Computers in Biology and Medicine 155 (2023) 106670

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Modeling risk of hypoglycemia during and following physical activity in people with type 1 diabetes using explainable mixed-effects machine learning

Clara Mosquera-Lopez [a,*], Katrina L. Ramsey [b], Valentina Roquemen-Echeverri [a], Peter G. Jacobs [a]

[a] Artificial Intelligence for Medical Systems (AIMS) Lab, Department of Biomedical Engineering, Oregon Health & Science University, Portland, Oregon, USA
[b] Biostatistics and Design Program, Oregon Health & Science University, Portland, Oregon, USA

ARTICLE INFO

Keywords:
Physical activity
Hypoglycemia risk
Type 1 diabetes
Explainable models
Mixed-effects machine learning

ABSTRACT

Background: Physical activity (PA) can cause increased hypoglycemia (glucose <70 mg/dL) risk in people with type 1 diabetes (T1D). We modeled the probability of hypoglycemia during and up to 24 h following PA and identified key factors associated with hypoglycemia risk.

Methods: We leveraged a free-living dataset from Tidepool comprised of glucose measurements, insulin doses, and PA data from 50 individuals with T1D (6448 sessions) for training and validating machine learning models. We also used data from the T1Dexi pilot study that contains glucose management and PA data from 20 individuals with T1D (139 session) for assessing the accuracy of the best performing model on an independent test dataset. We used mixed-effects logistic regression (MELR) and mixed-effects random forest (MERF) to model hypoglycemia risk around PA. We identified risk factors associated with hypoglycemia using odds ratio and partial dependence analysis for the MELR and MERF models, respectively. Prediction accuracy was measured using the area under the receiver operating characteristic curve (AUROC).

Results: The analysis identified risk factors significantly associated with hypoglycemia during and following PA in both MELR and MERF models including glucose and body exposure to insulin at the start of PA, low blood glucose index 24 h prior to PA, and PA intensity and timing. Both models showed overall hypoglycemia risk peaking 1 h after PA and again 5–10 h after PA, which is consistent with the hypoglycemia risk pattern observed in the training dataset. Time following PA impacted hypoglycemia risk differently across different PA types. Accuracy of hypoglycemia prediction using the fixed effects of the MERF model was highest when predicting hypoglycemia during the first hour following the start of PA ($AUROC_{VALIDATION}$ = 0.83 and $AUROC_{TESTING}$ = 0.86) and decreased when predicting hypoglycemia in the 24 h after PA ($AUROC_{VALIDATION}$ = 0.66 and $AUROC_{TESTING}$ = 0.68).

Conclusion: Hypoglycemia risk after the start of PA can be modeled using mixed-effects machine learning to identify key risk factors that may be used within decision support and insulin delivery systems. We published the population-level MERF model online for others to use.

1. Introduction

Physical activity (PA) is associated with many health benefits. For people living with type 1 diabetes (T1D), these benefits include improved glycemic control, blood lipid profile, cardiovascular fitness, muscle strength, and insulin sensitivity [1,2] as well as reduced frequency and severity of other diabetes related complications such as nephropathy, retinopathy, and neuropathy [3]. Several studies have demonstrated that regular PA can significantly reduce hemoglobin A1C (HbA1C) in children and adolescents [4]. And regular PA has been shown to be associated with increased time in target glucose range (70–180 mg/dL) [5] and decreased mortality [6] in adults. However, PA even under tightly controlled laboratory conditions, can result in highly variable glucose responses [1,6–11]; oftentimes leading to increased risk of hypoglycemia (<70 mg/dL) during exercise and also for hours after PA in people with T1D [12–14].

* Corresponding author. 3303 SW Bond Avenue, Portland, OR, 97239, USA.
E-mail address: mosquera@ohsu.edu (C. Mosquera-Lopez).

https://doi.org/10.1016/j.compbiomed.2023.106670
Received 29 August 2022; Received in revised form 19 January 2023; Accepted 10 February 2023
Available online 11 February 2023

Heightened risk of hypoglycemia during PA could be associated with increased muscle glucose uptake through insulin-dependent and insulin-independent mechanisms and limited endogenous glucose production (EGP) for some PA modalities [14–16]. On the other hand, increased risk of late-onset hypoglycemia following PA could be a consequence of changes in insulin sensitivity [17–19] and prolonged exercise sessions that can result in depletion of glycogen stores, which require increasing glucose uptake from the blood to be restored in the longer term [15].

Despite the benefits of PA on overall glucose control, many individuals with T1D do not engage in regular PA because of fear of hypoglycemia. Furthermore, there is insufficient knowledge of hypoglycemia risk factors and glucose management strategies to help people with T1D better manage their glucose during and after exercise [20–24]. Although there exist guidelines on insulin adjustment and carbohydrate consumption to help people with T1D maintain appropriate glucose control and avoid exercise-induced hypoglycemia [2,25], these recommendations are difficult to follow and hypoglycemia still occurs [26]. Thus, there is still a need to understand the variables that can impact the risk of hypoglycemia when performing different forms of PA and to develop new technologies that can predict future hypoglycemia and provide warning and recommendations to help people with T1D more effectively avoid hypoglycemia events.

Recent advances in continuous glucose monitoring (CGM), insulin delivery systems (e.g., insulin pumps and closed-loop systems), and decision support tools may help to improve glycemic control and substantially reduce hypoglycemia during PA in people with T1D [27]. The increasing use of CGM and other smart and connected devices (e.g., insulin pumps and pens, smart watches, and smart rings) among people with T1D has resulted in the collection of large amounts of multimodal free-living glucose management data. These rich datasets can be leveraged to develop data-driven glucose prediction models that may help anticipate and prevent adverse glucose excursions [28–41].

Several special-purpose population-based and personalized machine learning models have been developed to anticipate hypoglycemia events associated with PA from the start of the activity using datasets from relatively small studies and under controlled conditions. Some existing models predict maximum glucose change or minimum glucose during or up to 4 h after the start of exercise using multiple linear regression [42], autoregressive models (AR) [43,44], neural networks [45], and multivariate adaptive regression spline (MARS) [11]. These outcomes are proxies that can be used to anticipate the occurrence of hypoglycemia. Other works aimed at predicting hypoglycemia events using binary classification approaches based on logistic regression (LR) [46], decision trees, and random forest (RF) models [26]. For example, Reddy et al. [26] reported 87% correct classification during aerobic exercise. Tyler et al. [11] reported 81% accuracy during exercise and 56% accuracy up to 4 h after exercise. However, this work used a unique dataset collected in adults with T1D who participated in a four-arm crossover study evaluating insulin pump therapies, in which participants performed identically designed in-clinic aerobic exercise [47]. Because the exercise in this study was performed under highly controlled clinical settings, the accuracy of the developed algorithms may not be as high under real-world conditions.

In this work, we quantify the impact of key factors explaining hypoglycemia risk using explainable machine learning models. These models can be used to predict hourly probability of hypoglycemia up to 24 h after PA. Although it is well established that PA can lead to increased risk of low glucose not only during exercise but up to 24 h after a workout [14,48], to the best of our knowledge, this is the first attempt to model hourly short- and long-term hypoglycemia risk using a large free-living glucose management dataset across different forms of PA. The models described here may ultimately be incorporated into T1D decision support tools to provide recommendations on insulin dosing and meal intake while considering hypoglycemia risk. The models may also be used within the context of automated insulin delivery to modulate insulin infusion rate based on predicted probability of hypoglycemia following exercise events.

