Identifying Barriers to Self-Management in Type 1 Diabetes Using Momentary Assessment and Machine Learning

Introduction

Type 1 diabetes (T1D) is a prevalent chronic illness with increasing incidence rates reported worldwide [1,2]. It is an autoimmune disorder where the body does not produce insulin and requires patients to perform critical self-management tasks multiple times per day [3]. Two key self-management tasks inT1D are frequent monitoring of blood glucose (BG) and administering insulin. These tasks help manage glycemic control to avoid or delay serious short- and long- term consequences, such as retinopathy, neuropathy, and mortality [4–6]. Mealtimes are a critical time for diabetes self-management.

Adolescents and young adults have the worst glycemic control of any age group [4]. For young people with diabetes, living successfully with T1D is particularly hard due to many potential psychosocial and contextual barriers to self-management [7–9]. A recommended approach used to improve self-management involves promoting and supporting problem solving skills to reduce barriers [10]. To identify problems related to self-management, patients, caregivers, and clinicians must rely on blood glucose and insulin administration data from devices along with a patient recall of behavioral, emotional, and/or contextual events that could pose barriers to self-management. However, utilizing retrospective memory or recall for events that are days or weeks in the past has been identified as generally unreliable and potentially biased in nature [11]. Unreliable recall of events in diabetes problem solving could result in modifications to the insulin regimen that are not based on reliable information.

To address the limitations of human recall and bias in health behavior research, ecological momentary assessment (EMA) methods have been developed and successfully utilized in a range of health conditions. In contrast to traditional assessment methods, EMA utilizes more frequent and in-vivo ambulatory assessment of factors that impact health behaviors and decision-making. EMA methods provide a more proximal, and often more accurate, technology-mediated method to monitor and assess the contexts, subjective experiences, and processes that surround health decisions in daily life [12,13]. In particular, EMA methods provide more relevant and frequent observations per person and generates rich data to assess correlates of health behavior more accurately and identify novel correlates for intervention [14].

Many studies in the EMA literature typically use mixed effects or hierarchical linear modeling (HLM) [15,16]. That analytic approach does not provide a means to automate analyses or use learning algorithms that improve and integrate incoming data over time. A promising approach for identifying such a model involves integrating EMA with techniques and tools associated with machine learning, which is a data analysis method that automates statistical model building by identifying patterns and making decisions with minimal human intervention [17,18]. Machine learning has been used with wearable sensor data and may also be useful in analyzing intensive self-report data, such

as EMA. Machine learning techniques provide a viable means to examine both big and small data by providing automated classification and prediction for more feasible behavioral intervention.

The objective of our study was to develop a machine learning algorithm to predict risk for missed self-management. We sought to identify the momentary psychosocial and contextual factors that have an impact on T1D self-management assessed by EMA. To achieve these objectives, we trained and compared a number of machine learning models through a learned filtering architecture (LFA) to explore the extent to which EMA data could predict completion of two self-management behaviors: insulin administration (IA) and self-monitoring of blood glucose (SMBG). By integrating these two strategies (EMA and machine learning), we aim to provide researchers with not only a better understanding of what may hinder or promote adolescents' adherence of their T1D regimen from a behavioral perspective, but also an efficient and adaptive analytic computational method.

Methods

This study analyzed data from a feasibility trial of the mobile EMA and feedback app called MyDay[1], which is a self-management feedback and problem-solving tool designed for adolescent T1D patients [19]. Youth from the Vanderbilt Eskind Pediatrics Diabetes Clinic were invited to participate in a 30-day assessment period if (1) they were between the age of 13 and 19, (2) had been diagnosed of T1D for at least 6 months, (3) owned either an Android or iPhone smartphone, (4) understood and spoke English, and (5) were willing to use a Bluetooth blood glucose meter during the study.

A total of 48 participants were recruited for the pilot study. Three participants dropped out of the study noting competing demands, leaving 45 for our analyses. Subjects were randomized on a 2:1 ratio to the MyDay app + Bluetooth blood glucose (BG) meter group (n=31) and a control group (n=14). The control group provided SMBG data only using Bluetooth BG meters but did not use the MyDay app. Design processes, engagement, and momentary relationships results for MyDay were published previously [19–21].

Momentary Assessments and Glucose Meter Data

All SMBG data was objectively assessed using iHealth [22] glucometers. The iHealth glucometers are commercially available Bluetooth Low-Energy meters that can upload data automatically to the iHealth secure cloud server via their open API. Thirty-one participants were instructed to use the MyDay app at each mealtime and bedtime to answer questions that focused on factors likely to impact diabetes self-management.

MyDay provided notifications to complete the EMA assessment personalized to typical mealtimes identified by participants. Timestamps were associated with all data entries. Only mealtime EMA were used in analyses. Variables analyzed in relation to self-management outcomes were organized into the following subsets. The first two domains of variables were collected for all participants: (1) demographics obtained at baseline (i.e., gender, age, father's education, mother's education, family income, and

[1] The study was reviewed and approved by the Vanderbilt University Institutional Review Board (IRB). All parents provided consent before adolescents provided assent. Both consent and assent were obtained before study procedures commenced.

race) and (2) time variables that were coded using the original timestamps of the collected data entries, e.g., weekday, weekend, and mealtime (breakfast, lunch, dinner).

