

ARTICLE OPEN



A computational framework for discovering digital biomarkers of glycemic control

Abigail Bartolome¹ and Temiloluwa Prioleau¹

Digital biomarkers can radically transform the standard of care for chronic conditions that are complex to manage. In this work, we propose a scalable computational framework for discovering digital biomarkers of glycemic control. As a feasibility study, we leveraged over 79,000 days of digital data to define objective features, model the impact of each feature, classify glycemic control, and identify the most impactful digital biomarkers. Our research shows that glycemic control varies by age group, and was worse in the youngest population of subjects between the ages of 2–14. In addition, digital biomarkers like prior-day time above range and prior-day time in range, as well as total daily bolus and total daily basal were most predictive of impending glycemic control. With a combination of the top-ranked digital biomarkers, we achieved an average F1 score of 82.4% and 89.7% for classifying next-day glycemic control across two unique datasets.

npj Digital Medicine (2022)5:111 | <https://doi.org/10.1038/s41746-022-00656-z>

INTRODUCTION

Maintaining good glycemic control is the primary and daily objective for persons with diabetes. Tangibly, this equates to managing several factors that can affect blood glucose such as food, medication, activity, and much more^{1–4}. Research shows that persons with diabetes make over 100 health-related decisions and spend an average of 58 mins per day on self-care related to their condition^{5,6}. This daily effort is critical to minimize the occurrence of adverse glycemic events (i.e., hypoglycemia and hyperglycemia). In addition, maintaining good glycemic control is imperative to delay and possibly avoid long-term diabetes complications like kidney failure and peripheral arterial disease^{7,8}. Hence, the use of digital tools like continuous glucose monitors (CGMs) and insulin pumps is growing to support the complex task of managing diabetes^{9–12}.

CGMs and insulin pumps are clinically-validated minimally-invasive wearable systems that enable continuous monitoring of blood glucose concentrations, and semi-automatic administration of insulin for the goal of maintaining normoglycemia (i.e., normal blood glucose levels). The benefits of using these wearable medical devices in diabetes care is two-fold. First, CGMs and insulin pumps enable real-time awareness of blood glucose levels and in-the-moment response to current or impending adverse glycemic events^{13,14}. Second, CGMs and insulin pumps provide a rich data source for retrospective learning to guide future management strategy. However, research shows that these data sources are significantly underutilized^{15,16}. Given the growing national and global burden of diabetes^{17–20}, there are unique and untapped opportunities for intelligent computational methods that can leverage the vast amount of digital data from CGMs, insulin pumps, and other devices to increase our understanding of the association and impact of individualized daily factors on diabetes outcomes.

The importance of digital biomarkers in the health domain is irrefutable^{21–28}. Digital biomarkers refer to objective, quantifiable, physiological, and behavioral measures that are collected by means of digital devices, such as wearable devices, for the purpose of explaining, influencing, or predicting health outcomes^{21,29,30}.

However, there is limited research on digital biomarkers relevant to chronic conditions like diabetes. Despite the deluge of digital data that is collected from routine use of wearable medical devices like CGMs and insulin pumps, there are many unanswered questions about quantifiable measures that are associated with and/or predictive of daily glycemic control.

To support an increased use of routinely collected data from a myriad of digital devices and improve our understanding of complex factors that affect diabetes, this research makes important contributions by proposing and evaluating a scalable computational framework for discovering digital biomarkers of glycemic control. As a feasibility study, we leverage over 79,000 days of digital data from CGMs and insulin pump to: 1) define objective features from digital data, 2) model digital biomarkers and glycemic control, 3) classify glycemic control with digital biomarkers, and 4) identify the most impactful digital biomarkers for impending glycemic control (see Fig. 1). Our dataset is orders of magnitude larger than related research in the field^{2,27,31–34}. In addition, our proposed framework can be easily expanded to integrate and incorporate continuous streams of other relevant data sources, such as from wearable activity trackers like Fitbit, to discover the impact of daily activity level, sleep, stress, circadian rhythm, and much more, on glycemic control. Our proposed method is intuitive and it provides interpretable and actionable insights that can further guide daily management strategies to improve diabetes outcomes.

RESULTS

Defining objective features from digital data

In this work, our digital data comes from CGMs and insulin pumps used by 250 subjects with diabetes, primarily type 1 diabetes (ages 2–76 years, diabetes duration is between 0 and 60 years); see Table 2 in the Methods section for more details. This dataset represents two unique populations including 100 subjects that use sensor augmented pump (SAP) therapy and 150 subjects that use a hybrid closed loop (HCL) system for daily management of their diabetes. In SAP therapy, the patients' insulin pump is in

¹Dartmouth College, Computer Science, Hanover, NH 03755, USA. ✉email: tprioleau@dartmouth.edu

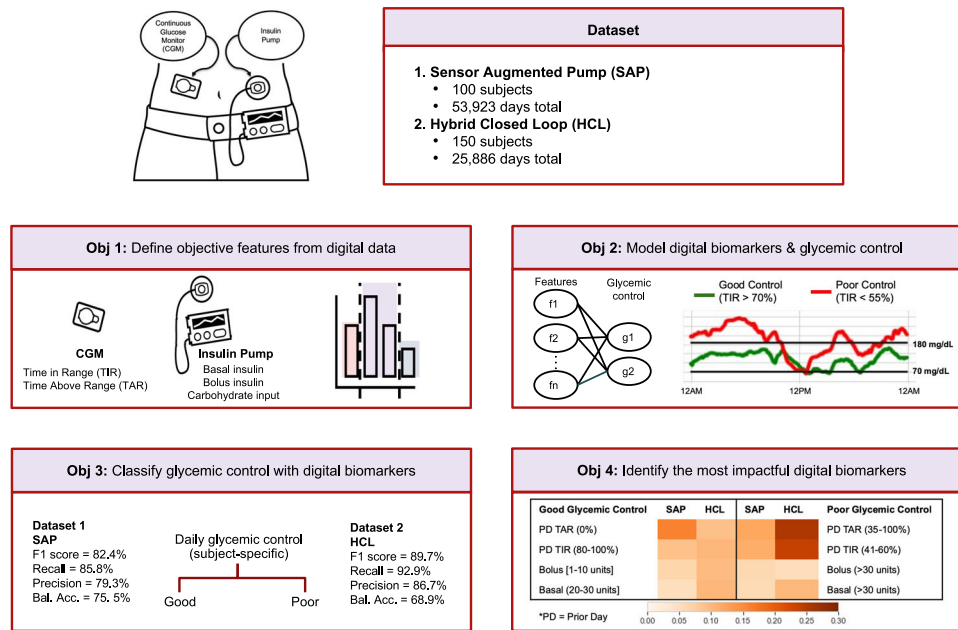


Fig. 1 Research overview. This study leverages over 79,000 days of CGM and insulin pump data from 250 subjects for four objectives: 1) define objective features from digital data, 2) model digital biomarkers and glycemic control, 3) classify glycemic control with digital biomarkers, 4) identify the most impactful digital biomarkers associated with good and poor glycemic control and consistent across two distinct datasets, i.e., sensor augmented pump (SAP) and hybrid closed loop (HCL).

communication with their CGM, thus the pump can suspend insulin delivery when blood glucose readings from the CGM are low (or getting low) to reduce hypoglycemia³⁵. Meanwhile, in the HCL case, the patients' CGM plus insulin pump system includes a more sophisticated algorithm that automatically adjusts primarily basal insulin (i.e., small amounts of background insulin) from the pump in response to blood glucose readings from the CGM^{36,37}. However, in both cases (SAP and HCL), a person with diabetes is expected to "announce meals" by entering the estimated carbohydrate amount into their insulin pump to determine and administer bolus insulin (i.e., larger amounts of fast-acting insulin) needed to account for food consumption^{38,39}. Thus, from both SAP and HCL populations, the diabetes-device dataset includes blood glucose data, recorded every ≈ 5 min from CGMs, as well as amounts and timestamps of bolus insulin doses, basal insulin doses, and carbohydrate amount entries from the insulin pumps.

