

ARTICLE OPEN



Engineering digital biomarkers of interstitial glucose from noninvasive smartwatches

Brinnae Bent¹, Peter J. Cho¹, Maria Henriquez², April Wittmann³, Connie Thacker³, Mark Feinglos³, Matthew J. Crowley³ and Jessilyn P. Dunn^{1,4}✉

Prediabetes affects one in three people and has a 10% annual conversion rate to type 2 diabetes without lifestyle or medical interventions. Management of glycemic health is essential to prevent progression to type 2 diabetes. However, there is currently no commercially-available and noninvasive method for monitoring glycemic health to aid in self-management of prediabetes. There is a critical need for innovative, practical strategies to improve monitoring and management of glycemic health. In this study, using a dataset of 25,000 simultaneous interstitial glucose and noninvasive wearable smartwatch measurements, we demonstrated the feasibility of using noninvasive and widely accessible methods, including smartwatches and food logs recorded over 10 days, to continuously detect personalized glucose deviations and to predict the exact interstitial glucose value in real time with up to 84% and 87% accuracy, respectively. We also establish methods for designing variables using data-driven and domain-driven methods from noninvasive wearables toward interstitial glucose prediction.

npj Digital Medicine (2021)4:89; <https://doi.org/10.1038/s41746-021-00465-w>

INTRODUCTION

There is currently no commercially-available noninvasive method for monitoring glycemic status; specifically, noninvasive glucose monitoring to inform glycemic self-management, particularly for prediabetics, is lacking. Prediabetes affects over one-third of people in the United States¹. While prediabetes is highly prevalent and has serious consequences, it is also seriously underdiagnosed—only ten percent of people with prediabetes are aware that they have the disease². For those who have been diagnosed, prediabetes is often poorly managed^{3–5}, leading to 70% of individuals with prediabetes to eventually develop type 2 diabetes (T2D) and to a 10% annual conversion rate from prediabetes to T2D⁶. Prediabetes is reversible with lifestyle modifications: the Diabetes Prevention Program reduced diabetes incidence by 58% through interventions aimed at weight loss, dietary change, and physical activity in patients with prediabetes⁷. Recently, monitoring of blood glucose levels has been added to several prediabetes treatment plans^{8,9} and tracking of blood glucose is even being used by those with normal glucose levels in order to better understand and track glycemic and metabolic health^{10,11}. Long-term lifestyle changes are more likely when patients self-monitor their blood glucose¹², and this practice can be upheld by easily accessible methods for glycemic health monitoring that aid in self-management of prediabetes.

Glucose control is often a main objective of diabetes management, which includes regulating glucose variability and avoiding glucose deviations. Definitions of glucose deviations, including ‘hyperglycemia’ (glucose that is too high) and ‘hypoglycemia’ (glucose that is too low), have been widely cited in literature pertaining to type 1 diabetes (T1D) and T2D^{5,13–24}. Hyperglycemia is traditionally defined as having non-fasting glucose above 180 mg/dL and hypoglycemia is traditionally defined as having non-fasting glucose below 70 mg/dL²⁵. These definitions were developed for diabetes management¹⁹. However, these may not be adequate to explain glucose deviations in individuals with

prediabetes and normal hemoglobin A1c (HbA1c) due to lower fasting glucose levels and lower glucose variability than in those with diabetes. The importance of personalization of glycemic health has been demonstrated previously using ‘glucotypes’ to describe intraindividual differences in glycemic responses²⁶. Management of glucose fluctuations begins with an understanding how specific behaviors influence a person’s own blood glucose levels. Current methods for monitoring blood glucose²⁷, including blood glucose meters and continuous glucose monitors, are not frequently utilized by non-insulin-dependent patients due to their inconvenience, invasiveness²⁸, and high cost. There is currently no commercially-available noninvasive tool to estimate interstitial glucose in real-time.

Non-invasive wrist-worn biometric monitoring technologies²⁹, often referred to as ‘wearables’³⁰, are becoming nearly ubiquitous in the United States, with 117 million currently in use and an expected 100% growth in the next three years³¹. Because of this widespread use, wearables can enable digital biomarker discovery which will facilitate detection and monitoring of chronic diseases³⁰. Digital biomarkers are digitally collected data (e.g., heart rate measurements from a wearable) that may be used as indicators of health outcomes (e.g., prediabetes)³². Digital biomarker algorithms support the aggregation of high-resolution, intra-individual data into summary metrics that are interpretable and actionable³³.

There are many competing factors that affect blood glucose levels, including diet³⁴, activity and exercise³⁵, stress³⁶, circadian rhythm³⁷, and biological sex³⁸. There are also factors that are related to glycemic health and glucose fluctuations, although they may not affect blood glucose directly, including heart rate³⁹, body temperature⁴⁰, and autonomic functions⁴¹ including the sudomotor response⁴². The factors affecting and relating to blood glucose are extremely personalized: individualized factors can have significant impacts on the blood glucose dynamic²⁵,

¹Department of Biomedical Engineering, Duke University, Durham, NC, USA. ²Department of Statistical Science, Duke University, Durham, NC, USA. ³Division of Endocrinology, Department of Medicine, Duke University Medical Center, Durham, NC, USA. ⁴Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC, USA. ✉email: jessilyn.dunn@duke.edu

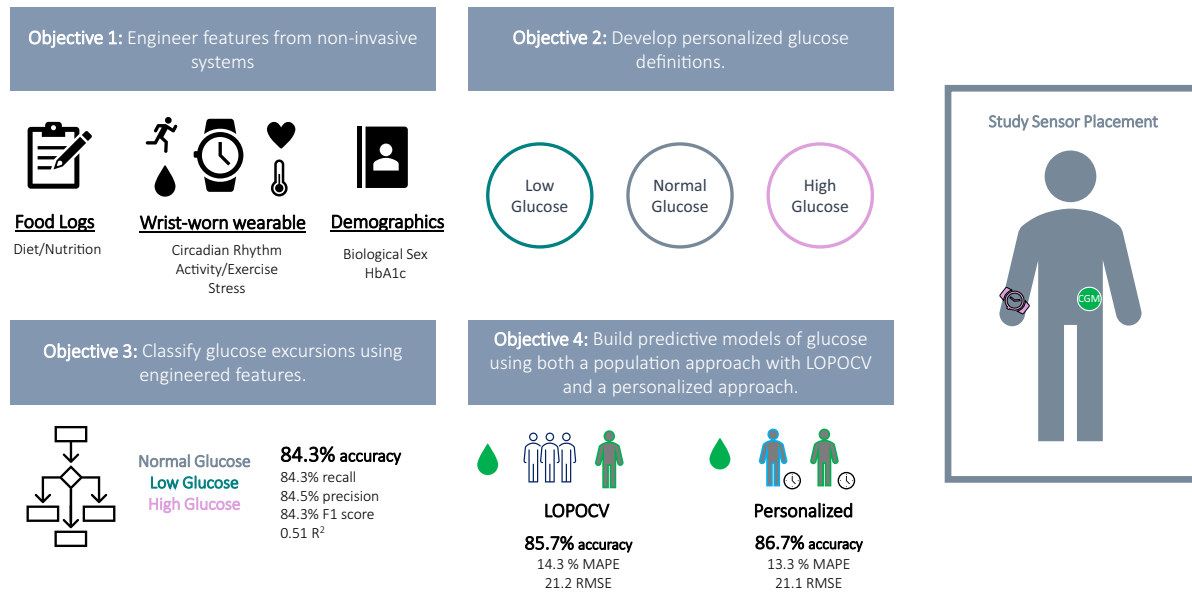


Fig. 1 Graphical abstract of study. This study has four objectives: (1) Engineer features from non-invasive systems using a combined data-driven and domain-driven feature engineering approach. (2) Develop personalized glucose excursions definitions. (3) Classify glucose excursions using engineered features. (4) Build predictive models of glucose using both a population approach with leave-one-person-out cross validation and a personalized approach. Sensor placement for the study (Empatica E4 on the wrist and a Dexcom G6 continuous glucose monitor on the abdomen) is also shown.

glycemic responses to foods are highly individual, and the foods that raise a person's blood glucose can vary dramatically⁴³.

The primary objective of this study is to determine whether we can build models from noninvasive wearables data combined with food logs to classify interstitial glucose levels and to predict interstitial glucose (Fig. 1). This would enable a frictionless and noninvasive method to monitor interstitial glucose in real time, allowing patients to engage with their glycemic health and actively monitor their progress while employing lifestyle modifications.