2. Materials and methods

2.1. Datasets

Training and validation datasets: We leveraged a free-living dataset from the Tidepool Big Data Donation Program (Tidepool, Palo Alto, CA, USA) to model hypoglycemia risk during and following PA using mixed-effects machine learning (MEML). The dataset contained demographic data, glucose management data including CGM measurements and insulin doses (10636 days; training: 9166 and validation: 1470), and self-reported PA type, duration, timing, and intensity (6448 sessions; training: 5457 and validation: 991). The data were collected from 50 people with T1D on sensor augmented pump and closed-loop therapies. In this work, we only analyzed data associated with PA sessions. Seventy five percent of PA sessions were of type cardio (aerobic) in the training dataset. The last 24–30 days of data from each of the 50 participants were held out for validation. This validation dataset was used to evaluate the accuracy of the trained models and select the best performing model. Data were aggregated from multivendor CGM systems, insulin pumps, smartphones, and wearable sensors through the *Tidepool.org* cloud platform. CGM data were collected at a 5-min sampling period. *Tidepool.org* did not provide information about the devices' vendors or models associated with collected data. Clinical information related to time since T1D diagnosis was provided. Demographic data were limited to age and biological sex. Table 1 presents a summary of demographic and clinical information as well as an overview of the

Table 1
Characteristics of participants (N = 50).

Demographics	
Age, years	38 ± 13
Biological sex, n (%)	Female: 21 (42)
	Male: 23 (46)
	Unknown: 6 (12)
Clinical information	
Duration of diabetes, years	21 ± 13
Glucose control metrics	
% Time below range (<70 mg/dL), median [range]	2.54 [0.05, 16.17]
% Time in range (70–180 mg/dL), median [range]	78.38 [30.09, 97.48]
% Time above range (>180 mg/dL), median [range]	17.81 [0.15, 69.33]
Days of glucose management data per participant, median [range]	TRaining: 136 [49, 700]
	Validation: 30 [24, 30]
Physical activity	
Physical activity sessions per participant, median [range]	Training: 78 [28, 387]
	Validation: 19 [3, 62]

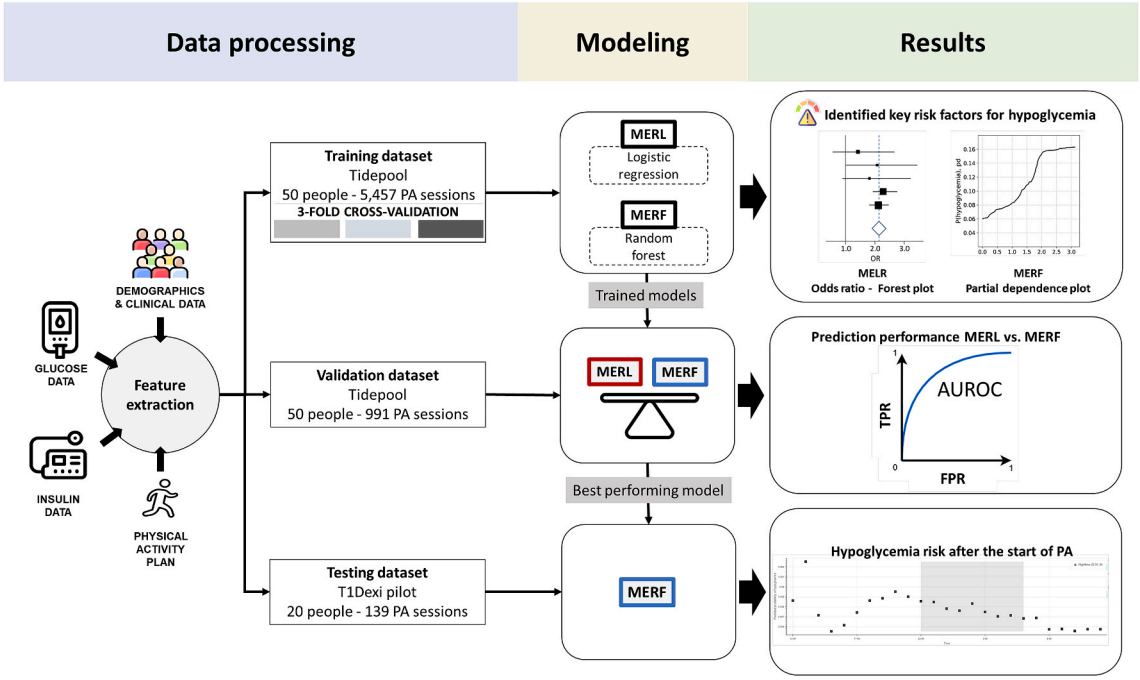

Fig. 1. Methodology workflow.

amount of available data in the training and validation datasets.

Testing dataset: We used an independent test dataset called T1Dexi pilot collected by the T1D Exercise Initiative group (Jaeb Center for Health Research, Tampa, FL, USA) [5,49]. The T1Dexi pilot dataset utilized in this work was collected from 20 people living with T1D on sensor augmented pump therapy or hybrid closed-loop therapy (age 32 ± 13 years, 17 males, 17 ± 14 years since T1D diagnosis). We used a total of 139 sessions of video-guided aerobic, high-intensity interval, and resistance exercise that lasted between 20 and 30 min to assess the generalization performance of the best performing population model (i.e., fixed-effects part of the mixed-effect model that yielded the best AUROC on the validation dataset) on a separated dataset that was not used for model development.

2.2. Hypoglycemia risk modeling approach

We used MEML [50] as a general framework for longitudinal supervised learning to model hourly hypoglycemia risk around PA (see Fig. 1). For each participant i, session j, and time t after the start of a given PA session, we observed a feature vector $x^d_{i,j,t}$ of fixed-effect covariates (i.e., the features x^d derived from demographics, clinical information on duration of diabetes, PA data, and glucose management data), a vector $z_{i,j}$ of random-effects (i.e., participant- and session-varying effects), and a binary response variable $y_{i,j,t}$ that indicates the occurrence of hypoglycemia (sensor glucose <70 mg/dL) measured from the start of PA at $t = 0$ up to $t = 24$ hours after PA or until a new PA session was reported within the 24-h analysis period. The reason for splitting the 24-h analysis time frame into 1-h segments is that the majority of reported PA sessions lasted 1 h or less. Conditioned on the random effects $z_{i,j}$, the mixed-effects MEML framework assumes that $y_{i,j,t}$ follows a distribution from the exponential family with expected value defined as follows [50]:

$$\mathrm{E}\left[y_{i,j,t} \middle| z_{i,j}\right] = g^{-1}\left(f\left(x^d_{i,j,t}, t\right) + Bz_{i,j}\right) \quad \text{Equation 1}$$

In Equation (1), $\mathrm{E}[y_{i,j,t} | z_{i,j}]$ is the expected value of the binary response $y_{i,j,t}$ given the random effects $z_{i,j}$; $f(\cdot)$ is a function that captures the relationship of fixed effects $x^d_{i,j,t}$ and t with the response variable $y_{i,j,t}$; B are the random-effects parameters; and $g(\cdot)$ is a monotonic differentiable pre-specified link function. For a binary response variable, $g(\cdot)$ is the logit link function presented in Equation (2).

$$g(p) = \mathrm{logit}(p) = \log\left(\frac{p}{1-p}\right) \quad \text{Equation 2}$$

In Equation (2), p is the probability of hypoglycemia, which was calculated to be 8.9% in the training set and 9.2% in the validation set across the entire datasets.

