico evaluation through basal-bolus insulin therapy for type 1 diabetes. *IEEE Transactions on Biomedical Engineering* **64**, 1249–1260 (2016).
34. Samadi, S. *et al*. Automatic detection and estimation of unannounced meals for multivariable artificial pancreas system. *Diabetes Technology & Therapeutics* **20**, 235–246 (2018).
35. Kölle, K., Biester, T., Christiansen, S., Fougner, A. L. & Stavdahl, Ø. Pattern recognition reveals characteristic postprandial glucose changes: Non-individualized meal detection in diabetes mellitus type 1. *IEEE Journal of Biomedical and Health Informatics* **24**, 594–602 (2019).
36. Vettoretti, M. & Facchinetti, A. Combining continuous glucose monitoring and insulin pumps to automatically tune the basal insulin infusion in diabetes therapy: a review. *Biomedical Engineering Online* **18**, 1–17 (2019).
37. Mosquera-Lopez, C. *et al*. Enabling fully automated insulin delivery through meal detection and size estimation using artificial intelligence. *npj Digital Medicine* **6**, 39 (2023).
38. Morton, S., Li, R., Dibbo, S. & Prioleau, T. Data-driven insights on behavioral factors that affect diabetes management. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 5557–5562 (IEEE, 2020).
39. Belsare, P., Lu, B., Bartolome, A. & Prioleau, T. Investigating temporal patterns of glycemic control around holidays. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (2022).
40. Vhaduri, S. & Prioleau, T. Adherence to personal health devices: A case study in diabetes management. In *Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 62–72 (2020).
41. Drecogna, M., Vettoretti, M., Del Favero, S., Facchinetti, A. & Sparacino, G. Data gap modeling in continuous glucose monitoring sensor data. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 4379–4382 (IEEE, 2021).
42. Zhang, Y., Chanana, K. & Dunne, C. Idmvis: Temporal event sequence visualization for type 1 diabetes treatment decision support. *IEEE Transactions on Visualization and Computer Graphics* **25**, 512–522 (2018).
43. Prioleau, T., Sabharwal, A. & Vasudevan, M. M. Understanding reflection needs for personal health data in diabetes. In *Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 263–273 (2020).
44. Katz, D. S., Price, B. A., Holland, S. & Dalton, N. S. Data, data everywhere, and still too hard to link: Insights from user interactions with diabetes apps. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12 (2018).
45. Raj, S., Lee, J. M., Garrity, A. & Newman, M. W. Clinical data in context: towards sensemaking tools for interpreting personal health data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **3**, 1–20 (2019).
46. Raj, S., Toporski, K., Garrity, A., Lee, J. M. & Newman, M. W. "My blood sugar is higher on the weekends" finding a role for context and context-awareness in the design of health self-management technology. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13 (2019).
47. Contreras, I. *et al*. Artificial intelligence for diabetes management and decision support: literature review. *Journal of Medical Internet Research* **20**, e10775 (2018).

48. Ellahham, S. Artificial intelligence: The future for diabetes care. *The American Journal of Medicine* **133**, 895–900 (2020).
49. Tyler, N. S. & Jacobs, P. G. Artificial intelligence in decision support systems for type 1 diabetes. *Sensors* **20**, 3214 (2020).
50. Prioleau, T., Bartolome, A., Comi, R. & Stanger, C. Diatrend: A dataset from advanced diabetes technology. *Synapse* https://doi.org/10.7303/syn38187184 (2022).
51. Tidepool. https://www.tidepool.org/ (2023).
52. Glooko. Glooko: Remote monitoring for diabetes and related conditions. https://glooko.com/ (2023).
53. Sora, N. D., Shashpal, F., Bond, E. A. & Jenkins, A. J. Insulin pumps: Review of technological advancement in diabetes management. *The American Journal of the Medical Sciences* **358**, 326–331 (2019).
54. Battelino, T. *et al*. Clinical targets for continuous glucose monitoring data interpretation: recommendations from the international consensus on time in range. *Diabetes Care* **42**, 1593–1603 (2019).
55. Danne, T. *et al*. International consensus on use of continuous glucose monitoring. *Diabetes Care* **40**, 1631–1640 (2017).
56. Burdick, J. *et al*. Missed insulin meal boluses and elevated hemoglobin a1c levels in children receiving insulin pump therapy. *Pediatrics* **113**, e221–e224 (2004).
57. Deeb, A. *et al*. Important determinants of diabetes control in insulin pump therapy in patients with type 1 diabetes mellitus. *Diabetes Technology & Therapeutics* **17**, 166–170 (2015).
58. Patton, S. R. *et al*. Frequency of mealtime insulin bolus predicts glycated hemoglobin in youths with type 1 diabetes. *Diabetes Technology & Therapeutics* **16**, 519–523 (2014).
59. Akturk, H. K., Agarwal, S., Hoffecker, L. & Shah, V. N. Inequity in racial-ethnic representation in randomized controlled trials of diabetes technologies in type 1 diabetes: critical need for new standards. *Diabetes Care* **44**, e121–e123 (2021).
60. Majidi, S. *et al*. Inequities in health outcomes in children and adults with type 1 diabetes: data from the t1d exchange quality improvement collaborative. *Clinical Diabetes* **39**, 278–283 (2021).
61. Bot, B. M. *et al*. The mpower study, parkinson disease mobile data collected using researchkit. *Scientific Data* **3**, 1–9 (2016).
62. Hershman, S. G. *et al*. Physical activity, sleep and cardiovascular health data for 50,000 individuals from the myheart counts study. *Scientific Data* **6**, 1–10 (2019).

Acknowledgements

C.S. acknowledges funding from the National Institute of Diabetes and Digestive and Kidney Diseases (R01DK124428).

Author contributions

T.P. conceived the Digital SMD study for cohort 1. T.P., A.B. and R.C. led efforts for data collection from cohort 1. C.S. conceived and led the SweetGoals study and data collection for cohort 2. A.B. cleaned and organized both datasets. T.P. and A.B. wrote the manuscript, and prepared figures & tables. All authors reviewed and contributed to the final approval of the manuscript.

Competing interests

The authors declare no completing interests.

Additional information

Correspondence and requests for materials should be addressed to T.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.