RESULTS

Developing personalized definitions of interstitial glucose excursions for detecting intraindividual excursions

Study participants were 35–65 years of age with elevated blood glucose in the normal range (HbA1c 5.2–5.6) or prediabetes (HbA1c 5.7–6.4) (Supplementary Table 2). Existing clinical definitions of blood glucose excursions, including 'hyperglycemia' (glucose that is too high) and 'hypoglycemia' (glucose that is too low), are defined at the population level and are largely used to describe glucose excursions in T1D and T2D patients^{5,13–19}; however, these classifications were established for diabetes management¹⁹ and may not be suitable to explain significant glucose excursions in normoglycemic or prediabetic individuals due to their lower overall fasting glucose levels and lower glucose variability as compared with diabetic individuals. Examples demonstrating this scenario are shown in Fig. 2, where three separate participants had clear glycemic responses to specific behaviors, including sugary food or drink intake (spike in blood sugar) or not eating (drop in blood sugar). These excursions were not sufficiently high or low as to be clinically categorized as hyper- or hypoglycemia and would therefore be considered normal as per the traditional definitions, even though they are high or low compared to the individual's own baseline. Personalized glucose excursion classes (Fig. 2 circles) provide more tailored information about glucose fluctuations and enable self-management and tracking of diet, exercise, and stress-related behaviors that affect blood glucose. Furthermore, the glycemic baseline is also dynamic

over time^{44–46}, so single-valued population-level thresholds to define glucose excursions are inadequate for understanding an individual's deviations over time from their typical state^{25,26,43}. Thus, we developed three personalized and dynamic designations to categorize each interstitial glucose measurement to indicate the presence and absence of personalized glucose excursions. We denote these categories as PersHigh, PersLow, and PersNorm, which correspond to an interstitial glucose measurement that is greater than, less than, or within one standard deviation of the 24-h personalized mean, respectively (Fig. 3a). These personalized and time-varying calculations account for circadian, intra-, and inter-day variability. Each of the three categories are approximately normally distributed (PersNorm Kolmogorov–Smirnov Normality Test (KS) Statistic=0.03 (Fig. 3c); PersHigh KS Statistic=0.05; PersLow KS Statistic=0.04) (Table 1). The PersHigh distribution is skewed moderately right and is leptokurtic, with more data located at the tails of the distribution rather than around the mean (Fig. 3d, Table 1). PersLow is skewed slightly left and is mesokurtic, with a distribution moderate in breadth and approximately normally distributed (Fig. 3e, Table 1). The wider distribution for PersHigh compared to PersLow likely reflects the fact that there is a wider range of possible hyperglycemic values than hypoglycemic values in our population. Interestingly, there is an overlap in all three distributions between interstitial glucose values of 66–164 mg/dL (Fig. 2b), supporting the idea that what may be considered a normal measurement for one person may actually be a low or high measurement for another person, which also points to the inadequacy of population-level thresholds.

Glucose excursion classification

For many patients with conditions like prediabetes, simply understanding which behaviors trigger high or low glucose excursions would greatly improve disease management. A number of measurable factors influence blood glucose levels, including diet^{34,47–57}, physical activity and exercise^{35,58–62}, stress^{36,63–69}, circadian rhythm^{37,55,70–73}, and biological sex^{38,74–80}. Additionally, physiological parameters like vital signs are associated with glycemic health and glucose fluctuations, including heart rate^{39,81,82}, core body temperature^{40,83}, and autonomic functions^{41,84–86} like the sudomotor

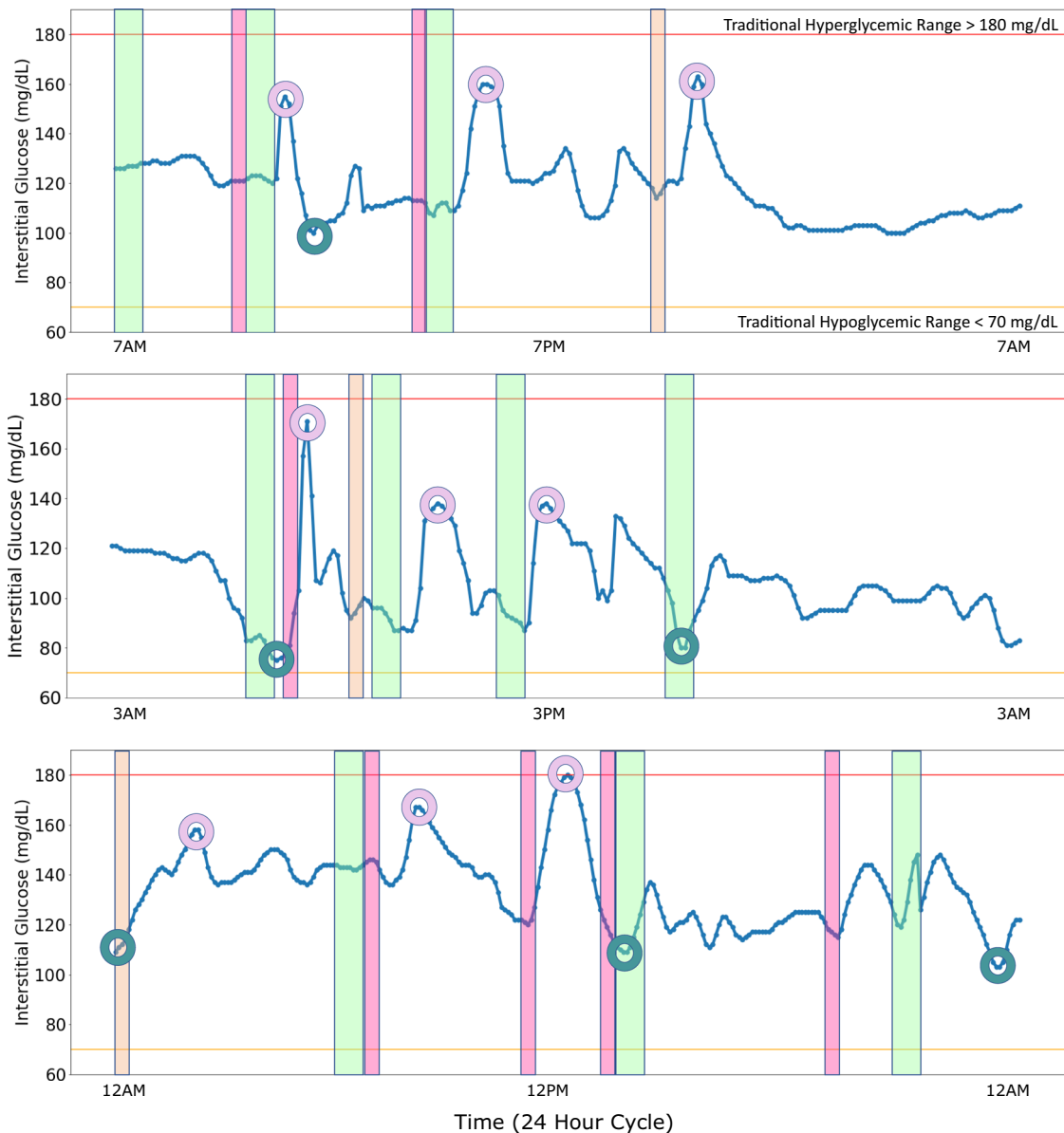


Fig. 2 Case studies of CGM data over 24 h for 3 participants highlighting eating habits and subsequent glucose excursions. Green bands indicate a meal or snack, pink bands indicate consumption of a sugary beverage or soda, and orange bands indicate a sugary snack or dessert. Shown in red is the traditional hyperglycemic range (>180 mg/dL) and shown in yellow is the traditional hypoglycemic range (<70 mg/dL). Note that the participants do not exceed the traditional definitions of glucose excursions. A system that would alert prediabetes patients based on traditional definitions of glucose excursions would be inadequate in these cases. The personalized classifications we propose in this paper would better inform prediabetes patients so they can begin to self-manage their diet, exercise, and stress levels based upon this information. As shown, the personalized classes PersHigh (shown in plum) and PersLow (shown in teal) would provide more personalized information about fluctuating glucose. For example, sugary drinks and snacks result in higher glucose fluctuations. By reducing the amount and frequency of sugary drinks and snacks, participants would be less likely to experience a glucose excursion.

response⁴². These relationships suggest that it may be possible to estimate glucose values from novel modes of measurement. To determine whether this is possible, we developed classification models using data from these alternative modes of measurement with the goal of detecting when interstitial glucose was outside of the personal norm.