For the objective of leveraging digital data to discover biomarkers of glycemic control (i.e., the outcome variable), we defined features using domain knowledge and data-driven insights^{2,14,40-43}. We define good glycemic control based on the clinical target for persons with type 1 diabetes, which is to maintain blood glucose between 70 and 180 mg/dL for greater than 70% of a 24-h day⁴⁰. Using the aforementioned target as an anchor, we define poor glycemic control as days when the time in range (TIR) is less than 55% in a 24-h day. Our definition is based on the intuition that if TIR greater than 70% is recognized clinically as good glycemic control, then TIR equal to 68% or 65% should not arbitrarily be classified as poor glycemic control. Thus, we consider the TIR between 55% and 70% as moderate glycemic control (i.e., not good and not poor). Additionally, we defined descriptive features for the input data (i.e., retrospective CGM data and insulin pump data). Based on the distribution of each data stream shown in Fig. 2, we identified patterns of behaviors that might impact impending glycemic control by stratifying each data stream into unique ranges. For example, total daily bolus insulin was described with one of four features: no entry, 1–10 units, 10–20 units, 20–30 units, and greater than 30 units. Large daily bolus insulin amounts (e.g., >30 units/day) could be representative of large meals, meanwhile low daily carbohydrate

inputs (e.g., 5–50g per day) could be representative of a low-carb diet^{43,44} or missed meal announcements⁴⁵⁻⁴⁷. In total, our input variable feature set comprising 26 features from five categories included two from the prior-day CGM data and three from the current-day insulin pump data. Supplementary Table 1 shows the full set of objective features.

The analysis in this step highlighted some key observations. From Fig. 2 and Table 1, we observe that the distribution of CGM data from the HCL population is more asymmetric and negatively skewed with a thinner tail on the left (mean TIR = 76.3%, median TIR = 80.4%) compared to the SAP population (mean TIR = 68.0%, median TIR = 72.2%). This indicates that there were more occurrence of high TIR (i.e., good glycemic control) in the HCL population compared to that of the SAP population. Prior work has found similar results and supports that patients with HCL systems often achieve better glycemic control and more time with their blood glucose levels in the target range of 70–180 mg/dL than patients with SAP therapy³⁹. Better glycemic control in the HCL population is likely due to the automatic and adjustable nature of basal insulin delivery, in contrast to the generally fixed nature of basal insulin delivery in the SAP population. In addition, from Fig. 2 and Table 1 respectively, we can also observe a smaller (or thinner) tail on the bolus insulin distribution with a slightly lower mean of 16.3 units per day in the HCL population compared to the SAP population with a mean of 20.4 units per day. This represents fewer occurrences of large daily bolus insulin in the HCL population compared to the SAP population. Conversely, we observe a wider tail on the basal insulin distribution and a mean of 25.5 units per day in the HCL population which represents more occurrences of larger daily basal insulin compared to the SAP population with a mean of 18.3 units per day. Based on the observed differences, we modeled the impact of the defined digital biomarkers on daily glycemic control to: 1) investigate the ability to classify good and poor glycemic control using a combination of those features, and 2) identify the most impactful digital biomarkers across both the SAP and HCL datasets.

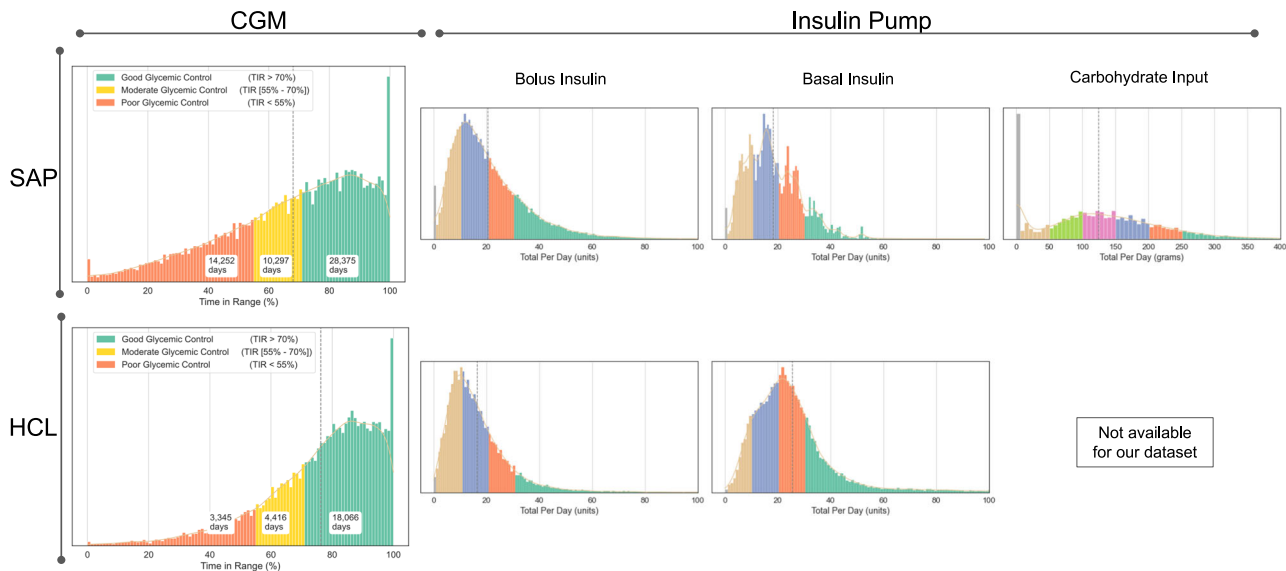


Fig. 2 Overview and distribution of digital data. This figure shows the distributions across subjects in the SAP and HCL population for time in range from the CGMs and total bolus, basal and carbohydrate input per day from the insulin pumps.

Table 1. Summary metrics for digital biomarkers in the SAP ($n = 100$) and HCL ($n = 150$) population.