The next three domains of EMA data were available only for the 31 participants using the MyDay app: (3) context related to who was with the youth at time of self-management (i.e., parent, sibling, alone, casual friend, close friend, other family, other person, strangers, and boyfriend/girlfriend) and where the youth was at time of self-management (i.e., home, school, work, restaurant, friends' house, or on the road), (4) stress, fatigue, mood levels at the reported self-management event: scored as 0-100 with higher scores indicating greater stress, more fatigue, and worse negative mood, and (5) selected situational barriers at time of self-management event (i.e., participant was rushing, feeling sick, on the road, hungry, wanting privacy, busy, without supplies, or having fun). Details of the EMA data collection process can be found in [20].

The dataset was preprocessed using the following statistical approaches. First, it was observed that the dataset contained missing values in demographic features: 8.89% missing for both father's education and household income, and 26.67% missing for mother's education (the percentage of missing values in each category is also reported as an N/A entry in Table 1). In this study, the missing values of a feature were imputed using the mode value for features of mother's education and father's education and median value for the feature of family income. Ordinal categorical variables whose order of the values are significant, such as parent education and family income level, were each transformed into a single feature with numeric values, whereas nominal variables whose significance cannot be assumed, such as participant race and day of week, were converted to numeric values using one-hot-encoding. Each feature was normalized using the min-max scaler such that all final values of that feature were between 0 and 1. The source code for data preprocessing is included in Appendix A.

Outcomes

We examined three self-management behavioral outcomes:

1. Daily SMBG frequency of "less than 4" or "4 or more" times a day. Four glucose checks per day is generally considered the minimum recommended [23],
2. Missed SMBG at mealtimes,
3. Insulin administration (IA) at mealtimes

Data from all subjects were available (n=45) for analyses examining daily number of SMBG from meters. The data that was available for all subjects were demographic and time variables. Analyses for outcomes 2 and 3 examined data from participants who used the MyDay EMA app (n=31), which obtained mealtimes.

The Learned Filtering Architecture

To extract domains of variables to predict IA and SMBG self-management behaviors via the training of a series of models, a learned filtering architecture (LFA) was created in this study as a byproduct, and a similar process was used in [24] but not formally constructed. For this study, the LFA created and compared four machine learning models: K-Nearest Neighbors (KNN), Logistic Regression (LR), Random Forest (RF), and Support Vector Machines (SVM). These models performed binary classification for each behavioral outcome observed in this study.

KNN classifies each sample by finding its K-most similar instances in the training set and chooses the class that majority of the neighboring instances belong to [25]. The value of K is determined by running KNN models with varying K values iteratively and selecting the K value that produced the most optimal model. LR is a statistical model that classifies a sample by predicting the probability of an output using the maximum likelihood estimation method and using a probability threshold ($p = 0.5$ was used in our study as the threshold such that an output with a probably of $p >= 0.5$ was classified as true and false otherwise) to separate the two classes [26]. RF which is a popular ensemble learning method that trains multiple decision trees on different parts of the dataset and then averages the results to improve classification accuracy [27]. The number of trees, or "estimators" is determined by running a number of RF models with varying estimator values, such as 10, 50, 100, etc, and selecting the value that produced the most performant model. SVM works by finding an optimal hyperplane in the feature space that optimally separates the data points into different classes [28].

Figure 1 presents the workflow of this LFA and shows that SMBG data and EMA data collected from the MyDay app were integrated as a complete dataset fed into the LFA (steps 1 and 2). The LFA then performed specified data pre-processing, such as normalizing numeric values, removing entries that were empty or had many missing features, and one-hot encoding, based on the type of each column (step 3). After step 3, a data filtering process began, where subsets of variables were extracted from the cleaned data either based on configurable user input, such as the names of columns which would be grouped to create a clinically meaningful, or to-be-observed, feature subset. The features were grouped as described above to create multiple data subsets. Due to the small sample size of the data available, the data subsets were each split further for evaluating each classification model using cross validation (steps 4a and 4b).

The LFA calculates the distribution of the target variable of each dataset. If the dataset is balanced, it evaluates each model using k-fold cross validation that further splits the data into training and validation sets k times and produces mean values of the performance metrics. Otherwise, if the classes are unevenly distributed, it uses the stratified k-fold cross validation to create k (k=7) splits, with each split of training and validation sets maintaining the original class distributions. The performance metrics are averaged across the results from the k different splits. The process then repeats for each of the machine learning models specified (step 6).

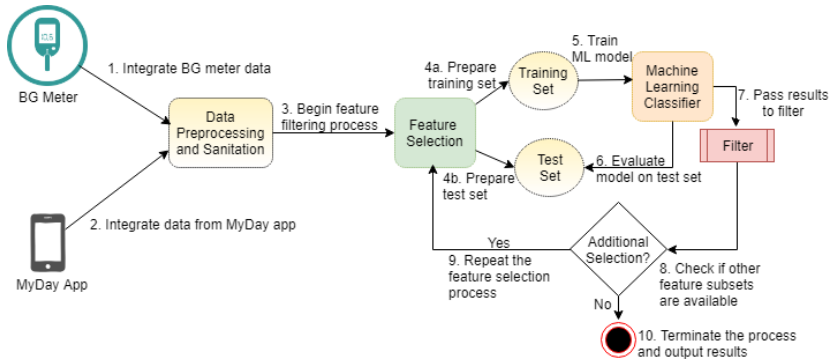

Figure 1. Iterative Process of The Learned Filtering Architecture (LFA)

Specifically, we used the following metrics to assess the models: (1) accuracy, which is the percentage of correct predictions, (2) precision, which is the ratio of true positives and all predicted positives that evaluates what proportion of predicted positives was actually correct, (3) recall, which is the ratio of true positives and all actual positives that calculates what proportion of actual positives was predicted correctly, (4) F1 score, which evenly weighs precision and recall, and (5) for imbalanced classification tasks, the Brier score, which is a continuous scoring loss function that evaluates the goodness of predicted probabilities in a classification task – a lower number corresponds to a stronger model and vice versa.