We engineered 69 variables based on the previously described literature^{34–42,47–86} as inputs into our prediction models using a combination of data-driven and domain-driven feature engineering (Supplementary Table 1). These variables were built using data collected from a noninvasive wearable, a food log, and electronic reports of demographics. Measurements included metrics of

stress, circadian rhythm, diet, activity and exercise, heart rate, skin temperature, and biological sex.

We defined the ground truth for intraindividual excursions, classified as PersHigh, PersLow, or PersNorm, based on a personalized, rolling basis from CGM measurements. We developed a multi-class model to classify between PersLow, PersHigh, and PersNorm interstitial glucose using a class-balanced dataset ($N = 8666$). The decision tree classifier with repeated stratified k-fold cross validation achieved an accuracy of $84.3 \pm 0.013\%$ (recall = $84.3 \pm 0.013\%$; precision = $84.5 \pm 0.013\%$; weighted F1 Score = $84.3 \pm 0.013\%$; $R^2 = 0.505 \pm 0.050$) (Table 2). We repeated this modeling task using a 70/30 train/test (TT) split. The decision

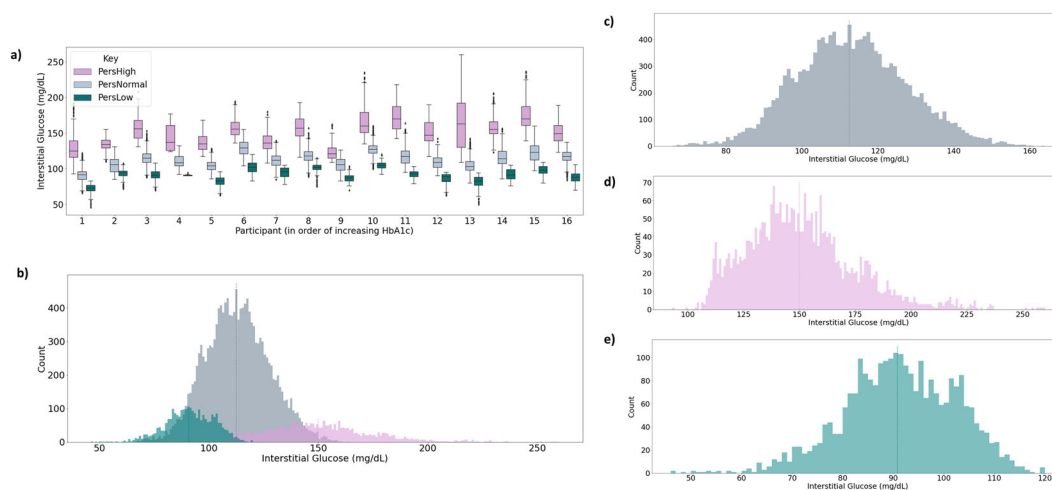


Fig. 3 Personalized glucose excursions. Glucose excursions are classified on a personalized, rolling basis examining the previous 24 h of each participant's historical data. **a** Boxplots of PersNorm, PersHigh, and PersLow for each participant in the dataset. **b** Histogram of all distributions over all participants. **c** PersNorm distribution. **d** PersHigh distribution. **e** PersLow distribution. Bin width for each histogram shown is 1 mg/dL.

tree classifier using the 70/30 TT split achieved 82.0% accuracy (recall = 82.0%; precision = 82.3%; F1 Score = 82.1%; $R^2 = 0.46$) (Table 2). The confusion matrix for the decision tree classifier using the 70/30 TT split is shown in Fig. 4. The per-class accuracies for each of the three interstitial glucose classes were similar: the class accuracy for PersHigh glucose was 82.6%, the class accuracy of PersNorm was 81.3%, and the class accuracy for PersLow was 82.1%. The decision tree models both outperformed logistic regression (accuracy = 52.0%; recall = 52.0%; precision = 52.3%; F1 Score = 52.0%; $R^2 = 0$ (Method of calculation for R^2 enabled negative values, which were thresholded at zero), indicating that more complex relationships in the data need to be captured to perform highly accurate classification (Table 2).

Glucose prediction

Prediction of precise interstitial glucose values as opposed to whether or not a person is experiencing a high or low excursion would give additional information for glucose self-monitoring and tracking. Thus, we extended our models to determine whether noninvasive wearables could serve as a proxy for a continuous glucose monitor. In order to predict glucose at 5-min intervals, we developed both a gradient-boosted population model validated with leave-one-person-out cross validation (LOPOCV) and a gradient-boosted personalized model, trained and tested on each individual's previously measured data. Our population model validated with LOPOCV had an average root mean squared error (RMSE) of 21.22 ± 4.14 mg/dL and an average mean average percent error (MAPE) of $14.33 \pm 3.25\%$. Average accuracy for the population model over all participants was 85.67%. Using the initial half of the participant's data for training, the personalized model trained and tested on each participant's data had an average RMSE of 21.10 ± 4.50 mg/dL and an average MAPE of $13.26 \pm 3.94\%$. Average accuracy for the personalized model across all participants was 86.74%. Both the personalized model and the traditional population model outperformed the naïve models (mean and median). That the population and personalized models performed comparably may indicate that there are common factors that are mostly sufficient for modeling.

Exploring feature importance for glucose prediction

In order to determine the contribution of each of the 69 variables to the interstitial glucose prediction, we calculated impurity-based importance for each variable in the LOPOCV random forest regression models with importances averaged across each fold.

The feature importances allow us to examine the extent that each measurement type contributed to the success of the model and therefore serve as a mechanism to generate hypotheses of potential physiologic relationships that can be tested directly through experimentation. Features were aggregated together into the following categories: 'food', 'circadian rhythm', 'stress', 'activity', 'temperature', 'heart rate', 'electrodermal activity', 'biological sex', 'HbA1c', and 'personalization'. They were further categorized by the source of the data: 'food log', 'wearable', 'user input', and 'model'. Based on the literature we anticipated that the most important features would be related to food, activity, circadian rhythm, and stress^{34–37,47–73} and found that indeed food had the highest importance, with an average of 37.0% (percent of total importance, where total importance sums to 100%), followed closely by activity (17.0%), circadian rhythm (10.6%), and stress (8.2%). Of the feature importances, 49.3% were derived from a wearable, 37.0% were sourced from the food log, 10.8% were user input (including biological sex and HbA1c), and 2.9% were personalization features of the model. This supports further development of multi-modal models using features from both a noninvasive wearable and a food diary to predict glucose. In terms of how the features were engineered, 66.8% of feature importances were domain-driven, 19.5% were data-driven, and 13.7% were neither (demographics data) (Fig. 5).

The fifteen most important features in the model (mean impurity-based feature importance >0.02), ordered from most to least important, included measures of circadian rhythm, diet, demographics, exercise/activity, and stress (Table 3, Supplementary Fig. 1). The variance among the LOPOCV models is illustrated in Supplementary Fig. 1 and highlights the variability in the most important features.

Overall, novel findings from this work included the development of personalized definitions of interstitial glucose excursions, demonstrating robust classification of these personalized interstitial glucose excursions from non-invasive wearables and food logs data and creating a noninvasive continuous glucose monitor "proxy" with high accuracy. We developed novel feature engineering methods that achieve high accuracy in interstitial glucose classification and regression models. These features highlight the existence of important relationships between physiologic measurements from non-invasive wearables and interstitial glucose fluctuations which serve as a basis for future experimental studies.

Table 1. Summary metrics of distributions of each personalized glucose definition.

Glucose class	Mean (mg/dL)	Standard deviation (mg/dL)	Range (mg/dL)	Skewness	Kurtosis
PersNorm	112.4	14.5	66.0 - 164.0	0.12	-0.10
PersHigh	149.9	24.4	93.0 - 260.0	0.75	0.93
PersLow	90.8	11.1	46.0 - 120.0	-0.32	0.17

The categories PersHigh, PersLow, and PersNorm are personalized glucose definitions that correspond to an interstitial glucose measurement that is greater than, less than, or within one standard deviation of the 24-h personalized mean, respectively.

Table 2. Model evaluation metrics for multi-class models.