Dataset	Digital biomarkers	Mean	Median	Std Dev	Range	Skewness	Kurtosis
SAP	Time in range (%)	68.0	72.2	23.2	0–100	−0.71	−0.20
	Time above range (%)	28.1	22.9	24.4	0–100	0.79	−0.18
	Bolus insulin (units per day)	20.4	17.2	14.0	0–163	1.53	4.05
	Basal insulin (units per day)	18.3	16.6	10.3	0–98	0.88	1.36
	Carb. input (grams per day)	124.3	120.0	88.6	0–598	0.49	0.10
HCL	Time in range (%)	76.3	80.4	18.5	0–100	−1.12	1.23
	Time above range (%)	20.1	14.9	19.4	0–100	1.23	1.33
	Bolus insulin (units per day)	16.3	13.7	12.0	0–244	2.78	18.62
	Basal insulin (units per day)	25.5	22.6	17.0	0–185	2.71	11.75

Modeling digital biomarkers and glycemic control

There are well-known behavioral and lifestyle factors that affect glycemic control^{42–44,48–54}. However, there is little research on understanding how trends and individualized patterns amongst lifestyle factors are associated with good and poor glycemic control. This knowledge can be very beneficial for patients with diabetes and can provide actionable insights on what exactly a person can change to achieve better glycemic control on a daily basis^{2,16,27}. To address this critical need, we modeled the impact of digital biomarkers from retrospective CGM data (i.e., using prior data only) and insulin pump data on the present-day to understand daily glycemic control. Our approach uses a form of topic modeling, particularly supervised latent Dirichlet allocation (sLDA)⁵⁵, which was originally proposed for text mining, but has since been extended to other data types and applications^{56–58}. Unlike the standard LDA with documents and words, our approach uses data streams from the CGM (using prior data only) and insulin pump as “documents” and patterns of behaviors (e.g., large daily bolus insulin) as “words” (or features) that are then mapped to impending glycemic control as the outcome variable. Using our defined digital biomarkers ($n = 26$) from the feature categories of prior-day glycemic control (quantified using prior-day TIR and prior-day TAR), and insulin pump features (quantified using daily bolus insulin, daily basal insulin, and daily carbohydrate input), we modeled good (TIR > 70%) and poor (TIR < 55%) glycemic control. Since sLDA is an extension of the unsupervised

latent Dirichlet allocation (LDA)⁵⁹, which employs a random seed for modeling, there can be variation in the results obtained from multiple trials. Hence, we repeated our modeling $k = 3$ times to ensure stable findings on digital biomarkers that impact daily glycemic control.

Figure 3 shows the top 10 digital biomarkers associated with good and poor glycemic control in the SAP and HCL populations. Supplementary Table 2 shows similar results from two additional runs with different random seeds. We observe that a combination of prior-day TIR between 80 and 100% and smaller daily basal and bolus insulin amounts, generally less than 30 units per day, are amongst the top digital biomarkers with the highest probabilities associated with good glycemic control. Meanwhile, a combination of prior-day time above range (TAR) between 36 and 100%, prior-day TIR between 41 and 60%, and larger daily bolus insulin greater than 30 units per day are amongst the top digital biomarkers with the highest probabilities associated with poor glycemic control. These results were found to be stable across repeated evaluations and across both the SAP and HCL dataset.

Classifying glycemic control with digital biomarkers

Predicting glycemic control with digital biomarkers is useful to further evaluate the importance and relevance of the defined features. This knowledge can also provide a greater understanding of glycemic control and actionable insights for persons with diabetes who use continuous digital devices for management of

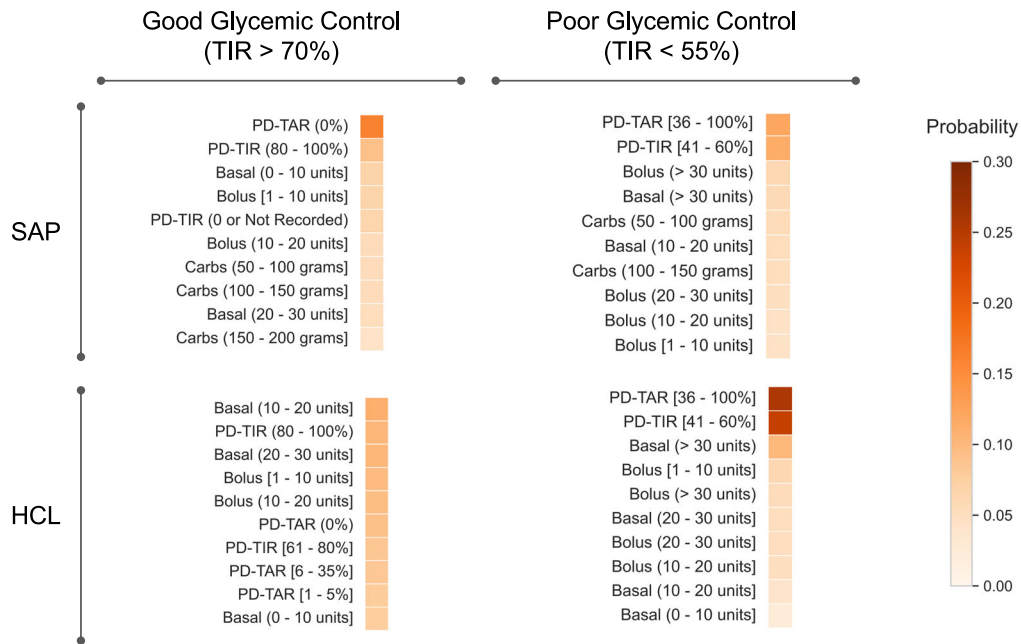


Fig. 3 Top 10 digital biomarkers associated with good and poor glycemic control for the SAP and HCL datasets. This figure shows the probabilities obtained from modeling the defined objective features and daily glycemic control using sLDA.

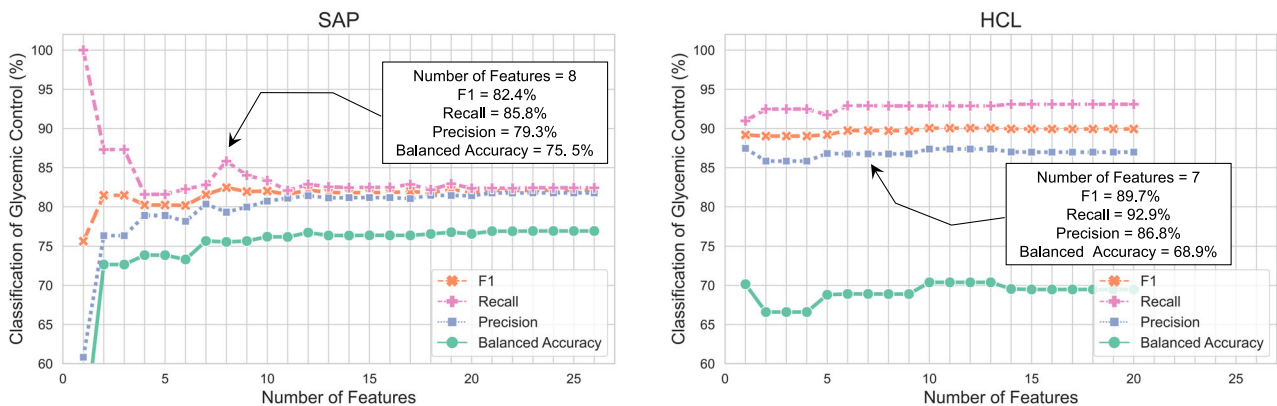


Fig. 4 Feature importance analysis for classification of glycemic control. The classification results for the SAP and HCL dataset plateaued after inclusion of the top 7–8 objective features. This figure shows the results obtained with a logistic regression classifier.