The classification results were then used by the filter component to compare them across all feature subsets (step 7). The filter component had a configurable tolerance value that was used to select feature subset(s) with relatively good classification results compared to the best performing model(s). Next, the LFA checked whether additional feature groups remained to be processed (step 8). If so, feature selection was repeated to create the next data subset (step 9). Otherwise, the filtering process would terminate and output the filtered results, i.e., variable groups with relatively strong predictive power of the outcomes (step 10).

The classification results were filtered to extract the best predictor group(s) for the target class variable. For example, if the performance metrics overall exceed the specified threshold values (such as 15% compared to the performance metrics of the model trained with all features together), the predictor group was added to the final output queue. When all variable groups were evaluated, LFA returned the final insights obtained from the input, i.e., feature groups that had significant predictive power of the outcomes observed in this study.

Although the number of observations per participant was substantial (average number of observations = 60), the overall number of participants was relatively small (n=45). The collected data thus had some imbalance in the distribution of the outcomes, with missed mealtime insulin being a relatively less frequent event. Classification models constructed using imbalanced datasets may result in the minority class being neglected [29]. Techniques such as Synthetic Minority Oversampling Technique (SMOTE) [30] and Tomek link (T-link) [31] have been used in the literature for training imbalanced data, especially for small datasets [32–35]. However, given the small size of the population under this study, using such sampling methods would risk introducing bias and misleading results. In this study we therefore employed a stratified K-fold (k=7) cross validation [36] evaluation method instead of randomly oversampling or introducing synthetic samples based on the existing data.

In stratified K-fold cross validation, the original dataset is randomly split into k folds. Each fold is further split into separate training and testing sets that are used to generate evaluation metrics of a model. The distributions of the majority and minority classes within each training and testing set follows the distribution of the majority and minority classes in the original dataset. After the model has been trained and tested against all k folds, the results are averaged to represent the overall classification performance.

In addition to the machine learning methods described above, we also employed a Bayesian hierarchical regression model for the entire EMA dataset that has a large

number of features, but a small sample size. This approach was applied to confirm the inferential power of the collected EMA data, rather than focusing on which specific category is the most predictive of the outcomes.

Hierarchical modeling can capture similarities of multiple subjects within a dataset while allowing estimations of individual parameters for data containing multiple subjects. With the Bayesian approach, the entire dataset is considered known information that is used to derive distributions of unknown parameters of the model. It is a probabilistic model that intends to estimate expected values or density.

In our analysis, we applied Markov Chain Monte Carlo (MCMC) methods [37] to assist with the model formation and sampling process. Monte Carlo is a method for randomly sampling a probability distribution to approximate some desired target function. Markov Chain is a sampling technique that can generate a sequence of random samples where the current sample is drawn based on the prior sample. The goal of MCMC is to construct a Markov Chain that eventually stabilizes on the desired quantity to be inferred. Specifically, we created a non-centered Bayesian hierarchical model to estimate the likelihoods of SMBG and IA.

Results

This section analyzes the results obtained from the LFA constructed in accordance with the methods described above.

Daily SMBG Frequency

The sample of n=45 participants were on average 13.33 years of age (SD 1.67), were 53.33% female, 84.44% White, 57.46% used an insulin pump and had a mean HbA1c (indicating overall glycemic control) of 9.03% (SD 1.91). Additional characteristics of the sample are summarized in Table 1.

Table 1. Characteristics of the Sample (n=45, entries marked as N/A represent data not reported)

Variable	Mean (SD) or %
Age	13.33 (1.67)
Female	53.33%
Male	46.67%
Race / ethnicity	
White	84.44%
African American	10.22%
Asian	2.22%
Hispanic	2.22%
Other	0.00%
Father education	
Less than high school	2.22%
High school / GED	28.89%
2-year college	15.56%
4-year college	33.33%
Master's degree	11.11%
Doctoral degree	0.00%

N/A	8.89%
Mother education	
Less than high school	0.00%
High school / GED	22.22%
2-year college	26.67%
4-year college	37.78%
Master's degree	4.44%
Doctoral degree	0.00%
N/A	26.67%
Household Income	
Less than $25,000	4.44%
25,001 – 35,000	6.67%
35,001 – 35,000	15.56%
75,001 – 35,000	31.11%
100,001 – 100,000	26.67%
More than $70,000	6.67%
N/A	8.89%
Duration of diabetes (years)	5.47 (3.59)
HbA1c	9.03 (1.91)
Use insulin pump (yes)	57.46%

A total of 4,475 blood glucose (BG) measurements were obtained from iHealth Bluetooth meters used by all participants (n=45). For this analysis the demographic and time variables were studied to identify if they had any impact on the outcome of SMBG frequency per day. The measurements were aggregated on a daily basis to obtain a new dataset of 1,231 entries, with each entry per participant being the total number of measurements an individual had each day during the study period. SMBG frequency ranged between 1-12 measurements per day. If a participant did not report an entry on a particular day, the entry for that day was not assumed to have an SMBG daily frequency of 0 and hence the entry for participant on that day was not created.