Model	Balanced accuracy	Recall	Precision	F1 Score	R ²
Multiclass Decision Tree Repeated Stratified K-Fold CV	84.3 ± 0.013%	84.3 ± 0.013%	84.5 ± 0.013%	84.3 ± 0.013%	0.505 ± 0.050
Multiclass Decision Tree 70/30 Train/Test Split	82.0%	82.0%	82.3%	82.1%	0.455
Multiclass Logistic Regression 70/30 Train/Test Split	52.0%	52.0%	52.3%	52.0%	0 ^a

^aMethod of calculation for R² enabled negative values, so the threshold was set to 0.

DISCUSSION

The primary objective of this study was to determine whether we can use noninvasive wearables with food diaries to detect personalized interstitial glucose excursions and predict interstitial glucose values. We integrated data-driven and domain-driven feature engineering methods to engineer 69 features, and we developed a personalized, rolling baseline to determine glucose excursions. We used our engineered features in machine learning models to classify glucose excursions. Finally, we used our engineered features to develop and compare population-level and personalized models for glucose prediction.

Typically, patients with diabetes or other glucose metabolism conditions monitor their glucose by measuring their blood sugar level periodically throughout the day with a blood glucose meters or continuous glucose monitors, both of which involve an invasive needle and are not typically recommended for individuals with prediabetes. Innovative, practical strategies to improve monitoring and management of glycemic health are desperately needed. One way we can improve management of glycemic health is by using noninvasive wearables and food logging to determine when a participant is experiencing interstitial glucose excursions (high glucose or low glucose).

There is a lack of prediabetes-specific thresholds for glucose deviations, which makes prediabetes monitoring and disease management challenging. Furthermore, the spectrum of prediabetes is highly personalized^{26,43} and personalized thresholds for glucose excursions do not currently exist, preventing personalized monitoring. To address this gap, we developed personalized definitions of interstitial glucose excursions. Defining glucose excursions on a personalized, rolling basis enables the use of a participant's historical data to better understand deviations from a personalized baseline over a defined window of time. This also allows accounting for intraday variability in addition to interday variability.

Here, we developed a multi-class model to detect PersHigh, PersLow, and PersNorm interstitial glucose events that achieved 84.3% balanced accuracy. Currently, these classification models could be employed in practice to alert patients with prediabetes when they may be experiencing a glucose excursion to check their blood glucose. Overall, optimizing the timing of blood glucose checks will better inform people about the effect of their behaviors and lifestyle habits on their blood glucose. For example,

a patient's own historical data can show them how reducing the amount and frequency of sugary drinks and snacks would reduce the likelihood of a glucose excursion. In the future, further validation of these models may enable their use alone without additional tools to self-monitor and track blood glucose.

Displaying exact glucose values to users has been shown to improve logical reasoning and influence their eating and activity behaviors⁸⁷, which could be directed toward improved glucose control. Having specific, numeric, goals have been demonstrated to increase and maintain motivation⁸⁸. Here, we demonstrate the feasibility of predicting glucose from non-invasive data at 5-min intervals (the same sampling rate as a continuous glucose monitor) which could further enable patients to understand how their lifestyle habits are influencing their blood glucose levels and help them manage their disease. Interestingly, while continuous glucose monitors are widely used to inform insulin delivery, their accuracy can vary substantially from a blood glucose meter. For example, the Dexcom G6 is considered sufficiently accurate if its measurements fall within ±20 mg/dL of a simultaneous blood glucose meter measurement for meter values <100 mg/dL and ±20% for meter values >100 mg/dL⁸⁹. As such, the accuracy of our newly proposed noninvasive glucose prediction models is within the accuracy realm of currently used technologies (RMSE = 21.22 ± 4.14 mg/dL).

The random forest-based glucose prediction models showed that, of the top features, 49.3% were derived from a wearable, 37.0% were sourced from the food log, 10.8% were from user inputs (including biological sex and HbA1c), and 2.9% were resulting from the model itself (personalization). The benefits of multiple measurement modalities can inform future decisions in digital biomarker research including data collection and wearable and mobile health device and application design. Of the 15 most important features, the overarching categories of circadian rhythm, demographics, diet, exercise, and stress were all represented, lending further credence to the usefulness of domain-driven feature engineering. There is ongoing debate about using HRV as a proxy measure of stress⁹⁰ which may impact interpretation of the key model features. Additionally, the glucose management indicator (GMI) has been demonstrated to approximate HbA1c using continuous glucose monitoring data⁹¹. In future models, GMI may be able to replace the clinical HbA1c measurement, avoiding the need for this clinical measurement.

		True		
		PersNorm	PersLow	PersHigh
Predicted	PersNorm	689	102	57
	PersLow	103	672	44
	PersHigh	110	52	771

Fig. 4 Confusion matrix: multiclass model. Confusion Matrix for multiclass decision tree model validated with a 70/30 train/test split. In this model, the class-accuracy for PersHigh glucose is 82.6%, the class accuracy for PersNorm is 81.3%, and the class accuracy for PersLow is 82.1%.

Noninvasive glucose monitoring is a complex challenge, and our findings highlight the necessity of having multi-disciplinary teams in digital biomarker discovery research to integrate domain knowledge for domain-specific feature engineering.

In future work, we recommend expanding upon this study to improve glucose prediction using a personalized modeling approach with larger datasets and other machine learning methods. With a larger dataset, the integration of deep learning with the personalized approach may further increase the efficacy of non-invasively predicting glucose using wearables. In larger cohort studies, we recommend exploring age and body mass index and/or body fat percentage as covariates in the models. We also recommend the evaluation of this technology following the V3 (verification, analytical validation, and clinical validation) framework²⁹. Here, we demonstrate the feasibility of using noninvasive methods for interstitial glucose classification and prediction. Follow up studies are needed to clinically evaluate this technology.

Glycemic health is at an all-time low: in the U.S., one in ten people have diabetes and one in three people have prediabetes; a dismal 20% of those with diabetes and 90% of those with prediabetes are undiagnosed⁹². In order to manage glucose fluctuations, it is important for patients to understand how their behaviors influence their blood glucose levels. There is a critical need for innovative, practical strategies to improve monitoring and management of glycemic health. In this study, we demonstrated the feasibility of using noninvasive and widely accessible mobile health and machine learning methods to non-invasively classify glucose excursions and predict glucose values.

METHODS

Dataset recruitment and collection protocol

The study was approved by the Duke University Health System (DUHS) Institutional Review Board and written informed consent was obtained from all participants (Pro00101398). All subjects consented to the study and were compensated a total of \$150 for their participation.

Patients ($N = 16$) were recruited for this prospective study from the Duke Endocrinology and Lipids Clinic through medical record review that identified patients between 35–65 years of age with high normal blood glucose (HbA1c 5.2–5.6) or prediabetes (HbA1c 5.7–6.4) (Supplementary Table 2). Exclusion criteria included cancer, COPD, cardiovascular disease, food allergies, or any antidiabetic drug use.

HbA1c was measured in the clinic on Day 0. Upon confirmation of A1C within range for this study, a continuous glucose monitor (CGM; Dexcom G6) and a non-invasive wearable smartwatch (Empatica E4) were worn continuously for 8–10 days (placement of study sensors shown in Fig. 1). A standardized breakfast meal with a high glycemic index (1.5 cups of

Frosted Flakes and 1 cup Lactaid 2% Milk) was ingested every other morning during the monitoring period prior to ingesting any other food, drink, or medication, enabling repeated intraindividual monitoring of glycemic response. All other meals and snacks during the monitoring period were recorded through comprehensive written diet logging.

Dataset

The Dexcom G6 records interstitial glucose concentration (mg/dL) every 5 min. The Empatica E4 contains four sensors: photoplethysmography (optical heart rate), electrodermal activity (galvanic skin response, related to sweat activity), skin temperature, and tri-axial accelerometry. Heart rate was recorded once per second, (calculated from photoplethysmography sampled at 64 Hz), electrodermal activity and skin temperature were recorded at 4 Hz, and accelerometry was recorded at 32 Hz.

In total, for this analysis we utilized over 25,000 interstitial glucose point measurements and 25,000 5-min epochs of wearable data measured over 8–10 days across 16 participants.