their condition. Toward this goal, we started with the single top-ranked feature in each dataset, as shown in Fig. 3 and Supplementary Table 3, then incrementally added one feature with the next highest probability for iterative classification of good and poor glycemic control. As shown in Fig. 2, our dataset includes a total of 42,817 and 20,879 days of good plus poor glycemic control for the SAP and HCL populations, respectively. We evaluated several classifiers using an 80/20 train/test split for classification of good versus poor glycemic control and observed that the results plateaued after inclusion of 7–8 features for both datasets. Figure 4 shows the results of this analysis for the logistic regression classifier while Supplementary Table 5 shows the results with other classifiers including support vector machine, decision tree, and a linear discriminant analysis classifier. For classification of glycemic control, we achieved an F1 score of 82.4% and balanced accuracy of 75.5% (recall = 85.8% and precision = 79.3%) for the SAP dataset using the top 8 features. Meanwhile, we achieved an F1 score of 89.7% and balanced accuracy of 68.9% (recall = 92.9% and precision = 86.8%) for the HCL dataset using the top 7 features. Supplementary Fig. 1 shows the confusion matrix for results obtained with the logistic

regression classifier. From Fig. 4, we note that the classification results for the HCL dataset was relatively stable (i.e., <5% difference) independent of the number of features used. This is likely due to significant class imbalance in the HCL dataset with approximately 4-times more occurrences of good glycemic control (TIR > 70%) as opposed to poor glycemic control (TIR < 55%) - see Fig. 2. The presence of more days with good glycemic control, which equates to class imbalance, in the HCL dataset is representative of real use of this system for diabetes management. Thus, we calculated and reported the balanced accuracy⁶⁰ in our analysis. Our results indicate that the top 7–8 features shown in Fig. 3 and Supplementary Table 3 are useful digital biomarkers that impact glycemic control. Some of these features are modifiable behaviors and thus provide actionable insights on what a person with diabetes can change to achieve better glycemic control.

Identifying the most impactful digital biomarkers that impact glycemic control

Similar to traditional biomarkers, digital biomarkers are objective measures that can be used to explain, influence, or predict

health-related outcomes³⁰. However, unlike traditional biomarkers that provide a “snapshot view” based on limited measurements collected over time, digital biomarkers are often derived from longitudinal and continuous measurements, and thus can capture dynamic changes in health and related outcomes^{21–26,30,61}. Consequently, an important but critical part of the process of developing digital biomarkers is sifting through large volumes of often heterogeneous data sources to identify pertinent features or measures. Through our systematic approach of defining objective features from digital data (particularly CGM and insulin pump data), modeling digital biomarkers and glycemic control, and classifying glycemic control with digital biomarkers, we learned that prior-day time above range (PD-TAR) and prior-day time in range (PD-TIR) are most predictive digital biomarkers for both good and poor glycemic control on a given day. As shown in Fig. 3 and in Supplementary Tables 2 and 3, good glycemic control with PD-TAR equal to 0% and PD-TIR in the range of 80–100% was most predictive of good glycemic control for the next day. Conversely, poor glycemic control with PD-TAR in the range of 36–100% and PD-TIR in the range of 41–60% was most predictive of poor glycemic control for the following day. The next most impactful digital biomarker observed through multiple iterations of modeling was related to the daily total units of basal and bolus insulin. More specifically, a daily total of less than 30 units of basal and bolus insulin was often associated with good glycemic control while a daily total of greater than 30 units was often associated with poor glycemic control. Although carbohydrate input data was only present for the SAP dataset, our results show daily total of carbs in the range of 50–100 g and 100–150 g was often associated with poor glycemic control. The aforementioned digital biomarkers were consistently amongst the top 5 features associated with good or poor glycemic control. Thus, a combination of these features proved to be particularly useful for predicting glycemic control.

DISCUSSION

Diabetes is a prevalent chronic condition that is increasingly affecting more children, youths, and adults, both nationally and globally^{17–20}. Management of this life-long condition is a daily challenge that requires maintaining good glycemic control to reduce the occurrence of short- and long-term complications⁷. However, research shows that less than one-third of people living with type 1 diabetes are consistently achieving the recommended goals for glycemic control¹⁰. As a result, digital technology like CGMs, insulin pumps, and automated insulin delivery systems are increasingly prescribed and used by persons with diabetes to reduce the occurrence of adverse health events and the burden of management^{9,10}. The routine use of ubiquitous digital devices provides unique opportunities to gain a greater understanding of the condition, the impact of behavioral and lifestyle choices, and develop digital biomarkers from longitudinal and continuous measurements to further support patients, families, and care teams with improving management.

In this work, we present a scalable computational framework for leveraging digital data from wearable devices to discover and identify digital biomarkers for understanding daily glycemic control. Our proposed framework includes defining objective features from digital data, modeling the digital features with clinically-validated metrics for glycemic control, and investigating the relative importance of the defined features for predicting glycemic control. From implementing and evaluating the proposed framework on large SAP and HCL datasets, our results showed that features from the CGM, particularly prior-day TIR and prior-day TAR, are the most predictive digital biomarkers for good and poor glycemic control. We also found that insulin-related features, particularly total daily basal insulin and total daily bolus insulin, are relevant digital biomarkers of

glycemic control. For classification of glycemic control, we found that the top-ranked 7–8 digital biomarkers shown in Fig. 3 were sufficient for prediction and achieved an F1 score of 82.4% in the SAP dataset and 89.7% in the HCL dataset. A natural extension of this work is to leverage the proposed framework for discovering more digital biomarkers of daily glycemic control by studying the relationship and impact of other relevant factors like activity level, sleep, stress, circadian rhythm, and more, many of which can be monitored with consumer-grade wearable devices like Fitbit^{62,63}.

Compared to prior and related work, our research is unique for three reasons. First, we propose and evaluate a scalable computational framework for leveraging large amounts of continuously sensed and grossly underutilized digital data to discover biomarkers of daily glycemic control. This work sets a precedence for use of routinely collected digital data to understand the complex combination of factors that impact personalized diabetes management. Second, this research leverages a large dataset with 250 subjects and over 79,000 days of data from two unique patient populations (i.e., SAP and HCL) to evaluate the proposed computational framework. Our dataset is orders of magnitude larger than those used in related studies^{2,27,31–34}. In addition, unlike prior work on digital biomarkers in various health domains^{23,26,56,57}, our study does not rely on self-reported data which is often biased, instead we use data from clinically-validated medical devices. Third, since our dataset was collected through routine use of wearable devices for daily management of diabetes, this observational data can reveal behaviors/factors in the natural environment that are not influenced or altered as a result of enrolling in a research study⁶⁴. For example, exploratory analysis on the distributions of CGM data in different populations showed more occurrence of high TIR (i.e., good glycemic control) for patients with HCL systems compared to patients with SAP therapy (see Fig. 2). This finding aligns with results from prior literature³⁹. Additionally, we explored differences in glycemic control across age groups by stratifying subjects in this study into four groups (i.e., ages 2–14, 15–29, 30–40, and 40+). Figure 5 shows that glycemic control improved with age and was worse in the youngest group of subjects between the ages of 2–14. This trend is particularly evident in the SAP population and less prevalent in the HCL population. It is worth noting that our proposed computational framework, modeling, and results are interpretable and provide actionable insights. For example, our research showed that prior-day glycemic control is the most predictive digital biomarker associated with present-day glycemic control. This finding is intuitive, interpretable, and in alignment with results from recent research involving patients with type 1 diabetes who use a closed-loop control system (as opposed to SAP or HCL as in this study) for daily management of their condition⁶⁵.