We chose the MEML modeling framework for this work over traditional machine learning methods because MEML allows us to include random effects within the models. MEML models, therefore, are trained to account for the dependencies between observations, which result in models that are accurate and robust to the variability introduced by correlated data (e.g., multiple measurements from the same participant). We trained a mixed-effects logistic regression (MELR) whereby $f(x^d_{i,j,t}, t) = A[x^d_{i,j,t}, t]$ is a linear combination of the input variables, and a mixed-effects random forest (MERF) whereby $f(x^d_{i,j,t}, t)$ is a purely data-driven non-linear function of fixed-effect covariates approximated using a RF model. The parameters of the MERF model were estimated using an iterative method described by Pellagatti et al. [51] that alternates, until convergence, the training of the RF that models the fixed effects, with the estimation of the random effects. We used participant-level three--fold cross-validation to select the hyperparameters of the RF that yielded the lowest cross-validation error. We ensured that data from a given participant was not simultaneously used for model training and validation in a given cross-validation iteration.

Feature extraction: A total of $D = 17$ features were derived from demographic and clinical information, CGM and insulin doses collected over the 24 h preceding PA, and information about the type, timing, duration, and intensity of PA. We studied the relationship between input features using correlation analysis to ensure highly correlated variables ($\rho \geq 0.6$) providing possibly redundant information were not included in the models. The analysis showed that there was moderate positive correlation ($\rho = 0.59$) between age (DEMOGRAPHICS.AGE) and the number of years living with diabetes (CLINICAL.YEARST1D); and a weak negative correlation ($\rho = -0.29$) between DEMOGRAPHICS.AGE

Table 2
Data features obtained from demographic, clinical, glucose management, and physical activity data.

Feature	Label	Name	Description
x^1	DEMOGRAPHICS.AGE	Age	Age in years
x^2	CLINICAL.YEARST1D	Years living with T1D	Time since T1D diagnosis in years
x^3	GLUCOSE.CV.24H.PRIOR	Glucose variability 24 h prior	Coefficient of variation of glucose data over the 24 h preceding PA
x^4	GLUCOSE.LBGI.24H.PRIOR	Low blood glucose index 24 h prior	Hypoglycemia risk estimated using the low blood glucose index (LBGI) [52], which increases with the frequency and duration of hypoglycemia excursions
x^5	GLUCOSE.EXCURSION.2H.PRIOR	Glucose excursion 2 h prior	Maximum excursion (used as a proxy for carbohydrate intake) in the 2 h prior PA with reference to 24-h mean glucose
x^6	GLUCOSE.TREND.30MIN.PRIOR	Glucose rate of change over last 30 min prior	Average glucose trend in the 30 min preceding the start of PA
x^7	GLUCOSE.CGM.START	Glucose at the start of PA	Sensor glucose at the start of PA
x^8	INSULIN.IOBTOTDIR.START	Ratio of insulin on board over total daily insulin requirement at the start of PA	Exposure to insulin calculated as the insulin on board at the start of PA to total daily insulin requirement estimated as the average insulin dosed over the preceding study days
x^{9-12}	PA.TYPE	Physical activity type	One-hot encoded feature to represent the different types of PA (cardio, strength, mixed, and other)
x^{13-14}	PA.HOUROFDAY	Physical activity time	The hour at the start of PA coded as $\cos(2\pi\frac{hour}{24}), \sin(2\pi\frac{hour}{24})$
x^{15}	PA.INTENSITY	Physical activity intensity	The intensity of PA in kilocalories per minute
x^{16}	PA.DURATION	Physical activity duration	Duration of PA session in minutes
x^{17}	PA.TIMESINCEPA	Time since previous PA session	Time since last reported PA session in hours
t	TIME.FROMSTART	Time from the start of PA in hours	$t \in \{0, 1, 2, \ldots 24\}$

Insulin on board (IOB) is estimated as the weighted sum of past insulin boluses with peak effect of 1 h after 30 min of injection and subsequent exponential decay [39, 53].

and reported PA intensity (PA.INTENSITY) indicating that older people may choose to exercise at a lower intensity.

The time variable t represents the hour following the start of a PA event when hypoglycemia is to be predicted and can range from 0 to 24. This variable was included as a fixed effect to model how hypoglycemia risk changes hourly across the 24 h following PA.

The dimension of the models' fixed effects depends on which features (and interactions) are included, and the method used to encode the time variable in MELR and MERF models. Table 2 shows a descriptive summary of the features used in this study.

Mixed-effects logistic regression: The MELR approach results in a directly interpretable model in which the estimated coefficients (i.e., odds ratio) represent the association of input features with hypoglycemia risk. The variable CLINICAL.YEARST1D was not included in the MELR model since it was correlated with DEMOGRAPHICS.AGE and had lower predictive power. DEMOGRAPHICS.AGE and GLUCOSE.CGM.START were mean centered by subtracting their mean across the entire training dataset from the observed feature value. The mean-centered version of GLUCOSE.CGM.START was scaled by 50 mg/dL (C.S50.GLUCOSE.CGM.START); therefore, the resulting odds ratio for the linear term is associated with 50 mg/dL difference from the overall mean of the glucose at the start of PA. PA.DURATION was scaled by 10 min (S10.PA.DURATION); thus, the estimated odds ratio for PA duration is associated with 10-min changes in the value of the variable.

Since hypoglycemia risk may change in a nonlinear way during hours 0–24 following exercise, the time variable t was coded with linear splines to allow the slope to change at 1 h, 3 h, and 6 h after PA. The coefficient associated with t^0 is the slope in the hour after the end of PA; and the coefficients associated with t^1, t^3, and t^6 correspond to the slopes from one to 3 h, three to 6 h, and six to 24 h following PA, respectively. Note that the 24-h time frame following the start of exercise is analyzed hourly. However, the time variable is transformed using splines to better model the nonlinear association between hypoglycemia risk and time following PA observed in the training dataset (see supplementary Table ST1 for details on how the time variable is transformed using linear splines and Supplementary Table ST2 for details on how the rate of hypoglycemia changed in the training dataset during the 24 h following PA).