Several distributions of SMBG daily frequency were observed. There were 591 entries with *Below 4* frequency and 640 entries with *4 or Above*. Out of all the classifiers trained with the same training data, RF was the best performing model based on the overall classification metrics using the same test data. The mean and standard deviation (SD) values of the evaluation results from the best performing RF model are shown in Table 2 for SMBG frequency *Below 4* (the source code comparing the performance of all machine learning models is included in Appendix A). The filter then compared the benchmark value with the outcome classification results obtained from each variable group. A tolerance value of 15% was configured for the filter to select subsets with significant predictive power. As shown in Table 2, demographics variable group for SMBG frequency resulted in a better performance than time variables and all variables.

Table 2. SMBG Below 4 Classification Results

Feature Group	Accuracy Mean (SD)	Precision Mean (SD)	Recall Mean (SD)	F1 Score Mean (SD)

Demographics	75.5% (0.044)	0.75 (0.078)	0.72 (0.066)	0.74 (0.062)
Time variables	49.3% (0.040)	0.46 (0.064)	0.21 (0.142)	0.28 (0.124)
All	67.9% (0.031)	0.67 (0.063)	0.68 (0.057)	0.67 (0.035)

Missed Mealtime SMBG and Insulin Administration

From the app group (n=31), a total of 1,869 entries were associated with breakfast, lunch, or dinner and used to analyze factor(s) that could impact SMBG and IA. Missed IA had a distribution of 1:5.72 for True (missed) vs False (administered) outcomes. In contrast, the outcome missed SMBG had a class distribution of 1:5.44 for True (missed) vs False (checked). LFA created classification models for each variable group (i.e., demographic, time, social context, and psychosocial) using the stratified K-fold approach as discussed previously. Similar to the previous experiment, the RF model resulted in the best classification performance in all metrics compared to other models (the source code comparing the performance of all machine learning models is included in Appendix B)..

Tables 3 and 4 present the average classification results of missed SMBG and missed IA, respectively. The results showed mixed sentiments on the predictive power of individual groups of indicators on the self-management behavior, but their combined effect can be used to infer when the lack of SMBG or IA occurred with a high accuracy and high precision.

Table 3. Missing Mealtime Blood Glucose Measurement Classification Results

Feature Group	Accuracy	Precision	Recall	F1 Score	Brier
Demographics	78.3%	0.38	0.62	0.47	0.22
Time variables	49.9%	0.13	0.42	0.20	0.51
Social Context	60.7%	0.21	0.55	0.30	0.25
Stress, Fatigue, Mood	74.1%	0.22	0.29	0.25	0.33
Barriers	72.5%	0.33	0.44	0.33	0.25
All	88.3%	0.78	0.35	0.48	0.12
All (MCMC)	87.3%	0.78	0.25	0.38	0.13

Table 4. Missing Mealtime Insulin Administration Classification Results

Feature Group	Accuracy	Precision	Recall	F1 Score	Brier
Demographics	64.8%	0.25	0.65	0.36	0.36
Time variables	58.9%	0.21	0.64	0.32	0.41
Social Context	49.0%	0.16	0.59	0.25	0.51
Stress, Fatigue, Mood	74.1%	0.22	0.28	0.25	0.32
Barriers	72.5%	0.26	0.44	0.32	0.27
All	85.8%	0.61	0.14	0.23	0.14
All (MCMC)	85.4%	0.54	0.15	0.24	0.15

Commented [DZ1]:

Discussion

To better understand the factors impacting self-management behavior of adolescents with T1D, this study applied machine learning analyses to construct a learning filter architecture (LFA) using demographic, and momentary psychosocial and self-management data. The relative association of five domains of variables for predictability of self-management behaviors was compared using all the variables collectively as the benchmark.

For the demographic data, the results indicated that demographics were most associated with average daily SMBG frequency. These results highlight the value of social determinants of health, as defined by demographics. While demographic factors are generally not modifiable, social determinants of health are increasingly used to adapt care to for those who are most vulnerable and may not receive the full benefit of current approaches to healthcare [36, 37].

The EMA data was able to infer non-adherence for SMBG and insulin with a high accuracy and precision. Although the recall score is low, there is high confidence that the non-adherence events identified by the model are truly non-adherence ones. One reason for the lower recall score has to do with the small datasets that have disparities in the frequencies of observed classes or outcomes. Nonetheless, this study shows promise in the collection of larger datasets that would more effectively power a classifier that is deployable in the real world.

These results support the feasibility and value of integrating EMA and machine learning to improve behavioral assessment and automate behavioral pattern recognition in healthcare [18,38]. Our learned models show promise to quantify the impact of psychosocial factors on self-management. In diabetes, stress and mood are modifiable factors that may be positively influenced through coping and problem-solving interventions [39,40]. The use of machine learning and EMA was also seen in a recent study on Tinnitus (the phantom perception of sounds), where a random forest classifier was applied on EMA data collected from the TrackYourTinnitus mobile app across devices to predict the mobile operating system used [41].

Social context also provides a frame for understanding risk and may be modified in by interventions focused on social competence and problem solving [39]. In previous studies [42,43] using behavioral observation in the context of identifying patterns of hand hygiene compliance monitoring, from which we obtained useful initial insights into which domains of variables had the most impact on compliance behavior. Based on the current findings, similar experiments are needed with larger samples to prioritize multiple potential domains of influence on health behaviors, and advance the assessment and analytic approaches utilized here.

Moving forward, the use of primarily intensive self-reported and passive psychosocial and behavioral data streams combined with machine learning could provide the basis for population-based monitoring systems to help guide automated pattern detection for clinical risk management. For example, experimental unobtrusive indicators of mealtimes are in development [44] and insulin administration is available via pumps [44]. If successful, additional passive data streams would greatly improve our methodological rigor and reach [45].