However, despite the contributions of this research, there are some limitations that are worth noting. One of such limitation relates to representation in our study cohort. More specifically, it is well-known that individuals who have access to digital devices may not represent the full spectrum of the population of interest, especially in health-related conditions⁶⁴. For example, Foster et al. found that racial differences and disparities exist in the use of technology for health management, including the use of insulin pumps and CGMs for diabetes management¹⁰. Although the race and ethnic distribution of subjects in our study is unknown, it is important to recognize that there may be some underlying disparities. In addition, adherence of personal health devices is often varied because people do not always use or wear these devices; this is also the case for wearable medical devices like those used in this study^{66,67}. In addition, our dataset has less than 10% of missing data, hence it may represent more motivated patients who are already interested in leveraging digital data to improve management of their diabetes. Lastly, it is well-known

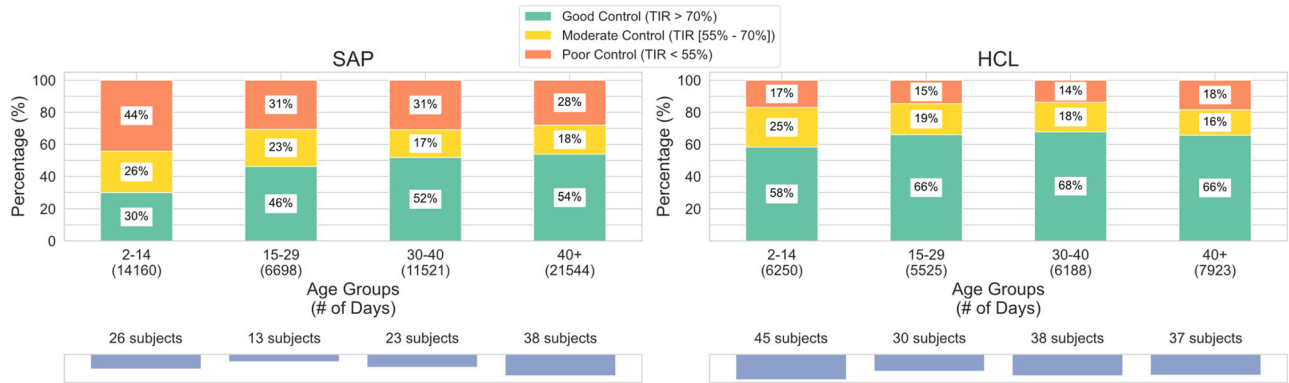


Fig. 5 Glycemic control by age group. Glycemic control improved with age in the SAP population and was worse in the youngest group of subjects between the ages 2–14 across both the SAP and HCL populations.

Table 2. Summary of study dataset ($n = 250$ subjects with diabetes) including descriptive characteristics of the sensor augmented (SAP) therapy and hybrid closed loop (HCL) populations.

		SAP		HCL	
*Note: NR = Not Reported		Mean \pm SD	Range	Mean \pm SD	Range
Subject	Total no. of subjects	100		150	
	Male/Female/NR	23/13/64		47/66/37	
	T1D/T2D/NR	41/0/59		142/0/8	
	Age of subjects	34.38 \pm 20.57	2–76	28.56 \pm 16.19	3–75
	Yrs. w/ diabetes	18.41 \pm 16.68	0–60	12.07 \pm 15.32	0–50
	No. of days/subject	521.73 \pm 179.71	306–1087	171.7 \pm 103.58	94–683
CGM	Total BG samples	14,416,242		7,181,253	
	Samples/subject	144,162 \pm 49,346	85,251–294,476	47,875 \pm 28,846	25,861–194,858
Insulin Pump	Total bolus doses	383,868		200,888	
	Bolus doses/subject	3838.68 \pm 1948.48	1227–10,857	1339 \pm 841	451–5467
	Total carb-inputs	208,582		Not in Dataset (N/A)	
	Carb-inputs/subject	2085.82 \pm 1529.87	0–9766	(N/A)	(N/A)
	Total 24-h basal doses	51,231		25,879	
24-h basal/subject	512.31 \pm 193.78	224–1139	172.53 \pm 104.59	94–699	

that differences in the type and quality of carbohydrates (and other macronutrients) can affect blood glucose differently. However, given that our dataset comes directly from CGMs and insulin pumps, it does not include information on the type and quality of meals.

Building on this research, there are several important avenues to be explored in future work. For example, this study only included digital data from CGMs and insulin pumps, thus, a compelling future direction is to integrate other streams of data that capture relevant behaviors that affect glycemic control such as activity, sleep, stress, etc. This data fusion can be achieved by leveraging wearable activity trackers in addition to CGMs and insulin pumps. Another future direction includes broadening representation in the patient cohort and conducting a prospective study to further validate the insights learned from this research. Finally, it is important to evaluate how knowledge from this and similar research can be useful to the larger population of people and patients who may not use continuous monitoring devices for daily management of their health. This includes the vast population of patients with type 2 diabetes who use glucometers for point-in-time measurements of the blood glucose and insulin pens for multiple daily injections.

METHODS

Dataset

Table 2 summarizes our dataset which includes 250 subjects with diabetes, primarily type 1 diabetes (age = 2–76 years, time with diabetes = 0–60 years). This dataset is from the Tidepool Big Data Donation Project⁶⁸. It was licensed and approved for use in this study by the Committee for Protection of Human Subjects at Dartmouth College. The experiments and methods are carried out in accordance with all ethical regulations and the license agreement. In addition, all subjects provided informed consent for use of their diabetes device data. The dataset represents two populations, namely, patients who use SAP therapy^{35,69} ($n = 100$) and patients who use HCL^{36,37,39} ($n = 150$) for daily management of their diabetes. On average, there was 522 days (≈ 1 year and 5 months) of data per subject in the SAP population and 172 days (≈ 5 –6 months) of data per subject in the HCL population. In both populations, patients use CGMs, which are clinically-validated and minimally-invasive wearable devices for continuous monitoring of their blood glucose^{14,40}. CGMs on the market today (e.g., those by Dexcom⁷⁰, Medtronic⁷¹, and Abbott⁷²) record 1 blood glucose sample every 1–5 min. Thus, our dataset includes an average of 144,162 blood glucose samples per subject (total = 14.4 million samples across 100 subjects) in the SAP population and an average of 47,875 blood glucose samples per subject (total = 7.1 million samples across 150 subjects) in the HCL population.