For the MELR model, in addition to the transformations done on the input features such as scaling and squaring, we evaluated interactions between variables. The interaction between different types of PA and the time from the start of the exercise event was of particular interest in this study to understand how the risk of hypoglycemia changed over time across different PA modalities. The MELR model is presented in Equation (3).

$$\begin{aligned}\text{Logit}(p) = {} & a_0 + a_1x^1 + a_2x^3 + a_3x^4 + a_4\left(x^4\right)^2 + a_5x^5 + a_6x^6 + a_7x^7 \\ & + a_8\left(x^7\right)^2 + a_9x^8 + a_{10}x^{10} + a_{11}x^{11} + a_{12}x^{12} + a_{13}x^{13} + a_{14}x^{14} \\ & + a_{15}x^{15} + a_{16}\left(x^{15}\right)^2 + a_{17}x^{16} + a_{18}x^{17} + a_{19}t^0 + a_{20}t^1 + a_{21}t^3 \\ & + a_{22}t^6 + a_{23}x^{10}t^0 + a_{24}x^{10}t^1 + a_{25}x^{10}t^3 + a_{26}x^{10}t^6 + Bz + \varepsilon\end{aligned}$$

Equation 3

In Equation (3), a_k are the MELR model coefficients; x^d are the input features described in Table 2; t^0, t^1, t^3, and t^6 are the components of a 4-element code that represent the time in hours after PA; and B are the parameters associated with the random effects z; and ε are the model residuals. Note that $z \sim N(0, \sigma_z)$, $\varepsilon \sim N(0, \sigma_\varepsilon)$ and $\text{Cov}(z, \varepsilon) = 0$.

Mixed-effects random forest: We used the raw input features as described in Table 2 to model the probability of hypoglycemia in the 0–24 h after the start of PA using the MERF modeling approach (Equation (4)).

$$\text{logit}(p) = RF\left(x^d_{i,j,t}, t\right) + Bz + \varepsilon \qquad \text{Equation 4}$$

In Equation (4), the RF function is obtained using a random forest algorithm; the time t is a single variable that represents the hour after the start of PA when a prediction of hypoglycemia probability will be done. For example, a value of $t = 0$ is used to predict hypoglycemia during the first hour following the start of PA while a value of $t = 20$ is used to predict hypoglycemia during the 20th following PA. As in Equation (3), B are the parameters associated with random effects z and ε are the model residuals.

The RF is an ensemble aggregation method that combines the predictions of a given number of decision trees [54]. An advantage of RF is that it has more flexibility than logistic regression to model non-linear relationships between input features and the target variable (i.e., hypoglycemia risk) without knowing the functional form of these relationships. For example, the RF does not need to use splines to model the non-linear effect of time on hypoglycemia risk as it was done with the MELR method. The hyperparameters of the RF include the number of trees in the forest, the maximum bootstrap sample size to train each base decision tree, the minimum number of samples at a leaf node, and the

Table 3
Mixed-effects logistic regression model estimates results.

Fixed effects			
Variable	Odds ratio estimate	P-value	Adjusted P-value
C.DEMOGRAPHICS.AGE	0.995	0.528	0.989
GLUCOSE.CV.24H.PRIOR	1.122	0.707	0.997
GLUCOSE.LBGI.24H.PRIOR	**1.432**	**<0.001**	**<0.001**
(GLUCOSE.LBGI.24H.PRIOR)^2	**0.967**	**<0.001**	**<0.001**
GLUCOSE.EXCURSION.2H.PRIOR	1.507	<0.001	<0.001
GLUCOSE.TREND.30MIN.PRIOR	0.991	0.708	0.997
C.S50.GLUCOSE.CGM.START	**0.772**	**<0.001**	**<0.001**
(C.S50.GLUCOSE.CGM.START)^2	**1.046**	**<0.001**	**<0.001**
INSULINTOTDIR.IOB.START	**3.625**	**0.001**	**0.014**
PA.TYPE.STRENGTH	0.930	0.741	0.997
PA.TYPE.MIXED	1.143	0.159	0.790
PA.TYPE.OTHER	0.919	0.419	0.978
PA.HOUROFDAY.SIN	0.932	0.053	0.480
PA.HOUROFDAY.COS	0.981	0.696	0.997
PA.INTENSITY	**1.101**	**<0.001**	**<0.001**
(PA.INTENSITY)^2	**0.996**	**0.001**	**0.014**
S10.PA.DURATION	1.029	<0.001	<0.001
PA.TIMESINCEPA	0.999	0.169	0.790
TIME.FROMSTARTPA.T0	**2.361**	**<0.001**	**<0.001**
TIME.FROMSTARTPA.T1	**0.583**	**<0.001**	**<0.001**
TIME.FROMSTARTPA.T3	**1.229**	**<0.001**	**<0.001**
TIME.FROMSTARTPA.T6	**0.976**	**<0.001**	**<0.001**
PA.TYPE.STRENGTH·TIME.FROMSTARTPA.T0	0.607	0.064	0.517
PA.TYPE.STRENGTH·TIME.FROMSTARTPA.T1	1.583	<0.001	<0.001
PA.TYPE.STRENGTH·TIME.FROMSTARTPA.T3	0.907	0.106	0.674
PA.TYPE.STRENGTH·TIME.FROMSTARTPA.T6	1.002	0.807	0.997
Random effects			
Variable			Variance
PARTICIPANT			0.414
SESSION			0.849

C.* identifies variables that are mean centered. C.S50.GLUCOSE.CGM.START is the mean-centered glucose at the start of exercise scaled by 50 mg/dL. S10.PA.DURATION is PA duration scaled by 10 min. Bolded variables were found to be significant in MELR and MERF models. P-values adjusted using the Holm-Sidak method [58].

portion of features to consider for node splits. The best RF hyperparameters were found to be the following: number of trees = 400, maximum bootstrap sample size = 20% of total number of examples, and minimum number of samples per leaf = 150, and the portion of features to consider for node splits = 33%.

We estimated the significance of the input features using the permutation importance approach described by Altman et al. [55] This method is based on repeated permutations of the response variable for estimating the distribution of measured importance for each input feature in a non-informative setting. The P-value of observed importance values provides a tool for assessing statistical significance of input features. In this work, we performed 500 permutation iterations.

We quantified the impact of fixed-effects variables of interest on hypoglycemia risk using partial dependence analysis [56]. Partial dependence analysis is conducted here to achieve explainability for the MERF model. Partial dependence pd_{X^S} is the marginal predicted effect of the features of interest on probability of hypoglycemia, and it is estimated using Equation (5).

$$pd_{X^S}\left(x^S\right) = \frac{1}{N}\sum_{n=1}^{N} RF\left(x^S, x^{C(n)}\right) \qquad \text{Equation 5}$$

In Equation (5), X^S represents the subset of variables whose target effect is of interest, X^C is the set of all other fixed-effects variables, and N is the total number of samples in the training dataset.

2.3. Model accuracy assessment metrics

We used log-likelihood to compare the quality of MELR and MERF models fit to training data, and the area under the receiver operating characteristic curve (AUROC) to assess the predictive accuracy of the developed models on the validation and testing datasets. From the training ROC curve, we selected the detection thresholds P_{TH} that maximized the harmonic mean of sensitivity and specificity [57] for three pre-specified time frames after a PA event (i.e., first hour after the start of PA, 1–4 h post-exercise, 5–24 post-exercise). Using the selected thresholds, we calculated sensitivity, specificity, and balanced accuracy (i.e., average of the sensitivity and specificity) on the validation dataset and an independent testing dataset.