The LFA machine learning methods employed here should be applied to a large diverse sample of patients to confirm and expand results reported in this paper. Although

passive methods are increasingly used to infer behavior and psychosocial status [46,47], there are important subjective experiences, such as mood, which may continue to require self-report. For the foreseeable future, both self-reported real-time data and passive data, such as social networking [48], may be integrated to optimize insights for healthcare.

Prior research using traditional retrospective questionnaire methods has focused largely on identifying psychosocial correlates and predictors of self-management in chronic illness in general and specifically in diabetes [9]. With few exceptions, little research using EMA has been conducted in diabetes. The few studies conducted have uniquely identified time-based factors, such as time of day and momentary negative mood, as related to self-management behaviors [49–51].

Machine learning analyses have been applied in various studies, focusing largely on the improvement of diabetes management and control. Earlier studies have constructed and fine-tuned different machine learning models to predict future blood glucose levels based on historical physiological data, [52–54], detect incorrect blood glucose measurements in [55], predict hypoglycemia [56,57], manage insulin dosing [58], and applied to provide lifestyle support integrating food recognition, and energy expenditures [59,60]. Our study results reported here advance the assessment and analysis of factors previously associated with self-management, including stress [49], mood [61,62], stigma [9,63], and social contexts [8,12]. Our study also uniquely assesses novel factors not previously studied in the T1D population, such as fatigue [64], location [65], social contexts [8], and contextual factors, such as rushing and traveling. The collected EMA data has a promising ability to infer the two diabetes self-management behaviors under study.

Future work will enhance MyDay's ability to utilize unobtrusive indicators of behaviors. For example, experimental unobtrusive indicators of mealtimes are in development and if successful would greatly enhance our methodological approach [45]. Finally, the LFA machine learning methods employed here will be applied to a large diverse sample of patients to confirm and expand results reported in this paper. Future systems will benefit from combining self-report of subjective human experiences together with passive indicators of factors that impact health behavior decision-making in daily life.

Acknowledgements

This work was supported by a grant from the National Institutes of Health, National Institute for Diabetes, Digestive, and Kidney Disease (NIDDK) DP3 DK097706, and the National Center for Advancing Translational Science (NCATS) UL1 TR000445.

Conflicts of Interest

None declared

Abbreviations

BG: blood glucose
CGM: continuous glucose monitoring
EMA: ecological momentary assessment
IA: insulin administration
KNN: k-nearest neighbors

LFA: learned filtering architecture
LR: logistic regression
MCMC: Markov Chain Monte Carlo
RF: random forest
SD: Standard Deviation
SMBG: self-monitoring of blood glucose
SMOTE: synthetic minority oversampling technique
SVM: support vector machines
T-link: Tomek link
T1D: type 1 diabetes

References

1. Borchers AT, Uibo R, Gershwin ME. The geoepidemiology of type 1 diabetes. Autoimmun Rev Elsevier; 2010;9(5):A355–A365.
2. Dabelea D, Mayer-Davis EJ, Saydah S, Imperatore G, Linder B, Divers J, Bell R, Badaru A, Talton JW, Crume T. Prevalence of type 1 and type 2 diabetes among children and adolescents from 2001 to 2009. Jama American Medical Association; 2014;311(17):1778–1786.
3. Wen L, Ley RE, Volchkov PY, Stranges PB, Avanesyan L, Stonebraker AC, Hu C, Wong FS, Szot GL, Bluestone JA. Innate immunity and intestinal microbiota in the development of Type 1 diabetes. Nature Nature Publishing Group; 2008;455(7216):1109–1113.
4. Wood JR, Miller KM, Maahs DM, Beck RW, DiMeglio LA, Libman IM, Quinn M, Tamborlane W V, Woerner SE, Network TEC. Most youth with type 1 diabetes in the T1D Exchange Clinic Registry do not meet American Diabetes Association or International Society for Pediatric and Adolescent Diabetes clinical guidelines. Diabetes Care Am Diabetes Assoc; 2013;36(7):2035–2037.
5. Cleary PA, Dahms W, Goldstein D, Malone J, Tamborlane W V. Beneficial effects of intensive therapy of diabetes during adolescence: outcomes after the conclusion of the Diabetes Control and Complications Trial (DCCT). J Pediatr 2001;139:804–812.
6. Group DC and CTR. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Engl J Med Mass Medical Soc; 1993;329(14):977–986.
7. Hilliard ME, De Wit M, Wasserman RM, Butler AM, Evans M, Weissberg-Benchell J, Anderson BJ. Screening and support for emotional burdens of youth with type 1 diabetes: strategies for diabetes care providers. Pediatr Diabetes Wiley Online Library; 2018;19(3):534–543.
8. Wiebe DJ, Helgeson V, Berg CA. The social context of managing diabetes across the life span. Am Psychol American Psychological Association; 2016;71(7):526.
9. Mulvaney SA, Hood KK, Schlundt DG, Osborn CY, Johnson KB, Rothman RL, Wallston KA. Development and initial validation of the barriers to diabetes adherence measure for adolescents. Diabetes Res Clin Pract Elsevier; 2011;94(1):77–83.
10. Fitzpatrick SL, Schumann KP, Hill-Briggs F. Problem solving interventions for diabetes self-management and control: a systematic review of the literature.