Feature engineering

Features were engineered on data collected from the Empatica E4 wrist-worn wearable and the food diary. In total, 69 data-driven, domain-driven, and demographic historical features were used in developing the models, including features linked to diet, stress, exercise, circadian rhythm, and behavioral habits: all features known to contribute to blood glucose fluctuations (Supplementary Table 1). These features were computed every 5 min on a rolling basis (when appropriate, as described below). All features used in modeling were historical (5 min to 24 h prior to the measurement being predicted). Thus, the models were only given historical data to make predictions from.

Two of the 69 features were demographic data that incorporate user input into the model: biological sex and HbA1c. Another feature, the ‘personalization’ feature, differentiates each participant with a unique number so that models can learn relationships that may be individualistic.

Data-driven features for each of the 5-min intervals of smart watch data include 7 summary statistics for each sensor: mean, standard deviation, minimum, maximum, first quartile, third quartile, and skew. These summary statistics are computed for heart rate, accelerometry (vector magnitude of the three axes), electrodermal activity, and skin temperature.

For the domain-driven feature engineering, we focused on 4 factors that have demonstrated effects on blood glucose: stress, exercise (short term and long-term effects), circadian rhythm, and diet (including timing of meals, frequency of meals, and the short, medium, and long-term effects of protein, sugar, carbohydrates, and calories). We will break each of these down and explain the metrics calculated below:

The effects of stress have been measured physiologically for decades in psychology research, and we borrow some of their methods to quantify stress here^{93–97}. Using the electrodermal activity data, we detected peaks and determined their prominences using the SciPy library in Python⁹⁸. We required a distance between peaks of 1 second (4 data points) and a prominence of 0.3 micro-siemens to be considered a ‘unique peak’. We determined the number of peaks in each 5-min interval. This data was then

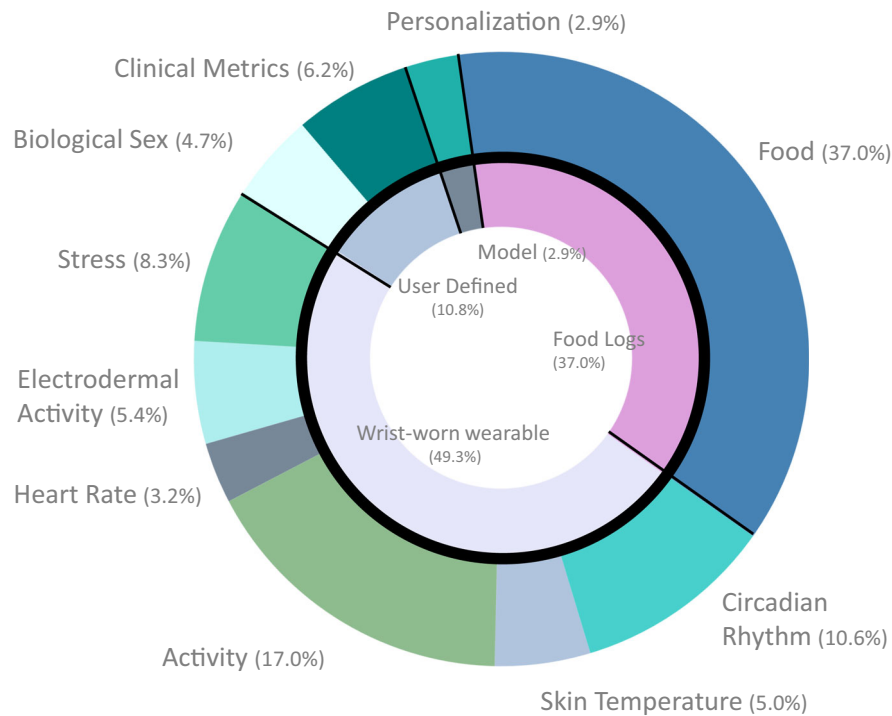


Fig. 5 Importance of 69 domain-driven and data-driven features in glucose predictions. Importance was determined from a random forest feature selection model using impurity-based features. The outer circle shows relative percent of importance for the categories food, circadian rhythm, activity, stress, gender, clinical metrics, personalization, electrodermal activity, heart rate, and skin temperature. The inner circle shows relative importance by source of features, including food logs, the wrist-worn wearable, user-defined features, and features defined for/by the model.

aggregated using a rolling window approach: for the previous 2 h, on a rolling window, we determined the total number of peaks and the average number of peaks in each 5-min interval.

Heart rate variability (HRV) has also been demonstrated to fluctuate relating to both acute and chronic stress⁹⁰. We utilized the inter-beat-interval data derived from the PPG and calculated 8 HRV metrics over each 5-min interval of data. Calculated metrics include mean HRV, median HRV, maximum HRV, minimum HRV, the standard deviation of intervals (SDNN), the root mean square of successive differences in the intervals (RMSSD), the number of successive intervals that differ by more than 50 ms (NN50), and the proportion of NN50 divided by the total number of intervals (pNN50). To calculate these metrics, we utilized open source code available in the Digital Biomarker Discovery Pipeline³² that has been validated against the state-of-the-art Kubios ECG software.

There have been shown to be both short-term and long-term effects of exercise on blood glucose levels^{58,59}, which we wanted to examine using both accelerometry and heart rate data. While it is difficult to quantify “exercise” without user-input, we can determine when participants are being more active. Thus, we developed a calculation for “activity bouts”, or bouts of activity.

To calculate an “activity bout”, we took the mean accelerometry vector magnitude and the mean heart rate over the 5-min interval and compared it to the average of the prior historical data from the individual. If both the mean accelerometry and mean heart rate values for the 5-min interval were above the previous average, that interval is said to be an “activity bout”. In order to account for the short- and longer- term effects of activity on the glucose metabolism, we used a rolling window to compute the total activity bouts in the last hour and the average activity bouts in the previous 24 h. Additionally, we found the mean and maximum accelerometry vector magnitude over the previous two hours using a rolling window approach.

Circadian rhythm is a confounding variable in most physiology and it is important to take this into account as a feature, especially because there is a known connection between circadian rhythm and blood glucose³⁷. We calculated the ‘minutes from midnight’ and the ‘hours from midnight’ as features indicative of circadian rhythm.

The time when an individual wakes can affect their circadian rhythm⁹⁹. Heart rate and accelerometry have been used previously to determine waking versus sleep states³². For each participant and for each day, we examined when the accelerometry and heart rate mean and standard deviation were less than the average for that day. When two of the four measures were less than the average for the day, we assigned that interval a ‘0’. All other intervals were assigned a ‘1’. This data was then averaged over 3 h using a rolling window approach. We assigned the ‘Wake Time’ to when the slope of this data sharply changed and remained consistently higher 25 and 75 min after the wake time.

Diet is known to have strong effects on glucose metabolism. Here, we determined features explaining the short-, medium-, and long-term effects of food on glucose. We used a rolling window to sum the number of calories, grams of protein, grams of carbohydrates, and grams of sugar over three windows: 2 h, 8 h, and 24 h.

Each time a participant consumed a unique meal, snack, or caloric beverage, we assigned that interval a binary ‘1’. This allows us to determine timing of meals and how that, in conjunction with the circadian rhythm features, may affect the glucose metabolism. We then took a rolling sum of how many times an individual consumed over 2 h, 8 h, and 24 h. We also looked at the average amount of times an individual consumed over the 2-h, 8-h, and 24-h rolling windows.

Developing personalized definitions of interstitial glucose excursions for detecting intraindividual excursions

The traditional definitions of glucose values that are ‘too high/low’, or hyperglycemia and hypoglycemia, have been defined from populations of people with diabetes^{13,19–24}. These previous thresholds are insufficient and require personalization²⁶. Thus, we wanted to re-define this and create personalized “high” and “low” glucose thresholds for each participant at each point in time. All outliers in the dataset were examined in exploratory data analysis relative to the food logs and other features, in addition to surrounding glucose values to ensure they were within reasonable bounds. We found no outliers in our dataset that were not explained by other features. We defined personalized glucose excursions as “PersHigh” and “PersLow”. We used a rolling window approach: for a glucose value to be

Table 3. Exploring feature importance for glucose prediction: the 15 most important features in the random forest model.