3. Results

3.1. Mixed-effects logistic regression model

The estimated odds ratios for the MELR model are presented in Table 3. Key risk factors identified by the logistic regression model included hypoglycemia risk over the 24 h preceding exercise, high glucose excursion in the 2 h prior to the start of PA, glucose and body exposure to insulin (i.e., ratio of insulin on board to total daily insulin requirement) at the start of PA, and PA intensity and duration. Regarding time effects, the MELR model showed highest risk of hypoglycemia in the first hour after PA. There is a difference in time effects between strength and other types of activities that is statistically significant between 1 and 3 h post PA, which leads to a more prominent peak for PA of type strength between 5 and 10 h after PA.

3.2. Mixed-effects random forest

Table 4 shows the normalized variable importance for the MERF model and its statistical significance. Variable importance Δacc_v is a measure of how much the accuracy of a model is affected when randomly permuting a variable or a set of variables v in the training

Table 4
Estimated variable importance for the MERF model.

Fixed effects			
Variable	Normalized variable importance	P-value	Adjusted P-value
DEMOGRAPHICS.AGE	0.088	<0.001	<0.001
CLINICAL.YEARST1D	0.058	<0.001	<0.001
GLUCOSE.CV.24H.PRIOR	0.086	<0.001	<0.001
GLUCOSE.LBGI.24H.PRIOR	**1.000**	**<0.001**	**<0.001**
GLUCOSE.EXCURSION.2H.PRIOR	0.040	0.962	0.999
GLUCOSE.TREND.30MIN.PRIOR	0.075	<0.001	<0.001
GLUCOSE.CGM.START	**0.205**	**<0.001**	**<0.001**
INSULIN.IOBTOTDIR.START	0.026	0.999	1.000
{GLUCOSE.CGM.START, INSULIN.IOBTOTDIR.START}	**0.227**	**<0.001**	**<0.001**
PA.TYPE	0.004	0.835	0.999
PA.HOUROFDAY	**0.063**	**<0.001**	**<0.001**
PA.INTENSITY	**0.100**	**<0.001**	**<0.001**
PA.DURATION	0.036	0.997	0.999
PA.TIMESINCEPA	0.030	0.999	1.000
TIME.FROMSTARTPA	**0.126**	**<0.001**	**<0.001**
{PA.TYPE, TIME.FROMSTARTPA}	**0.128**	**<0.001**	**<0.001**
Random effects			
Variable			Variance
PARTICIPANT			0.632
SESSION			0.822

Bolded variables were found to be significant in MELR and MERF models. P-values adjusted using the Holm-Sidak method [58].

dataset. We calculated normalized variable importance as the ratio $\frac{\Delta acc_v}{\max_v \Delta acc_v}$. In the MERF model, the most important feature is the exposure to low glucose levels over the 24 h preceding PA (GLUCOSE.LBGI.24H.PRIOR) and its normalized variable importance is 1. Other risk factors include glucose at the start of PA, body exposure to insulin at the start of PA, glucose variability 24 h prior to PA, average glucose trend in the 30 min prior to PA, age, duration of diabetes, and PA timing and intensity. The impact of relevant fixed-effects variables on hypoglycemia risk was quantified using the partial dependence analysis discussed hereinafter.

Partial dependence analysis: Fig. 2(a)-(d) show 1D partial dependence plots representing the marginal effect of the most relevant risk factors on probability of hypoglycemia predicted by the MERF model. From these figures we see that hypoglycemia risk nearly triples with higher LBGI in the 24-h preceding exercise (Fig. 2(a)). Hypoglycemia risk also increases for lower glucose values at the start of exercise (Fig. 2(b)), for higher exercise intensities (Fig. 2(c)), and in the evenings (Fig. 2(d)). Fig. 2(e)–(h) show 2D partial dependence plots of the marginal effect of risk factors and the time since the start of PA. Fig. 2(i)–(l) show the combined effect of glucose and exposure to insulin at the start of PA on predicted probability of hypoglycemia during and several hours post PA.

3.3. Comparative MELR and MERF accuracy assessment

Table 5 shows comparative summary metrics for the fit of MELR and MERF models to training data, as well as their prediction accuracy on the Tidepool training and validation datasets. Overall, the MELR and MERF models performed comparatively in terms of quality of model fit to training data and prediction accuracy. The fixed-effects part of the MERF model slightly outperformed that of the MELR model in terms of $AUROC_{FIXED}$ in the training set ($AUROC_{FIXED\text{-}MERF} = 0.71$ vs. $AUROC_{FIXED\text{-}MELR} = 0.65$) and validation set ($AUROC_{FIXED\text{-}MERF} = 0.66$ vs. $AUROC_{FIXED\text{-}MELR} = 0.63$). Also, we note that the random effects play a crucial role in the accuracy of both models. When random effects are included in the prediction, the accuracy of predicting hypoglycemia improves to $AUROC_{MIXED} = 0.83$. This suggests that if the factors associated with random effects (e.g., demographics or other person-specific characteristics) can be modeled using additional variables, then prediction accuracy may be substantially improved. Other known risk factors that may impact hypoglycemia risk during and after PA include nutrient intake before and after exercise, aerobic fitness, presence of comorbidities, stress, and anxiety [11,59]. It may be possible for such factors to be learned over time using personalized machine learning, which may then lead to increased prediction accuracy.

Because the overall performance of the fixed-effects part of the MERF model is better than the fixed-effects part of the MELR model in the validation dataset, the MERF model is determined to be the best choice for a prediction method for hypoglycemia risks associated with PA events. Therefore, we further evaluated the generalizability of the fixed-effects part of the MERF model on the test dataset that was not used for training the model. We further integrated the fixed-effects part of the MERF algorithm into a publicly available online browser-based tool for hypoglycemia risk assessment as further described below.

3.4. Prediction accuracy of the fixed-effects part of the MERF model on the validation dataset and independent test dataset

Table 6 shows the hypoglycemia prediction accuracy of the fixed-effects part of the MERF model or the fixed-effects RF (FERF) on the Tidepool validation dataset and on the independent T1Dexi pilot test dataset. We report AUROC and hypoglycemia prediction accuracy for the time intervals corresponding to the first hour from the start of PA, 1–4 h post PA, and 5–24 h post PA. Hourly AUROC results are also shown in supplementary Table ST2.