Diabetes Res Clin Pract Elsevier; 2013;100(2):145–161.
11. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol Annual Reviews; 2008;4:1–32.
12. Dunton G, Dzubur E, Li M, Huh J, Intille S, McConnell R. Momentary assessment of psychosocial stressors, context, and asthma symptoms in hispanic adolescents. Behav Modif SAGE Publications Sage CA: Los Angeles, CA; 2016;40(1–2):257–280.
13. Linas B, Genz A, Westergaard RP, Chang LW, Bollinger RC, Latkin C, Kirk GD. Ecological momentary assessment of illicit drug use compared to biological and self-reported methods. JMIR mHealth uHealth JMIR Publications Inc., Toronto, Canada; 2016;4(1):e27.
14. Brannon EE, Cushing CC, Crick CJ, Mitchell TB. The promise of wearable sensors and ecological momentary assessment measures for dynamical systems modeling in adolescents: a feasibility and acceptability study. Transl Behav Med Oxford University Press; 2016;6(4):558–565.
15. Myers TC, Wonderlich SA, Crosby R, Mitchell JE, Steffen KJ, Smyth J, Miltenberger R. Is multi-impulsive bulimia a distinct type of bulimia nervosa: Psychopathology and EMA findings. Int J Eat Disord Wiley Online Library; 2006;39(8):655–661.
16. Hedeker D, Mermelstein RJ, Demirtas H. Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models. Stat Med Wiley Online Library; 2012;31(27):3328–3336.
17. Bishop CM. Pattern recognition. Mach Learn 2006;128(9).
18. Kim H, Lee S, Lee S, Hong S, Kang H, Kim N. Depression prediction by using ecological momentary assessment, actiwatch data, and machine learning: observational study on older adults living alone. JMIR mHealth uHealth JMIR Publications Inc., Toronto, Canada; 2019;7(10):e14149.
19. Mulvaney SA, Vaala S, Hood KK, Lybarger C, Carroll R, Williams L, Schmidt DC, Johnson K, Dietrich MS, Laffel L. Mobile momentary assessment and biobehavioral feedback for adolescents with type 1 diabetes: feasibility and engagement patterns. Diabetes Technol Ther Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA; 2018;20(7):465–474.
20. Zhang P, Schmidt D, White J, Mulvaney S. Towards precision behavioral medicine with IoT: Iterative design and optimization of a self-management tool for type 1 diabetes. 2018 IEEE Int Conf Healthc Informatics IEEE; 2018. p. 64–74.
21. Mulvaney SA, Vaala SE, Carroll RB, Williams LK, Lybarger CK, Schmidt DC, Dietrich MS, Laffel LM, Hood KK. A mobile app identifies momentary psychosocial and contextual factors related to mealtime self-management in adolescents with type 1 diabetes. J Am Med Informatics Assoc Oxford University Press; 2019;26(12):1627–1631.
22. iHealth Labs Inc [Internet]. Available from: https://ihealthlabs.com
23. Miller KM, Foster NC, Beck RW, Bergenstal RM, DuBose SN, DiMeglio LA, Maahs DM, Tamborlane W V. Current state of type 1 diabetes treatment in the US: updated data from the T1D Exchange clinic registry. Diabetes Care Am Diabetes Assoc; 2015;38(6):971–978.
24. Zhang P, White J, Schmidt D. Architectures and patterns for leveraging high-

frequency, low-fidelity data in the healthcare domain. 2018 IEEE Int Conf Healthc Informatics IEEE; 2018. p. 463–464.
25. Dudani SA. The distance-weighted k-nearest-neighbor rule. IEEE Trans Syst Man Cybern IEEE; 1976;(4):325–327.
26. Seber GAF, Lee AJ. Linear regression analysis. John Wiley & Sons; 2012. ISBN:1118274423
27. Liaw A, Wiener M. Classification and regression by randomForest. R news 2002;2(3):18–22.
28. Suykens JAK, Vandewalle J. Least squares support vector machine classifiers. Neural Process Lett Springer; 1999;9(3):293–300.
29. Chawla N V, Japkowicz N, Kotcz A. Special issue on learning from imbalanced data sets. ACM SIGKDD Explor Newsl ACM New York, NY, USA; 2004;6(1):1–6.
30. Chawla N V, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 2002;16:321–357.
31. Tomek I. AN EXPERIMENT WITH THE EDITED NEAREST-NIEGHBOR RULE. 1976;
32. Lusa L. Evaluation of smote for high-dimensional class-imbalanced microarray data. 2012 11th Int Conf Mach Learn Appl IEEE; 2012. p. 89–94.
33. Elhassan T, Aljurf M. Classification of imbalance data using tomek link (t-link) combined with random under-sampling (rus) as a data reduction method. Glob J Technol Optim S 2016;1.
34. Thai-Nghe N, Nghi DT, Schmidt-Thieme L. Learning optimal threshold on resampling data to deal with class imbalance. Proc IEEE RIVF Int Conf Comput Telecommun Technol 2010. p. 71–76.
35. Kotsiantis S, Kanellopoulos D, Pintelas P. Handling imbalanced datasets: A review. GESTS Int Trans Comput Sci Eng 2006;30(1):25–36.
36. Zeng X, Martinez TR. Distribution-balanced stratified cross-validation for accuracy estimation. J Exp Theor Artif Intell Taylor & Francis; 2000;12(1):1–12.
37. Qian SS, Stow CA, Borsuk ME. On monte carlo methods for Bayesian inference. Ecol Modell Elsevier; 2003;159(2–3):269–277.
38. Roux AVD, Katz M, Crews DC, Ross D, Adler N. Social and behavioral information in electronic health records: new opportunities for medicine and public health. Am J Prev Med Elsevier; 2015;49(6):980–983.
39. Whittemore R, Jaser SS, Jeon S, Liberti L, Delamater A, Murphy K, Faulkner MS, Grey M. An internet coping skills training program for youth with type 1 diabetes: six-month outcomes. Nurs Res NIH Public Access; 2012;61(6):395.
40. Kumah-Crystal YA, Hood KK, Ho Y-X, Lybarger CK, O'Connor BH, Rothman RL, Mulvaney SA. Technology use for diabetes problem solving in adolescents with type 1 diabetes: relationship to glycemic control. Diabetes Technol Ther Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA; 2015;17(7):449–454.
41. Probst T, Pryss R, Langguth B, Schlee W. Emotional states as mediators between tinnitus loudness and tinnitus distress in daily life: Results from the "TrackYourTinnitus" application. Sci Rep Nature Publishing Group; 2016;6(1):1–8.