Feature	Feature category	Impurity-based feature importance (mean \pm standard deviation)
Minutes from midnight	Circadian rhythm	0.069 \pm 0.004
Sugar intake (previous 24 h)	Food	0.067 \pm 0.038
Hemoglobin A1c (HbA1c)	Clinical/Demographics	0.062 \pm 0.018
Carbohydrates (previous 2 h)	Food	0.047 \pm 0.016
Biological sex	Demographics	0.047 \pm 0.033
Activity (previous 24 h)	Activity/Exercise	0.042 \pm 0.005
Sugar intake (previous 2 h)	Food	0.038 \pm 0.022
Sugar intake (previous 8 h)	Food	0.031 \pm 0.004
Accelerometry (mean previous 2 h)	Activity/Exercise	0.030 \pm 0.002
Wake time	Circadian rhythm	0.029 \pm 0.005
ID	Personalization	0.029 \pm 0.016
Accelerometry (max previous 2 h)	Activity/Exercise	0.026 \pm 0.002
Protein (previous 24 h)	Food	0.025 \pm 0.002
Peak electrodermal activity (mean previous 2 h)	Stress	0.024 \pm 0.009
Carbohydrates (previous 24 h)	Food	0.024 \pm 0.003

classified as PersHigh, it had to exceed one standard deviation above the mean for the last 24 h. To be considered PersLow, the value had to be below one standard deviation below the mean for the last 24 h. The third category, "PersNorm" had to fall within one standard deviation above or below the mean for the last 24 h. We tested the extent to which the distributions of these categories are normal using the Kolmogorov-Smirnov Test to test between our distribution and a normal distribution.

Classification of glucose excursions

We balanced our classes of PersNorm, PersHigh, and PersLow for a total $N = 8666$ because the entire dataset ($N = \sim 25,000$) was highly imbalanced with the majority of data points being in the PersNorm category. Our model was implemented in a repeated stratified k-fold cross validation schema with 10 splits and 3 repeats. Within each fold, we implemented recursive feature selection to select the 20 most important features, which were used to train the model. We iterated through several estimator methods to determine the optimal method for the recursive feature elimination, including logistic regression, perceptron, decision tree, random forest, and gradient boosting classifier. We utilized the estimator that resulted in the highest accuracy for our final model, the decision tree estimator. Finally, for each fold, using the features selected with the recursive feature elimination, we trained a decision tree classifier.

In addition to the primary, cross-validated model, we developed a 70/30 train/test split decision tree classifier model and a logistic regression model.

We evaluated our model using balanced accuracy, weighted precision, weighted recall, and weighted f1 score, from the python package scikit-learn¹⁰⁰ at each fold in the cross validation and report the mean and standard deviation of each metric. We also examined the amount of variance explained by the model by reporting R^2 .

Glucose prediction

Using the same features derived above in our feature engineering, we developed predictive models for predicting actual interstitial glucose values from noninvasive wearables and food diary data. We developed both a population model with LOPOCV and a personalized model trained and tested on each participant's own data. We utilized a dataset of over 25,000 glucose measurements and 25,000 5-min intervals of smart watch data measured over 8–10 days across 16 participants. There were an average of 1500 glucose measurements per participant.

We developed a regression gradient boosting decision-tree-based regression model using the XGBoost¹⁰¹ algorithm and used tuned hyperparameters for all models (maximum depth=6, number of estimators=100, learning rate = 0.1). The prediction target of our models was

interstitial glucose at every 5-min interval. The model was trained using the 69 engineered features and feature selection (cutoff = 0.005) was performed with a random forest regression model (1000 trees) using impurity-based feature importances for each fold of our LOPOCV. Feature importance was taken at each fold in our LOPOCV model and averaged to determine the most important features in predicting glucose.

For the population, LOPOCV model, we iterated over each participant (fold), using all other participants as the 'training set' and using each participant as the 'test set'. For the personalized model, we trained on the first contiguous half (50%) of the participant's data and tested on the remaining half of the participant's data.

Models were evaluated using root mean squared error (RMSE), mean average percent error (MAPE), and accuracy (100-MAPE) for each fold of our LOPOCV models. We report the mean and standard deviation over the population of these evaluation metrics.

Analysis of features

In order to determine important features for predicting glucose, we trained a random forest regression model (1000 trees) with LOPOCV and averaged impurity-based feature importances for each fold of our LOPOCV to determine the most important features across participants.

Features were aggregated together into the following categories: 'food', 'circadian rhythm', 'stress', 'activity', 'temperature', 'heart rate', 'electrodermal activity', 'biological sex', 'HbA1c', and 'personalization'. They were further categorized by the source of the data: 'food log', 'wearable', 'user input', and 'model'. We also divided the features into categories based on how the feature engineering was performed: 'data-driven', 'domain-driven', and 'other'. Feature importances across these categories were aggregated by averaging the importance across each fold in our LOPOCV random forest regression model. We determined the percent importance of each of these feature categories out of the total feature importance; thus, we report importance as a percent when evaluating the features.

Reporting summary

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

DATA AVAILABILITY

The data sets generated during and/or analyzed during the current study will be submitted one year from the publication date to the Digital Health Data Repository in the Digital Biomarker Discovery Pipeline.

CODE AVAILABILITY

The code generated for this study are available open-access in the Digital Biomarker Discovery Pipeline (dbdp.org).

Received: 12 March 2021; Accepted: 11 May 2021;