3.5. Online browser-based tool for hypoglycemia risk assessment using the developed fixed-effects random forest model

We developed an online browser-based tool for providing an objective hourly hypoglycemia risk score for PA decision support leveraging the fixed-effects part of the MERF model (i.e., the RF model). The development of the web tool was done with Bokeh [60], a python library for creating interactive visualizations for web browsers. Our tool allows users to interactively change the value of hypoglycemia risk factors to predict the hourly probability of hypoglycemia using the population-level MERF model. Predictions are shown for up to 24 h after performing the indicated type of PA. A prediction with a green marker indicates that the predicted probability of hypoglycemia is lower than the mean predicted hypoglycemia in the training set used to fit the RF model for the given hour. On the other hand, a red marker indicates that the predicted probability of hypoglycemia is higher or equal to the mean

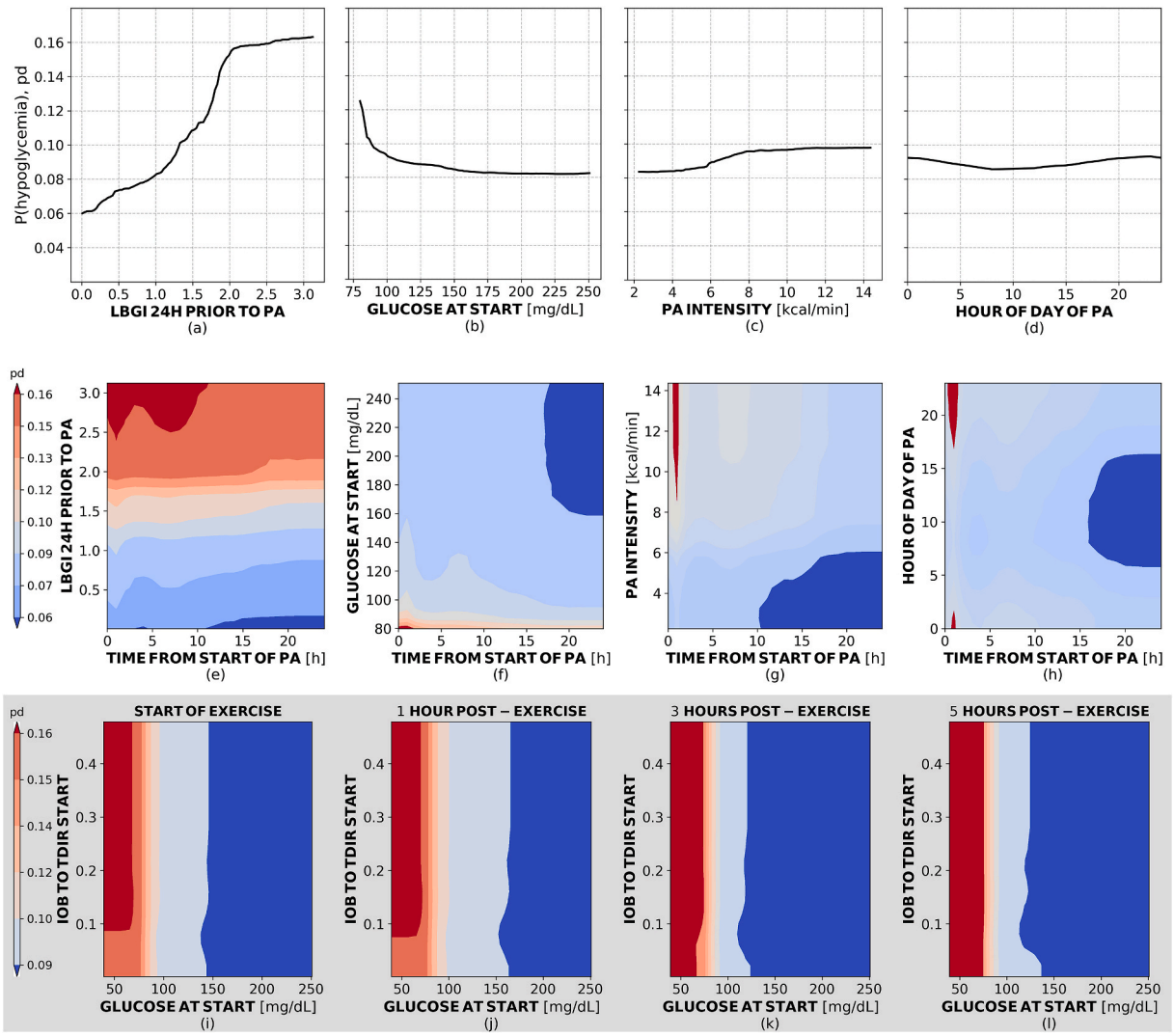

Fig. 2. Partial dependence plots for the MERF model. (a)–(d) are partial dependence plots of individual effect of the most relevant variables GLUCOSE.LBGI.24H. PRIOR, GLUCOSE.CGM.START, PA.INTENSITY, and PA.HOUROFDAY on predicted probability of hypoglycemia. (e)–(h) are 2D partial dependence plots of most relevant risk factors during physical and up to 24 h after the end of PA. (i)–(l) are the partial dependence plots showing the combined effect of GLUCOSE.CGM.START and INSULIN.IOBTOTDIR.START during PA and one, three, and 5 h post-exercise. Note that although INSULIN.IOBTOTDIR.START is not one of the most relevant factors identified by the MERF model, larger values of this variable are associated with higher probability of hypoglycemia during PA and during the first few hours immediately after exercise, particularly at lower glucose levels.

predicted hypoglycemia in the training set for the given hour. Nighttime, defined as starting at 22:00 and duration of 8 h, is highlighted within a gray box (see Fig. 3). This tool is publicly available and can be accessed from https://clara.mosqueralopez.com/pahypoglycemiarisk.

4. Discussion

This study uses tools from mixed effects linear regression and mixed effects random forests to identify the most relevant risk factors associated with hypoglycemia during and following PA that can be derived from CGM, insulin, and PA data. Several of the most significant factors associated with heightened risk of hypoglycemia identified by both MELR and MERF models may explain the increased probability of hypoglycemia during or immediately after the end of PA. The variables that are most relevant to increased hypoglycemia risk are the glucose at the start of exercise and the body exposure to insulin (calculated as the ratio of insulin on board to total daily insulin requirement) at the start of exercise. These two variables reflect the combined impact of individual's metabolic state and body exposure to insulin at the start of PA [42]. The estimated odds ratio in the MELR model indicate that the odds of experiencing post-exercise hypoglycemia more than tripled (3.6x) relative to the ratio of IOB to total daily insulin. As an example, a person with a total daily insulin requirement of 40 units could have an IOB of 7 units. By decreasing their IOB by 5 units prior to exercise they could reduce their risk of hypoglycemia by 45% (i.e., $3.6 \times \frac{5\ u}{40\ u} \times 100 = 45\%$) Conversely, for a 50 mg/dL increase in glucose at the start of PA from the mean (i.e., 148 mg/dL calculated from the training dataset), there is a 23% decrease in the odds of experiencing a hypoglycemia event over the 24 h post-exercise period. In the MERF model, body exposure to insulin was not significantly related to increased hypoglycemia risk by itself, but when combined with glucose at the start of exercise, it was associated with higher predicted probability of hypoglycemia both during exercise and up to 3 h post-exercise (see Fig. 2(i)–(l)).

Another relevant risk factor and predictor of hypoglycemia around PA is the exposure to low glucose levels (i.e., LBGI) over the 24-h period preceding PA. LBGI was highly related to exercise-induced hypoglycemia both in the short term and beyond 4 h after the end of PA (Fig. 2(e)). Increased LBGI can nearly double the marginal modeled probability of hypoglycemia from the average of 8.9% in the training set to 16%.

The MELR and MERF models described here showed overall

Table 5
Comparative accuracy analysis of MELR and MERF models on training and validation data sets.