In addition, all subjects also used insulin pumps for daily management of their diabetes. Hence, the insulin pump data includes basal, bolus, and

meal-related entries (or carb-inputs) for each subject. Basal insulin³⁸ refers to small amounts of background insulin infused from the pump either on a programmed schedule as in the SAP population or automatically adjusted by an algorithm as in the HCL population. Meanwhile, bolus insulin refers to larger amounts of rapid-acting insulin required to cover meals and correct for high blood glucose levels³⁸. Our dataset includes an average of 3838 bolus insulin doses per subject and 2085 carb-inputs per subjects (total = 383,868 insulin doses and 208,582 carb inputs across 100 subjects) in the SAP population, and 1339 bolus insulin doses per subject and unknown carb-input data (total = 200,888 bolus insulin doses across 150 subjects) in the HCL population. Since basal insulin is administered at an infusion rate (i.e., units per hour) which can vary for different time periods in the day, we calculated basal insulin amounts for 24-h periods in our dataset, thus the 24-h basal per subject in both dataset is approximately equal to the number of days of data per subject. This rich dataset represents an example of the prevalence of routinely collected digital data that is grossly underutilized in healthcare. Our research provides a scalable computational framework to facilitate improved use of such data to understand the complex combination of factors that affect diabetes management and discover digital biomarkers of glycemic control.

Defining objective features from digital data

In this work, we use CGM and insulin pump data described in Table 2 to define objective features for modeling and predicting daily glycemic control. Our feature set comprises of digital biomarkers from five categories outlined in Table 1, including two from prior-day CGM (i.e., time in range and time above range) and three from the current-day insulin pump (i.e., total bolus, basal, and carb inputs per day). Each feature category was split into 4–7 unique ranges based on the data distribution across samples from all subjects, thus yielding discrete features like prior-day TIR between 41 and 60% or total bolus insulin per day between 10 and 20 units. Supplementary Table 1 outlines all 26 features defined in this work. Prior-day CGM features were based on domain-knowledge which supports the use of time in different glycemic ranges as a validated and clinically-relevant metric for accessing diabetes management^{40,73,74}. More specifically, TIR - defined as the percentage of blood glucose readings from a CGM between 70 and 180 mg/dL, and TAR - defined as the percentage of blood glucose readings greater than 180 mg/dL⁴⁰, were used. Both metrics were calculated on a daily basis (i.e., within a 24-h window from 12AM - 11:59PM). In this study, we did not include time below range (TBR) as a separate feature because it is indirectly accounted for by the inclusion of TIR and TAR. In addition to the aforementioned retrospective data from the CGMs, we also included insulin pump data from the current day of interest for the goal of learning data-driven insights that can inform in-the-moment decision making for persons with diabetes. All features were defined to quantify behaviors/factors such as adherence to self-management tasks like meal announcements⁴⁵ or bolus insulin dosing⁴⁶ and dietary habits like conforming to low-carb or high-carb diets^{48,75}, each of which can have an impact on daily glycemic control and long-term diabetes outcomes.

Modeling digital biomarkers and glycemic control

To further understand how our defined features are associated with impending glycemic control, we used sLDA⁵⁵, a form of probabilistic topic modeling for analysis⁷⁶. Our approach builds on prior research that has shown the applicability of LDA and sLDA for modeling different health-relevant data types^{56–58}. However, in this study, sLDA was used to infer latent features (i.e., input variables) that are predictive of good and poor glycemic control (i.e, the response variable). The clinically-validated metric, TIR, was used to stratify both good and poor glycemic control^{40,73,74}. More specifically, the clinical target of TIR > 70% was classified as good glycemic control, while TIR < 55% was classified as poor glycemic control (see Fig. 2). In the SAP population, there were a total of 17,915 days with poor glycemic control and 24,902 days with good glycemic control. Conversely, in the HCL population, there were 4178 days with poor glycemic control and 16,701 days with good glycemic control. Every digital feature and the associated response variable was calculated on a daily basis. There were a total of 26 features from the SAP dataset and 20 features from the HCL dataset which was missing the carbohydrate input data stream (see Supplementary Tables 1 and 3). The tomtopy toolkit⁷⁷ in Python was used to implement sLDA for modeling. Following this, term-frequency inverse document

frequency (tf-idf) was used to weigh each input to adjust for the fact that some features appear more frequently in the dataset⁷⁸. Finally, as a validation step, sLDA was run multiple times ($k=3$) with different random seeds to ensure stable and reproducible findings on the ranking and associations between digital biomarkers and glycemic control (see Fig. 3 and Supplementary Table 2).

Classifying glycemic control with the most impactful digital biomarkers

To further investigate the potential of the defined objective features as digital biomarkers of glycemic control, we leveraged these features for classification. More specifically, we partitioned both datasets (i.e., SAP & HCL) into a training and test set using an 80/20 split (see Supplementary Table 4 for a breakdown of the number of samples in each partition). Following this, we evaluated and compared the classification results using 4 interpretable and explainable classic models^{79,80}, namely, logistic regression, support vector machine with a radial basis function, decision tree, and linear discriminant analysis. In this work, we also evaluated the importance of each feature by starting with the single top-ranked feature in each dataset and incrementally adding another feature with the next highest probability for iterative classification of glycemic control until all features were exhausted. Figure 3 shows that the top-ranked feature in the SAP dataset was prior-day TAR = 0% which is associated with good glycemic control while the top-ranked feature in the HCL dataset is prior-day TAR between 36 and 100% which is associated with poor glycemic control. The scikit-learn machine learning library in Python was used for implementation⁸¹. For performance evaluation, we used established metrics including F1 score, Recall, and Precision. However, we also reported the balanced accuracy⁶⁰ to account for inherent skewness and class imbalance in the datasets as shown in Fig. 2. As a final validation step, we compared top-ranked features across both the SAP and HCL datasets, to assess consistency in our findings and identify the most impactful digital biomarkers for daily glycemic control.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

The data analyzed in this study is available through the Tidepool Big Data Donation Project⁶⁸ but restrictions may apply. While data sharing agreements prohibit the authors from making the dataset publicly available, confidential access may be granted to those who meet specific criteria, pending permission from Tidepool. Access can be requested through the corresponding author, T.P.

CODE AVAILABILITY

The code generated for this study is available upon request at the Augmented Health Lab at Dartmouth College (www.ah-lab.cs.dartmouth.edu/).

Received: 16 November 2021; Accepted: 14 July 2022;

Published online: 08 August 2022

REFERENCES

1. Brown, A. 42 factors that affect blood glucose. <https://diatribe.org/42factors>. Accessed Nov 2021.
2. 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).
3. American Diabetes Association and others. Good to know: factors affecting blood glucose. *Clin. Diabetes* **36**, 202–202 (2018).
4. Wolkowicz, K. L. et al. A review of biomarkers in the context of type 1 diabetes: biological sensing for enhanced glucose control. *Bioeng. Transl. Med.* **6**, e10201 (2021).
5. Safford, M. M., Russell, L., Suh, D.-C., Roman, S. & Pogach, L. How much time do patients with diabetes spend on self-care? *J Am Board Family Pract.* **18**, 262–270 (2005).
6. Digitale, E. New research shows how to keep diabetics safer during sleep. <https://scopeblog.stanford.edu/2014/05/08/new-research-keeps-diabetics-safer-during-sleep/>. Accessed Aug 2021.