Published online: 02 June 2021

REFERENCES

- Prediabetes: You Could Be That 1 in 3 (accessed 2 March 2020); <https://www.cdc.gov/diabetes/library/features/prediabetes-1-in-3.html>.
- New CDC report: more than 100 million Americans have diabetes or prediabetes (CDC Online Newsroom, CDC, accessed 16 February 2020); <https://www.cdc.gov/media/releases/2017/p0718-diabetes-report.html>.
- Aldossari, K. K. et al. Prevalence of prediabetes, diabetes, and its associated risk factors among males in Saudi Arabia: a population-based survey. *J. Diabetes Res.* **2018**, 2194604 (2018).
- Mellbin, L. G., Anselmino, M. & Rydén, L. Diabetes, prediabetes and cardiovascular risk. *Eur. J. Cardiovasc. Prev. Rehabil.* **17** (Suppl 1), S9–14 (2010).
- Bansal, N. Prediabetes diagnosis and treatment: a review. *World J. Diabetes* **6**, 296 (2015).
- Tabák, A. G., Herder, C., Rathmann, W., Brunner, E. J. & Kivimäki, M. Prediabetes: a high-risk state for diabetes development. *Lancet* **379**, 2279–2290 (2012).
- Group, T. D. P. P. (DPP) R. The Diabetes Prevention Program (DPP): description of lifestyle intervention. *Diabetes Care* **25**, 2165–2171 (2002).
- Personalised Online Diabetes Education Program To Get You On Track (accessed 12 February 2021); <https://www.myhealthexplained.com/>.
- Prediabetes and Monitoring (accessed 12 February 2021); <https://www.myhealthexplained.com/diabetes-information/diabetes-articles/prediabetes-monitoring>.
- January.ai | Optimize your blood sugar intelligently (accessed 12 February 2021); <https://january.ai/landing/>.
- Levels - Metabolic Fitness Program (accessed 12 February 2021); <https://www.levelshealth.com/>.
- Russell Koenigsberg, M. & Corliss, J. Diabetes self-management: facilitating lifestyle change. *Am. Fam. Physician* **96**, 362–370 (2017).
- Umpierrez, G. E. & P. Kovatchev, B. Glycemic variability: how to measure and its clinical implication for type 2 diabetes. *Am. J. Med. Sci.* **356**, 518–527 (2018).
- Tamborlane, W. V. et al. Continuous glucose monitoring and intensive treatment of type 1 diabetes. *N. Engl. J. Med.* **359**, 1464–1476 (2008).
- Aronoff, S. L., Berkowitz, K., Shreiner, B. & Want, L. Glucose metabolism and regulation: beyond insulin and glucagon. *Diabetes Spectr.* **17**, 183–190 (2004).
- Rodbard, D. Glucose variability: a review of clinical applications and research developments. *Diabetes Technol. Ther.* **20**, S2–S15 (2018).
- Kovatchev, B. Glycemic variability: risk factors, assessment, and control. *J. Diabetes Sci. Technol.* **13**, 627–635 (2019).
- Jean-Marie, E. Diagnosis and classification of diabetes mellitus. in *Encyclopedia of Endocrine Diseases* **43**, 105–109 (Elsevier, 2018).
- Suri, P. & Aurora, T. Hyperglycemia. in *Observation Medicine: Principles and Protocols* 225–230 (Cambridge University Press, 2017). <https://doi.org/10.1017/9781139136365.043>
- Chiu, C.-J. & Taylor, A. Dietary hyperglycemia, glycemic index and metabolic retinal diseases. *Prog. Retin. Eye Res.* **30**, 18–53 (2011).
- Porcellati, F., Lucidi, P., Bolli, G. B. & Fanelli, C. G. Thirty years of research on the dawn phenomenon: lessons to optimize blood glucose control in diabetes. *Diabetes Care* **36**, 3860–3862 (2013).
- Monnier, L., Lapinski, H. & Colette, C. Contributions of fasting and postprandial plasma glucose increments to the overall diurnal hyperglycemia of type 2 diabetic patients: Variations with increasing levels of HbA1c. *Diabetes Care* **26**, 881–885 (2003).
- Suh, S. & Kim, J. H. Glycemic variability: how do we measure it and why is it important? *Diabetes Metab. J.* **39**, 273–282 (2015).
- Johnson, E. L. et al. Standards of medical care in diabetes—2019 abridged for primary care providers. *Clin. Diabetes* **37**, 11–34 (2019).
- Woldaregay, A. Z. et al. Data-driven blood glucose pattern classification and anomalies detection: machine-learning applications in type 1 diabetes. *J. Med. Internet Res.* **21**, e11030 (2019).
- Hall, H. et al. Glucotypes reveal new patterns of glucose dysregulation. *PLoS Biol.* **16**, e2005143 (2018).
- Gonzales, W. V., Mobashsher, A. T. & Abbosh, A. The progress of glucose monitoring—a review of invasive to minimally and non-invasive techniques, devices and sensors. *Sensors* **19**, 800 (2019).
- Heinemann, L. Finger pricking and pain: a never ending story. *J. Diabetes Sci. Technol.* **2**, 919–921 (2008).
- Dunn, J. C. et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). *npj Digit. Med.* **3**, 1–15 (2020).
- Dunn, J., Runge, R. & Snyder, M. Wearables and the medical revolution. *Per. Med.* **15**, 429–448 (2018).
- Older Americans Drive Growth of Wearables. *eMarketer*. Available at: <https://www.emarketer.com/content/older-americans-drive-growth-of-wearables> (2018).
- Bent, B. et al. The Digital Biomarker Discovery Pipeline: An open source software platform for the development of digital biomarkers using mHealth and wearables data. *J. Clin. Transl. Sci.* 1–28. <https://doi.org/10.1017/cts.2020.511> (2020).
- Witt, D. R., Kellogg, R. A., Snyder, M. P. & Dunn, J. Windows into human health through wearables data analytics. *Curr. Opin. Biomed. Eng.* **9**, 28–46 (2019).
- Russell, W. R. et al. Impact of diet composition on blood glucose regulation. *Crit. Rev. Food Sci. Nutr.* **56**, 541–590 (2016).
- Boulé, N. G. et al. Metformin and exercise in type 2 diabetes: examining treatment modality interactions. *Diabetes Care* **34**, 1469–1474 (2011).
- Lloyd, C., Smith, J. & Weinger, K. Stress and diabetes: a review of the links. *Diabetes Spectr.* **18**, 121–127 (2005).
- Qian, J. & Scheer, F. A. J. L. Circadian system and glucose metabolism: implications for physiology and disease. *Trends Endocrinol. Metab.* **27**, 282–293 (2016).
- Mauvais-Jarvis, F. Gender differences in glucose homeostasis and diabetes. *Physiol. Behav.* **187**, 20–23 (2018).
- Panzer, C., Lauer, M. S., Brieke, A., Blackstone, E. & Hoogwerf, B. Association of fasting plasma glucose with heart rate recovery in healthy adults: a population-based study. *Diabetes* **51**, 803–807 (2002).
- Kenny, G. P., Sigal, R. J. & McGinn, R. Body temperature regulation in diabetes. *Temperature* **3**, 119–145 (2016).
- Grandinetti, A. et al. Impaired glucose tolerance is associated with post-ganglionic sudomotor impairment. *Clin. Auton. Res.* **17**, 231–233 (2007).
- Maarek, A., Rao, G. & Gandhi, P. Detection of neuropathy using a sudomotor test in type 2 diabetes. *Degener. Neurol. Neuromuscul. Dis.* **5**, 1 (2015).
- Zeevi, D. et al. Personalized nutrition by prediction of glycemic responses. *Cell* **163**, 1079–1094 (2015).
- Sampson, M. et al. Discordance in glycemic categories and regression to normality at baseline in 10,000 people in a Type 2 diabetes prevention trial. *Sci. Rep.* **8**, 1–11 (2018).
- Joshi, A. et al. Patterns of glycemic variability during a diabetes self-management educational program. *Med. Sci.* **7**, 52 (2019).
- Wilboux, M., Wölnerhanssen, B. K., Meyer-Gerspach, A. C., Beglinger, C. & Pfister, M. Characterizing the dynamic interaction among gastric emptying, glucose absorption, and glycemic control in nondiabetic obese adults. *Am. J. Physiol. Integr. Comp. Physiol.* **312**, R314–R323 (2017).
- Franz, M. J. Protein: metabolism and effect on blood glucose levels. *Diabetes Educ.* **23**, 643–651 (1997).
- Lamothe, L. M. et al. The scientific basis for healthful carbohydrate profile. *Crit. Rev. Food Sci. Nutr.* **59**, 1058–1070 (2019).
- Capuano, E. The behavior of dietary fiber in the gastrointestinal tract determines its physiological effect. *Crit. Rev. Food Sci. Nutr.* **57**, 3543–3564 (2017).
- Clark, M. J. & Slavin, J. L. The effect of fiber on satiety and food intake: a systematic review. *J. Am. Coll. Nutr.* **32**, 200–211 (2013).
- Wheeler, M. L. & Pi-Sunyer, F. X. Carbohydrate issues: type and amount. *J. Am. Diet. Assoc.* **108** (4 Suppl 1), S34–9 (2008).
- Cozma, A. I. et al. Effect of fructose on glycemic control in diabetes: a systematic review and meta-analysis of controlled feeding trials. *Diabetes Care* **35**, 1611–1620 (2012).
- Ter Horst, K. W., Schene, M. R., Holman, R., Romijn, J. A. & Serlie, M. J. Effect of fructose consumption on insulin sensitivity in nondiabetic subjects: a systematic review and meta-analysis of diet-intervention trials^{1,2}. *Am. J. Clin. Nutr.* **104**, 1562–1576 (2016).
- Bell, K. J. et al. Estimating insulin demand for protein-containing foods using the food insulin index. *Eur. J. Clin. Nutr.* **68**, 1055–1059 (2014).
- Beebe, C. A. et al. Effect of temporal distribution of calories on diurnal patterns of glucose levels and insulin secretion in NIDDM. *Diabetes Care* **13**, 748–755 (1990).
- Lopez-Minguez, J., Gómez-Abellán, P. & Garaulet, M. Timing of breakfast, lunch, and dinner. Effects on obesity and metabolic risk. *Nutrients* **11**, 2624 (2019).
- Poggialle, E., Jamshed, H. & Peterson, C. M. Circadian regulation of glucose, lipid, and energy metabolism in humans. *Metabolism* **84**, 11–27 (2018).
- Peter Adams, O. The impact of brief high-intensity exercise on blood glucose levels. *Diabetes Metab. Syndr. Obes. Targets Ther.* **6**, 113–122 (2013).