42. Zhang P, Rodriguez-Cancio M, Schmidt DC, White J, Dennis T. Discussions of a Preliminary Hand Hygiene Compliance Monitoring Application-as-a-service. HEALTHINF 2017. p. 537–544.
43. Zhang P, White J, Schmidt D, Dennis T. Applying machine learning methods to predict hand hygiene compliance characteristics. 2017 IEEE EMBS Int Conf Biomed Heal Informatics IEEE; 2017. p. 353–356.
44. Farooq M, Sazonov E. Accelerometer-based detection of food intake in free-living individuals. IEEE Sens J IEEE; 2018;18(9):3752–3758.
45. Samadi S, Rashid M, Turksoy K, Feng J, Hajizadeh I, Hobbs N, Lazaro C, Sevil M, Littlejohn E, Cinar A. Automatic detection and estimation of unannounced meals for multivariable artificial pancreas system. Diabetes Technol Ther Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA; 2018;20(3):235–246.
46. Gimpel H, Regal C, Schmidt M. myStress: Unobtrusive smartphone-based stress detection. 2015;
47. Asselbergs J, Ruwaard J, Ejdys M, Schrader N, Sijbrandij M, Riper H. Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an explorative study. J Med Internet Res JMIR Publications Inc., Toronto, Canada; 2016;18(3):e5505.
48. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, Mendes GA, Lau AYS. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Informatics Assoc Oxford University Press; 2015;22(1):243–256.
49. Merwin RM, Dmitrieva NO, Honeycutt LK, Moskovich AA, Lane JD, Zucker NL, Surwit RS, Feinglos M, Kuo J. Momentary predictors of insulin restriction among adults with type 1 diabetes and eating disorder symptomatology. Diabetes Care Am Diabetes Assoc; 2015;38(11):2025–2032.
50. Mulvaney SA, Rothman RL, Dietrich MS, Wallston KA, Grove E, Elasy TA, Johnson KB. Using mobile phones to measure adolescent diabetes adherence. Heal Psychol American Psychological Association; 2012;31(1):43.
51. Merwin RM, Moskovich AA, Honeycutt LK, Lane JD, Feinglos M, Surwit RS, Zucker NL, Dmitrieva NO, Babyak MA, Batchelder H. Time of Day when Type 1 Diabetes Patients with Eating Disorder Symptomatology Most Commonly Restrict Insulin. Psychosom Med NIH Public Access; 2018;80(2):222.
52. Georga EI, Protopappas VC, Fotiadis DI. Glucose prediction in type 1 and type 2 diabetic patients using data driven techniques. Knowledge-oriented Appl data Min InTech; 2011;277–296.
53. Wang Y, Wu X, Mo X. A novel adaptive-weighted-average framework for blood glucose prediction. Diabetes Technol Ther Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA; 2013;15(10):792–801.
54. Ståhl F. Diabetes mellitus glucose prediction by linear and Bayesian ensemble modeling. Lund University; 2012.
55. Bondia J, Tarín C, García-Gabin W, Esteve E, Fernández-Real JM, Ricart W, Vehí J. Using support vector machines to detect therapeutically incorrect measurements by the MiniMed CGMS®. J Diabetes Sci Technol SAGE Publications; 2008;2(4):622–629.

56. Sudharsan B, Peeples M, Shomali M. Hypoglycemia prediction using machine learning models for patients with type 2 diabetes. J Diabetes Sci Technol SAGE Publications Sage CA: Los Angeles, CA; 2014;9(1):86–90.
57. Biester T, Kordonouri O, Holder M, Remus K, Kieninger-Baum D, Wadien T, Danne T. "Let the algorithm do the work": reduction of hypoglycemia using sensor-augmented pump therapy with predictive insulin suspension (SmartGuard) in pediatric type 1 diabetes patients. Diabetes Technol Ther Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA; 2017;19(3):173–182.
58. Bastani M. Model-free intelligent diabetes management using machine learning. 2014;
59. Kawano Y, Yanai K. Real-time mobile food recognition system. Proc IEEE Conf Comput Vis Pattern Recognit Work 2013. p. 1–7.
60. Ellis K, Kerr J, Godbole S, Lanckriet G, Wing D, Marshall S. A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. Physiol Meas IOP Publishing; 2014;35(11):2191.
61. Lansing AH, Berg CA, Butner J, Wiebe DJ. Self-control, daily negative affect, and blood glucose control in adolescents with Type 1 diabetes. Heal Psychol American Psychological Association; 2016;35(7):643.
62. Pugach O, Hedeker D, Richmond MJ, Sokolovsky A, Mermelstein R. Modeling mood variation and covariation among adolescent smokers: application of a bivariate location-scale mixed-effects model. Nicotine Tob Res Society for Research on Nicotine and Tobacco; 2013;16(Suppl_2):S151–S158.
63. Schabert J, Browne JL, Mosely K, Speight J. Social stigma in diabetes. Patient-Patient-Centered Outcomes Res Springer; 2013;6(1):1–10.
64. Cai RA, Beste D, Chaplin H, Varakliotis S, Suffield L, Josephs F, Sen D, Wedderburn LR, Ioannou Y, Hailes S. Developing and evaluating JIApp: acceptability and usability of a smartphone app system to improve self-management in young people with juvenile idiopathic arthritis. JMIR mHealth uHealth JMIR Publications Inc., Toronto, Canada; 2017;5(8):e7229.
65. Li T, Lin G. Examining the role of location-specific associations between ambient air pollutants and adult asthma in the United States. Health Place Elsevier; 2014;25:26–33.