7. Harding, J. L., Pavkov, M. E., Magliano, D. J., Shaw, J. E. & Gregg, E. W. Global trends in diabetes complications: a review of current evidence. *Diabetologia* **62**, 3–16 (2019).
8. Ravaut, M. et al. Predicting adverse outcomes due to diabetes complications with machine learning using administrative health data. *NPJ Digital Med.* **4**, 1–12 (2021).
9. van den Boom, L. et al. Temporal trends and contemporary use of insulin pump therapy and glucose monitoring among children, adolescents, and adults with type 1 diabetes between 1995 and 2017. *Diabetes Care* **42**, 2050–2056 (2019).
10. Foster, N. C. et al. State of type 1 diabetes management and outcomes from the t1d exchange in 2016–2018. *Diabetes Technol. Therapeutics* **21**, 66–72 (2019).
11. Landau, Z., Raz, I., Wainstein, J., Bar-Dayana, Y. & Cahn, A. The role of insulin pump therapy for type 2 diabetes mellitus. *Diabetes Metabol. Res. Rev.* **33**, e2822 (2017).
12. Taylor, P. J., Thompson, C. H. & Brinkworth, G. D. Effectiveness and acceptability of continuous glucose monitoring for type 2 diabetes management: a narrative review. *J. Diabetes Invest.* **9**, 713–725 (2018).
13. DeSalvo, D. & Buckingham, B. Continuous glucose monitoring: current use and future directions. *Current Diabetes Rep.* **13**, 657–662 (2013).
14. Danne, T. et al. International consensus on use of continuous glucose monitoring. *Diabetes Care* **40**, 1631–1640 (2017).
15. Wong, J. C., Neinstein, A. B., Spindler, M. & Adi, S. A minority of patients with type 1 diabetes routinely downloads and retrospectively reviews device data. *Diabetes Technol. Therapeutics* **17**, 555–562 (2015).
16. Bartolome, A., Shah, S. & Prioleau, T. Glucomine: a case for improving the use of wearable device data in diabetes management. *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.* **5**, 1–24 (2021).
17. Centers for Disease Control and Prevention and Others. *National diabetes statistics report, 2020* (Atlanta, GA: Centers for Disease Control and Prevention, US Department of Health and Human Services, 2020) pp. 12–15.
18. Lawrence, J. M. et al. Trends in prevalence of type 1 and type 2 diabetes in children and adolescents in the US, 2001–2017. *JAMA* **326**, 717–727 (2021).
19. Xu, G. et al. Prevalence of diagnosed type 1 and type 2 diabetes among US adults in 2016 and 2017: population based study. *BMJ* **362**, <https://doi.org/10.1136/bmj.k1497> (2018).
20. Mobasser, M. et al. Prevalence and incidence of type 1 diabetes in the world: a systematic review and meta-analysis. *Health Promotion Perspect.* **10**, 98 (2020).
21. Kourtis, L. C., Regele, O. B., Wright, J. M. & Jones, G. B. Digital biomarkers for alzheimer's disease: the mobile/wearable devices opportunity. *NPJ Digital Med.* **2**, 1–9 (2019).
22. Mandryk, R. L. & Birk, M. V. The potential of game-based digital biomarkers for modeling mental health. *JMIR Mental Health* **6**, e13485 (2019).
23. Jacobson, N. C., Weingarden, H. & Wilhelm, S. Digital biomarkers of mood disorders and symptom change. *npj Digital Med.* **2**, 1–3 (2019).
24. Dagum, P. Digital biomarkers of cognitive function. *NPJ Digital Med.* **1**, 1–3 (2018).
25. Mahadevan, N. et al. Development of digital biomarkers for resting tremor and bradykinesia using a wrist-worn wearable device. *NPJ Digital Med.* **3**, 1–12 (2020).
26. Jacobson, N. C., Summers, B. & Wilhelm, S. Digital biomarkers of social anxiety severity: digital phenotyping using passive smartphone sensors. *J. Med. Internet Res.* **22**, e16875 (2020).
27. Bent, B. et al. Engineering digital biomarkers of interstitial glucose from non-invasive smartwatches. *NPJ Digital Med.* **4**, 1–11 (2021).
28. Avram, R. et al. A digital biomarker of diabetes from smartphone-based vascular signals. *Nat. Med.* **26**, 1576–1582 (2020).
29. Dorsey, E. R. Digital biomarkers. Basel, Karger. <https://www.karger.com/Journal/Home/271954> (2018).
30. Babrak, L. M. et al. Traditional and digital biomarkers: two worlds apart? *Digital Biomarkers* **3**, 92–102 (2019).
31. Gadaleta, M., Facchinetti, A., Grisan, E. & Rossi, M. Prediction of adverse glycemic events from continuous glucose monitoring signal. *IEEE J. Biomed. Health Inform.* **23**, 650–659 (2018).
32. Dave, D. et al. Feature-based machine learning model for real-time hypoglycemia prediction. *J. Diabetes Sci. Technol.* **15**, 842–855 (2021).
33. Mujahid, O., Contreras, I. & Vehi, J. Machine learning techniques for hypoglycemia prediction: trends and challenges. *Sensors* **21**, 546 (2021).
34. Kodama, S. et al. Ability of current machine learning algorithms to predict and detect hypoglycemia in patients with diabetes mellitus: meta-analysis. *JMIR Diabetes* **6**, e22458 (2021).
35. Steineck, I., Ranjan, A., Nørgaard, K. & Schmidt, S. Sensor-augmented insulin pumps and hypoglycemia prevention in type 1 diabetes. *J. Diabetes Sci. Technol.* **11**, 50–58 (2017).
36. Leelarathna, L. et al. Hybrid closed-loop therapy: where are we in 2021? *Diabetes Obesity Metabol.* **23**, 655–660 (2021).
37. Weaver, K. W. & Hirsch, I. B. The hybrid closed-loop system: evolution and practical applications. *Diabetes Technol. Therapeutics* **20**, S2–16–S2–23 (2018).
38. McAdams, B. H. & Rizvi, A. A. An overview of insulin pumps and glucose sensors for the generalist. *J. Clin. Med.* **5**, 5 (2016).
39. Akturk, H. K. et al. Long-term real-life glycaemic outcomes with a hybrid closed-loop system compared with sensor-augmented pump therapy in patients with type 1 diabetes. *Diabetes Obesity Metabol.* **22**, 583–589 (2020).
40. 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).
41. Brazeau, A. et al. Carbohydrate counting accuracy and blood glucose variability in adults with type 1 diabetes. *Diabetes Res Clin Pract* **99**, 19–23 (2013).
42. Lehmann, V. et al. Lower daily carbohydrate intake is associated with improved glycemic control in adults with type 1 diabetes using a hybrid closed-loop system. *Diabetes Care* **43**, 3102–3105 (2020).
43. Schmidt, S. et al. Low versus high carbohydrate diet in type 1 diabetes: a 12-week randomized open-label crossover study. *Diabetes Obesity Metabol* **21**, 1680–1688 (2019).
44. Wylie-Rosett, J., Aebersold, K., Conlon, B., Isasi, C. R. & Ostrovsky, N. W. Health effects of low-carbohydrate diets: where should new research go? *Current Diabetes Rep.* **13**, 271–278 (2013).
45. O'Connell, M., Donath, S. & Cameron, F. Poor adherence to integral daily tasks limits the efficacy of CSII in youth. *Pediatric Diabetes* **12**, 556–559 (2011).
46. Olinder, A. L., Kernell, A. & Smide, B. Missed bolus doses: devastating for metabolic control in csii-treated adolescents with type 1 diabetes. *Pediatric Diabetes* **10**, 142–148 (2009).
47. Ramkissoon, C. M., Herrero, P., Bondia, J. & Vehi, J. Unannounced meals in the artificial pancreas: detection using continuous glucose monitoring. *Sensors* **18**, 884 (2018).
48. Evert, A. B. et al. Nutrition therapy recommendations for the management of adults with diabetes. *Diabetes Care* **37**, S120–S143 (2014).
49. Sardinha, L. B., Magalhães, J. P., Santos, D. A. & Júdice, P. B. Sedentary patterns, physical activity, and cardiorespiratory fitness in association to glycemic control in type 2 diabetes patients. *Front. Physiol.* **8**, 262 (2017).
50. Nguyen, T. et al. Fitness and physical activity in youth with type 1 diabetes mellitus in good or poor glycemic control. *Pediatric Diabetes* **16**, 48–57 (2015).
51. Reutrakul, S. et al. Sleep characteristics in type 1 diabetes and associations with glycemic control: systematic review and meta-analysis. *Sleep Med.* **23**, 26–45 (2016).
52. Perez, K. M. et al. Sleep in type 1 diabetes: implications for glycemic control and diabetes management. *Current Diabetes Rep.* **18**, 1–8 (2018).
53. Lloyd, C., Smith, J. & Weinger, K. Stress and diabetes: a review of the links. *Diabetes Spectrum* **18**, 121–127 (2005).
54. Sharif, K. et al. Psychological stress and type 1 diabetes mellitus: what is the link? *Expert Rev. Clin. Immunol.* **14**, 1081–1088 (2018).
55. McAuliffe, J. D. & Blei, D. M. Supervised topic models. *Adv. Neural Inf. Process. Syst.* **20**, 1–8 (2007).
56. Nosakhare, E. & Picard, R. Probabilistic latent variable modeling for assessing behavioral influences on well-being. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2718–2726 (ACM, 2019).
57. Nosakhare, E. & Picard, R. Toward assessing and recommending combinations of behaviors for improving health and well-being. *ACM Trans. Comput. Healthcare* **1**, 1–29 (2020).
58. Yazdavar, A. H. et al. Semi-supervised approach to monitoring clinical depressive symptoms in social media. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, 1191–1198 (ACM, 2017).
59. Blei, D. M., Ng, A. Y. & Jordan, M. I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003).
60. Sokolova, M., Japkowicz, N. & Szpakowicz, S. Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation. In *Australasian Joint Conference on Artificial Intelligence*, 1015–1021 (Springer, 2006).
61. Coravos, A., Khozin, S. & Mandl, K. D. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. *npj Digital Med.* **2**, 1–5 (2019).
62. Lee, J.-M., Byun, W., Keill, A., Dinkel, D. & Seo, Y. Comparison of wearable trackers' ability to estimate sleep. *Int. J. Environ. Res. Public Health* **15**, 1265 (2018).
63. Reddy, R. K. et al. Accuracy of wrist-worn activity monitors during common daily physical activities and types of structured exercise: evaluation study. *JMIR mHealth uHealth* **6**, e10338 (2018).
64. Hicks, J. L. et al. Best practices for analyzing large-scale health data from wearables and smartphone apps. *npj Digital Med.* **2**, 1–12 (2019).
65. Schoelwer, M. J. et al. Predictors of time-in-range (70–180 mg/dl) achieved using a closed-loop control system. *Diabetes Technol. Therapeutics* **23**, 475–481 (2021).
66. 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 (ACM, 2020).
67. Giani, E., Snelgrove, R., Volkening, L. K. & Laffel, L. M. Continuous glucose monitoring (CGM) adherence in youth with type 1 diabetes: associations with biomedical and psychosocial variables. *J. Diabetes Sci. Technol.* **11**, 476–483 (2017).