59. Boulé, N. G., Haddad, E., Kenny, G. P., Wells, G. A. & Sigal, R. J. Effects of exercise on glycemic control and body mass in type 2 diabetes mellitus: A meta-analysis of controlled clinical trials. *J. Am. Med. Assoc.* **286**, 1218–1227 (2001).
60. Umpierre, D. et al. Physical activity advice only or structured exercise training and association with HbA1c levels in type 2 diabetes: A systematic review and meta-analysis. *J. Am. Med. Assoc.* **305**, 1790–1799 (2011).
61. Snowling, N. J. & Hopkins, W. G. Effects of different modes of exercise training on glucose control and risk factors for complications in type 2 diabetic patients: a meta-analysis. *Diabetes Care* **29**, 2518–2527 (2006).
62. Van Dijk, J. W. et al. Effect of moderate-intensity exercise versus activities of daily living on 24-hour blood glucose homeostasis in male patients with type 2 diabetes. *Diabetes Care* **36**, 3448–3453 (2013).
63. Wiesli, P. et al. Acute psychological stress affects glucose concentrations in patients with type 1 diabetes following food intake but not in the fasting state. *Diabetes Care* **28**, 1910–1915 (2005).
64. Yitshak-Sade, M., Mendelson, N., Novack, V., Codish, S. & Liberty, I. F. The association between an increase in glucose levels and armed conflict-related stress: a population-based study. *Sci. Rep.* **10**, 1–6 (2020).
65. Wong, H., Singh, J., Go, R. M., Ahluwalia, N. & Guerrero-Go, M. A. The effects of mental stress on non-insulin-dependent diabetes: determining the relationship between catecholamine and adrenergic signals from stress, anxiety, and depression on the physiological changes in the pancreatic hormone secretion. *Cureus* **11**, e5474 (2019).
66. Usselman, C. W. et al. Hormone phase influences sympathetic responses to high levels of lower body negative pressure in young healthy women. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **311**, R957–R963 (2016).
67. Bo, S. et al. Effects of meal timing on changes in circulating epinephrine, norepinephrine, and acylated ghrelin concentrations: a pilot study. *Nutr. Diabetes* **7**, 303 (2017).
68. Tryon, M. S., Carter, C. S., DeCant, R. & Laugero, K. D. Chronic stress exposure may affect the brain's response to high calorie food cues and predispose to obesogenic eating habits. *Physiol. Behav.* **120**, 233–242 (2013).
69. Zamani-Alavijeh, F., Araban, M., Koohestani, H. R. & Karimy, M. The effectiveness of stress management training on blood glucose control in patients with type 2 diabetes. *Diabetol. Metab. Syndr.* **10**, 39 (2018).
70. Morris, C. J., Yang, J. N. & Scheer, F. A. J. L. The impact of the circadian timing system on cardiovascular and metabolic function. *Prog. Brain Res.* **199**, 337–358 (2012).
71. Nedeltcheva, A. V. & Scheer, F. A. J. L. Metabolic effects of sleep disruption, links to obesity and diabetes. *Curr. Opin. Endocrinol. Diabetes Obes.* **21**, 293–298 (2014).
72. Mattson, M. P. et al. Meal frequency and timing in health and disease. *Proc. Natl. Acad. Sci. USA* **111**, 16647–16653 (2014).
73. Radziuk, J. & Pye, S. Diurnal rhythm in endogenous glucose production is a major contributor to fasting hyperglycaemia in type 2 diabetes. Suprachiasmatic deficit or limit cycle behaviour? *Diabetologia* **49**, 1619–1628 (2006).
74. Nordström, A., Hadrévi, J., Olsson, T., Franks, P. W. & Nordström, P. Higher prevalence of type 2 diabetes in men than in women is associated with differences in visceral fat mass. *J. Clin. Endocrinol. Metab.* **101**, 3740–3746 (2016).
75. Chen, L., Magliano, D. J. & Zimmet, P. Z. The worldwide epidemiology of type 2 diabetes mellitus - Present and future perspectives. *Nat. Rev. Endocrinol.* **8**, 228–236 (2012).
76. Tracey, M. L. et al. The prevalence of Type 2 diabetes and related complications in a nationally representative sample of adults aged 50 and over in the Republic of Ireland. *Diabet. Med.* **33**, 441–445 (2016).
77. Wild, S., Roglic, G., Green, A., Sicree, R. & King, H. Global Prevalence of Diabetes: estimates for the year 2000 and projections for 2030. *Diabetes Care* **27**, 1047–1053 (2004).
78. Anish, T. et al. Gender difference in blood pressure, blood sugar, and cholesterol in young adults with comparable routine physical exertion. *J. Fam. Med. Prim. Care* **2**, 200 (2013).
79. Kautzky-Willer, A., Harreiter, J. & Pacini, G. Sex and gender differences in risk, pathophysiology and complications of type 2 diabetes mellitus. *Endocr. Rev.* **37**, 278–316 (2016).
80. Kautzky-Willer, A., Kosi, L., Lin, J. & Mihaljevic, R. Gender-based differences in glycaemic control and hypoglycaemia prevalence in patients with type 2 diabetes: Results from patient-level pooled data of six randomized controlled trials. *Diabetes Obes. Metab.* **17**, 533–540 (2015).
81. Valensi, P. et al. Influence of blood glucose on heart rate and cardiac autonomic function. The DESIR study. *Diabet. Med.* **28**, 440–449 (2011).
82. Frier, B. M., Scherthaner, G. & Heller, S. R. Hypoglycemia and cardiovascular risks. *Diabetes Care* **34**, S132 (2011).
83. Molnar, G. W. & Read, R. C. Hypoglycemia and body temperature. *J. Am. Med. Assoc.* **227**, 916–921 (1974).
84. Tesfaye, S. et al. Diabetic neuropathies: update on definitions, diagnostic criteria, estimation of severity, and treatments. *Diabetes Care* **33**, 2285–2293 (2010).
85. Wu, J.-S. et al. Epidemiological evidence of altered cardiac autonomic function in subjects with impaired glucose tolerance but not isolated impaired fasting glucose. *J. Clin. Endocrinol. Metab.* **92**, 3885–3889 (2007).
86. Low, P. A., Tomalia, V. A. & Park, K.-J. Autonomic function tests: some clinical applications. *J. Clin. Neuro.* **9**, 1–8 (2013).
87. Peters, E., Slovic, P., Västfjäll, D. & Mertz, C. K. Intuitive numbers guide decisions. *Judgment Decision Making* **3**, 31 (2008).
88. Ambrose, M. L. & Kulik, C. T. Old friends, new faces: motivation research in the 1990s. *J. Manag.* **25**, 231–292 (1999).
89. Is my Dexcom CGM sensor accurate? (accessed 21 April 2021); <https://www.dexcom.com/faqs/is-my-dexcom-sensor-accurate>.
90. Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y. H. & Koo, B.-H. Stress and heart rate variability: a meta-analysis and review of the literature. *Psychiatry Investig.* **15**, 235–245 (2018).
91. Bergenstal, R. M. et al. Glucose management indicator (GMI): a new term for estimating A1C from continuous glucose monitoring. *Diabetes Care* **41**, 2275–2280 (2018).
92. CDC. *National Diabetes Statistics Report 2020. Estimates of diabetes and its burden in the United States.* (2020).
93. Benedek, M. & Kaernbach, C. A continuous measure of phasic electrodermal activity. *J. Neurosci. Methods* **190**, 80–91 (2010).
94. Farnsworth, B. Skin Conductance Response - What it is and How to Measure it - iMotions. (2019) (accessed 15 October 2020); <https://imotions.com/blog/skin-conductance-response/>.
95. Braithwaite, J. J., Derrick, D., Watson, G., Jones, R. & Rowe, M. *A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments.* (2015).
96. Boucsein, W. et al. Publication recommendations for electrodermal measurements. *Psychophysiology* **49**, 1017–1034 (2012).
97. Healey, J. A. & Picard, R. W. Detecting stress during real-world driving tasks using physiological sensors. in *IEEE Transactions on Intelligent Transportation Systems* (2004).
98. SciPy — SciPy v1.5.2 Reference Guide. (accessed 15 October 2020); <https://docs.scipy.org/doc/scipy/reference/> (2020).
99. Waterhouse, J., Fukuda, Y. & Morita, T. Daily rhythms of the sleep-wake cycle. *J. Physiol. Anthropol.* **31**, 1–14 (2012).
100. scikit-learn: machine learning in Python — scikit-learn 0.24.1 documentation (accessed 4 February 2021); <https://scikit-learn.org/stable/>.
101. Python XGBoost Documentation. Available at: <https://xgboost.readthedocs.io/en/latest/> (2020).

ACKNOWLEDGEMENTS

This work was supported by Duke MEDx. B.B. is a Duke Forge predoctoral fellow. J.D. is a MEDx Investigator and a Whitehead Scholar.

AUTHOR CONTRIBUTIONS

B.B. was involved in study design, data collection, data analysis and interpretation, model development, and manuscript preparation. P.C. and A.W. were involved in data collection. M.H. was involved with data analysis. C.T. and A.W. were involved in study design. M.F. was involved in concept development, funding, and study design. M.C. was involved with data interpretation and manuscript preparation. J.D. was involved with concept development, funding, study design, data 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/s41746-021-00465-w>.

Correspondence and requests for materials should be addressed to J.P.D.

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) 2021