Appendix A – Source Code for Comparing Models of Daily SMBG Frequency

```
def runStratKFold(X, y):
    kfold = StratifiedKFold(n_splits=5, shuffle=True, random_state=1)
    # enumerate the splits and summarize the distributions
    for train_ix, test_ix in kfold.split(X, y):
        # select rows
        train_X, test_X = X[train_ix], X[test_ix]
        train_y, test_y = y[train_ix], y[test_ix]
        # summarize train and test composition
        train_0, train_1 = len(train_y[train_y==0]), len(train_y[train_y==1])
        test_0, test_1 = len(test_y[test_y==0]), len(test_y[test_y==1])
        print('>Train: 0=%d, 1=%d, Test: 0=%d, 1=%d' % (train_0, train_1, test_0, test_1))

def evaluateKFold(X, y, model):
    cv = KFold(n_splits=10, random_state=1, shuffle=True)

    accs = cross_val_score(model, X, y, scoring='accuracy', cv=cv, n_jobs=-1)
    pres = cross_val_score(model, X, y, scoring='precision', cv=cv, n_jobs=-1)
    recs = cross_val_score(model, X, y, scoring='recall', cv=cv, n_jobs=-1)
    fls = cross_val_score(model, X, y, scoring='f1', cv=cv, n_jobs=-1)
    rocs = cross_val_score(model, X, y, scoring='roc_auc', cv=cv, n_jobs=-1)
    # report performance
    print('Accuracy: %.3f (%.3f)' % (mean(accs), std(accs)))
    print('Precision: %.3f (%.3f)' % (mean(pres), std(pres)))
    print('Recall: %.3f (%.3f)' % (mean(recs), std(recs)))
    print('F1: %.3f (%.3f)' % (mean(fls), std(fls)))


models = [LogisticRegression(),
          RandomForestClassifier(random_state=seed, max_features='auto', n_estimators=50,
                                 max_depth=20, criterion='gini'),
          SVC()]

Xs = {'all': bg_below4.drop('below4', axis=1).columns.values,
      'demo': bg_below4.drop(['day_of_week','weekend','below4'], axis=1).columns.values,
      'time': ['day_of_week','weekend']
     }

for model in models:
    print('Results from model ', str(model))
    for name in Xs:
        print('Results from', name)
        X = bg_below4[Xs[name]]
        y = bg_below4['below4']
        evaluateKFold(X, y, model)
        print('==========================\n\n')
```

Appendix B – Source Code for Comparing Models of SMBG and IA

```
seed = 42
models = {
    'RF_10': RandomForestClassifier(random_state=seed, max_features='auto', n_estimators=10,
                                    max_depth=8, criterion='gini', class_weight='balanced'),
    'RF_50': RandomForestClassifier(random_state=seed, max_features='auto', n_estimators=50,
                                    max_depth=20, criterion='gini', class_weight='balanced'),
    'RF_100': RandomForestClassifier(random_state=seed, max_features='auto', n_estimators=100,
                                     max_depth=8, criterion='gini',class_weight='balanced'),
    'RF_120': RandomForestClassifier(random_state=seed, max_features='auto', n_estimators=120,
                                     max_depth=15, criterion='gini',class_weight='balanced'),
    'SVC': SVC(kernel='rbf', C = 0.1, gamma=1,class_weight='balanced'),
    'KNN_5': KNeighborsClassifier(n_neighbors=5),
    'KNN_7': KNeighborsClassifier(n_neighbors=7),
    'KNN_10': KNeighborsClassifier(n_neighbors=10)
}

def run_multi_model(features, data_scaled, target_name, models, cv, seed=7):
    dfs = {}
    plots = {}

    for name, subset in features.items():
        print('=========>>> Examining feature: ' + name)
        X = data_scaled[subset]
        y = data_scaled[target_name]
        dfs[name], plots[name] = runKFold(X, y, models, cv, seed)
    return dfs, plots


cv_strategies = {'strat_kfold_5': StratifiedKFold(n_splits=5, random_state=seed, shuffle=True),
                 'strat_kfold_7': StratifiedKFold(n_splits=7, random_state=seed, shuffle=True),
                 'strat_kfold_10': StratifiedKFold(n_splits=10, random_state=seed, shuffle=True)}

# Analysis with stratified kfold
for strategy in cv_strategies:
    smbg_dfs, smbg_plots = run_multi_model(features, data_scaled, 'glucose_no', models,
                                           cv_strategies[strategy], seed)
    isln_dfs, isln_plots = run_multi_model(features, data_scaled, 'insulin_no', models,
                                           cv_strategies[strategy], seed)
```