Metric	Dataset	Model							
		MELR				MERF			
		Fixed effects	Mixed effects (participant-level random effects)		Mixed effects (participant- and session level random effects)	Fixed effects	Mixed effects (participant-level random effects)		Mixed effects (participant- and session level random effects)
Log-likelihood	Train		−27270.1		−26057.7		−27219.8		−26340.2
AUROC	Train	0.65	0.69		0.83	0.71	0.69		0.83
	Validation	0.63	0.68		0.83	0.66	0.68		0.83

hypoglycemia risk peaking 1 h following exercise and again 5–10 h after PA. This result is consistent with the distribution of the hourly rate of hypoglycemia observed in the training dataset (see mean hourly hypoglycemia presented in Supplementary Table ST2). The delayed peak in hypoglycemia risk may be associated with previous hypoglycemia over the 24 h prior to the start of PA and the glucose level at the start of exercise (see Fig. 2(e) and (f)). Based on the partial dependence analysis, the hypoglycemia risk might be mitigated by avoiding PA if an individual with T1D has experienced long periods with low glucose levels or multiple hypoglycemia events in the 24 h preceding the start of PA and by ensuring a glucose level of 160 mg/dL or higher at the start of PA.

Physical activity intensity and timing are the main exercise-related factors impacting hypoglycemia risk. Intensity levels greater than 6 kcal/min result in heightened risk of hypoglycemia as shown in Fig. 2(c) and (g). The partial dependence plot of exercise timing in Fig. 2(d) and (h) show that the risk of hypoglycemia is reduced when PA is performed in the morning or early afternoon. It is not surprising that hypoglycemia risk was found to be higher when PA was performed later in the day, since the body's exposure to insulin is typically lower in the morning and can increase throughout the day as the person eats. Consensus statements that provide guidance for safe exercise also encourage exercising in the morning when insulin levels are oftentimes low [2,25].

We were not able to demonstrate that the type of PA (i.e., cardio, strength, or mixed) has a significant impact on predicted hypoglycemia risk. However, an interesting finding of this study is that there exists a significant difference in the time effect between strength and other types of activities. Particularly, the MELR model showed that strength PA might have been associated with a protective effect during and immediately after PA. However, strength exercise was associated with a significantly higher risk of late-onset hypoglycemia when compared with other forms of exercise. The estimated odds ratio for the interaction of strength exercise and the slope of the time spline between one to 3 h post-exercise was found to be 1.583 ($P < 0.001$), which indicates a 58% higher likelihood of hypoglycemia for strength exercise after 3 h post-exercise. As discussed by Valli et al. [61], this study could provide support to the hypothesis that while aerobic exercise tends to lead to more hypoglycemia during exercise, it might be recommended for better glycemic control during the hours following exercise.

In terms of model fit and prediction accuracy, MELR and MERF performed comparably. Given the observed variability of random effects, it is likely that a model designed for personalization and adaptation that accounts for inter- and intra-participant variability may yield better prediction accuracy as indicated by the improvement of 26–32% in AUROC when random effects are included in the prediction. Personalization to account for participant-level random effects can improve AUROC by 3–8% relative to $\text{AUROC}_{\text{FIXED}}$.

We found that the MELR and MERF yielded similar prediction accuracy. The main difference between MELR and MERF models is how inputs are processed and the assumptions of each modeling approach. In the MELR model, it is assumed that the fixed effects are linearly related to the logit-transformed probability of hypoglycemia (i.e., $\text{logit}(p)$). Therefore, to model nonlinear relationships, raw input features must be transformed, and all interactions of interest must be specified as model inputs. On the other hand, the function that maps the fixed effects to the target variable in the MERF model is found using an RF ensemble learning algorithm, which does not assume a linear relationship between inputs and output. A RF can seamlessly find important interactions among input features. An advantage of the MELR model is that, compared with the MERF, it can be more easily interpreted because the associations between inputs and output variables can be estimated directly from the model coefficients. However, the MERF modeling approach has the advantage that it does not require substantial preprocessing of the input data and is more flexible in modeling the relationships between input variables and the response variable. The MERF model also had higher prediction accuracy than the MELR model when using only fixed effects. The hypoglycemia prediction accuracy of

Table 6
Prediction accuracy of the fixed-effects random forest (FERF) on the validation and testing datasets.

Time frame	P_{TH} [%]	AUROC		Sensitivity [%]		Specificity [%]		Balanced accuracy [%]	
		Validation dataset	Testing dataset	Validation dataset	Testing dataset	Validation dataset	Testing dataset	Validation dataset	Testing dataset
first hour from the start of PA	9.8	0.83	0.86	81.4	93.3	67.3	58.3	74.4	75.8
1–4 h post PA	8.0	0.66	0.71	70.1	75.0	51.4	56.1	60.7	65.5
5–24 h post PA	7.7	0.66	0.64	59.3	70.8	61.3	59.1	60.3	64.9

P_{TH} values are selected using the ROC curve from the training dataset. The values of sensitivity, specificity, and balanced accuracy can change based on the strategy for selecting an operating point on the ROC curve.
The time frames used to report performance of the MERF model are selected to match those used in previously published works on hypoglycemia prediction following PA.