68. Tidepool. Tidepool big data donation project. <https://www.tidepool.org/bigdata>. Accessed Nov 2021.
69. Ly, T. T. et al. Effect of sensor-augmented insulin pump therapy and automated insulin suspension vs standard insulin pump therapy on hypoglycemia in patients with type 1 diabetes: a randomized clinical trial. *JAMA* **310**, 1240–1247 (2013).
70. Dexcom. Dexcom continuous glucose monitoring. <https://www.dexcom.com/>. Accessed Nov 2021.
71. Medtronic. The guardian connect system. <https://www.medtronicdiabetes.com/products/guardian-connect-continuous-glucose-monitoring-system>. Accessed Nov 2021.
72. Abbott. Freestyle libre continuous glucose monitor. <https://www.abbott.com/freestyle-libre-2-continuous-glucose-monitor-cgm.html>. Accessed Nov 2021.
73. Beck, R. W. et al. Validation of time in range as an outcome measure for diabetes clinical trials. *Diabetes Care* **42**, 400–405 (2019).
74. Dovc, K. & Battelino, T. Time in range centered diabetes care. *Clin. Pediatric Endocrinol.* **30**, 1–10 (2021).
75. Russell, W. R. et al. Impact of diet composition on blood glucose regulation. *Critical Rev. Food Sci. Nutr.* **56**, 541–590 (2016).
76. Blei, D. M. Probabilistic topic models. *Commun. ACM* **55**, 77–84 (2012).
77. bab2min & Fenstermacher, D. bab2min/tomotopy: 0.10.0. <https://doi.org/10.5281/zenodo.4362292> (2020).
78. Rajaraman, A. & Ullman, J. D. *Data Mining*, 1–17 (Cambridge University Press, 2011).
79. Ahmad, M. A., Eckert, C. & Teredesai, A. Interpretable machine learning in healthcare. In *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, 559–560 (ACM, 2018).
80. Vellido, A. The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural Comput. Appl.* **32**, 18069–18083 (2020).
81. Pedregosa, F. et al. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
82. JDRF. Drugs and devices for glucose control. <https://www.jdrf.org/drugs-and-devices-for-glucose-control/>. Accessed Jul 2022.

ACKNOWLEDGEMENTS

The authors would like to thank Tidepool⁶⁸ and JDRF⁸² for data contribution and support which makes this and other diabetes-related research possible. This work is partly supported by the National Science Foundation (award #2037267).

AUTHOR CONTRIBUTIONS

All authors edited and reviewed the final manuscript. A.B. conceived the study, conducted data analysis, data interpretation, and contributed to manuscript preparation. T.P. was involved in concept development, funding, data collection, analysis, and interpretation, and manuscript preparation.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41746-022-00656-z>.

Correspondence and requests for materials should be addressed to Temiloluwa Prioleau.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022, corrected publication 2022