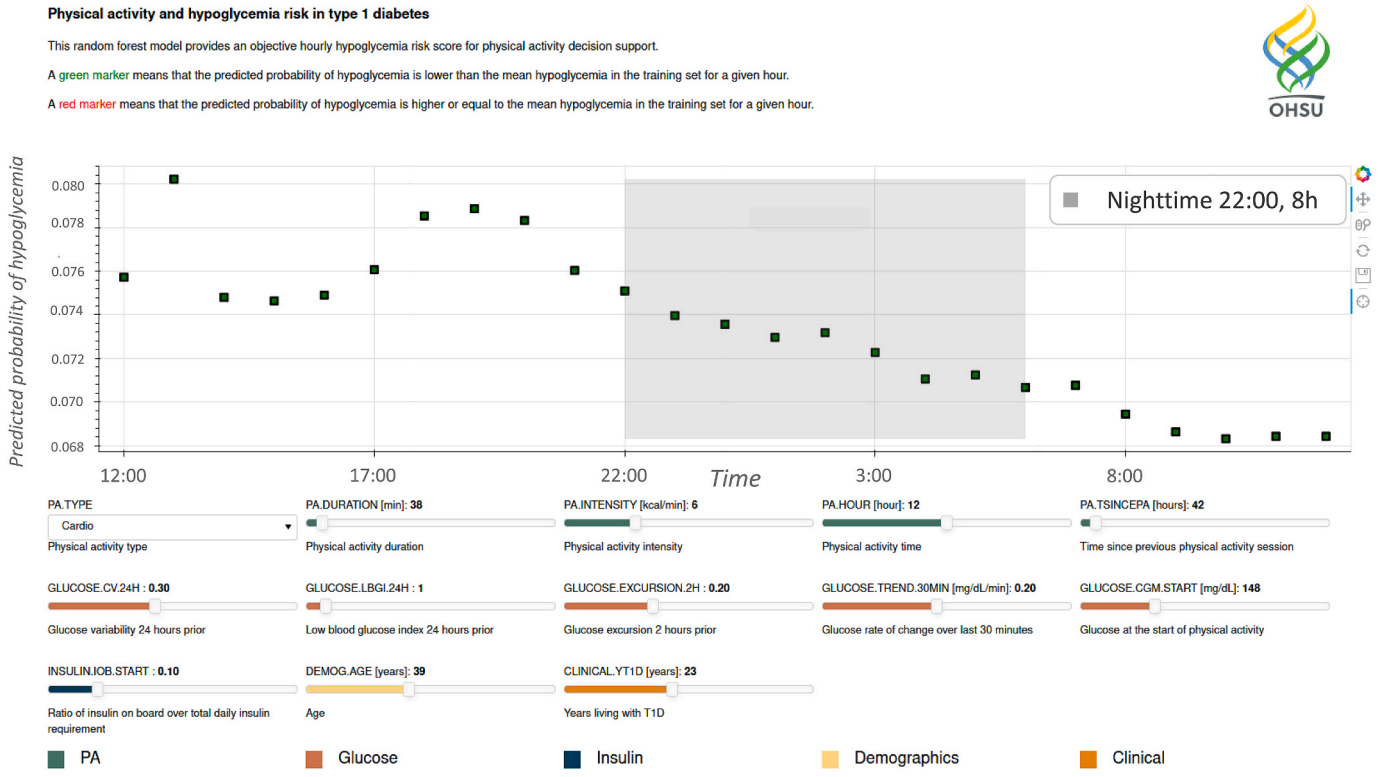

Fig. 3. Screenshot of the web tool for hypoglycemia risk assessment.

the FERF model (see Table 4) is in the same range as those found in prior work using other prediction methods, and the accuracy of our method is higher for an extended prediction horizon of 4 h after the start of PA. To compare the performance of our proposed model with previously published algorithms, we used the reported accuracy by Reddy et al. [26] and Tyler et al. [11] Reddy et al. [26] reported 86.7% accuracy in predicting hypoglycemia during exercise using a RF model, but an extended prediction horizon was not considered. Furthermore, the Reddy et al. model required heart rate data during the first 10 min of exercise as an input to the model. On the other hand, Tyler et al. [11] reported 72%–81% accuracy in predicting hypoglycemia during exercise and 47%–56% accuracy in predicting hypoglycemia 4 h after exercise using MARS, AR, and LR models. Both Reddy et al. and Tyler et al. trained and evaluated models on aerobic exercise done under highly controlled clinical study conditions. Therefore, comparison of the current results with these other studies should be made with caution.

Overall, the fixed-effects RF predicts hypoglycemia with high balanced accuracy over a short-term prediction horizon of 1 h after the start of exercise, but the performance degrades by approximately 14–17% as the prediction horizon is extended (see Table 6). We have made the RF model publicly available as part of this manuscript. This model can be potentially used as a population management tool by health care providers, diabetes educators, patients, or incorporated into systems to provide decision support to individuals living with T1D to better manage their glucose around PA or to modulate insulin delivery in automated insulin delivery systems based on an objective metric of exercise-associated hypoglycemia risk.

One limitation of this study is the lack of data from individuals on multiple daily injections therapy as all the training data from the Tidepool data set is from insulin pump or closed-loop users. Also, the Tidepool dataset used for training and validation of the developed models contains glucose data collected using multiple CGM systems. Although the accuracy of commercially available CGM systems to accurately collect and document glucose levels is widely accepted by the clinical community [62], and multiple CGM systems has been found to have comparable clinical accuracy in past comparative studies [63], the variability associated with the sensing device was not included as a random effect in the MELR and MERF models because *Tidepool.org* does not provide information on the manufacturer and model of CGM sensors. Another limitation of the study is that because the dataset is a free-living

dataset, the exercise type, duration, and intensities are all self-reported, meaning that there can be errors in these values. And lastly, because the cohort of people who choose to donate to the Tidepool Big Data Donation Program, additional data must be considered to better represent the population of people living with T1D as a whole. In the future, there is an opportunity to expand this analysis to larger and more diverse datasets.

5. Conclusions

We modeled short- and long-term post-exercise hypoglycemia risk using a comprehensive mixed-effects machine learning modeling approach. Two explainable models MELR and MERF were fitted using a large free-living glucose management dataset that can be used to predict the probability of hypoglycemia associated with different PA modalities with high accuracy during exercise and with fair performance in the subsequent 24 h post-exercise. We identified key risk factors significantly associated with hypoglycemia including glucose and body exposure to insulin at start of PA, hypoglycemia risk as measured by the low blood glucose index over the 24 h preceding PA, and exercise intensity and timing. Both models showed that hypoglycemia risk is expected to peak 1 h and again between 5-10 h after exercise, with differences in hypoglycemia risk between strength and other types of PA. Results indicate that strength exercises lead to a more prominent peak of late-onset hypoglycemia risk.

We found that random effects, particularly session-level effects, accounted for a large portion of the variance in hypoglycemia risk, confirming that glucose response to PA has high inter- and intra-individual variability that can't be modeled with the fixed effects covariates considered in this study. The significance of random effects indicates that personalization and adaptation may help improve the accuracy of long-term hypoglycemia prediction during and after exercise.

We made the fixed-effects part of the MERF model publicly available online. This model provides prediction of hourly probability of hypoglycemia during and following PA under varied conditions. The goal of this tool is to facilitate population-based exercise management approaches to help prevent post-exercise hypoglycemia in people living with T1D.

Funding

This research was funded by The Oregon Medical Research Foundation and The National Institutes of Health NIH/NIDDK (grant 5R21DK128582-02).

Authors' contributions

C.M.L. and P.G.J. designed the study. C.M.L. processed the datasets and developed analysis tools. C.M.L. and K.L.R. developed and performed formal evaluation of hypoglycemia risk models. C.M.L, K.L.R., and P.G.J. analyzed study results. V.R.E. developed the web tool for exercise-induced hypoglycemia risk prediction that we have made publicly available with this manuscript. C.M.L. wrote the first draft and revisions of the manuscript. K.L.R., V.R.E., and P.G.J. edited the manuscript. C.M.L. acquired funding and administered the project.

Authors' disclosures

P.G.J. has a financial interest in Pacific Diabetes Technologies Inc., a company that may have a commercial interest in the results of this research and technology. For all other authors, no competing interests exist.

Statement of conflict of interest

Peter G. Jacobs has a financial interest in Pacific Diabetes Technologies Inc., a company that may have a commercial interest in the results of this research and technology. For all other authors, no competing interests exist.

Acknowledgments

The guarantor of this research is Clara Mosquera-Lopez who takes responsibility for the content of the article. The authors thank Tidepool for providing the datasets and technical support during the study. Preliminary findings of this study were presented at the 15th International Conference on Advanced Technologies & Treatments for Diabetes, April 2022, in Barcelona, Spain.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2023.106670.

References

[1] J.E. Yardley, J. Hay, A.M. Abou-Setta, S.D. Marks, J. McGavock, A systematic review and meta-analysis of exercise interventions in adults with type 1 diabetes, Diabetes Res. Clin. Pract. 106 (3) (2014) 393–400, https://doi.org/10.1016/j.diabres.2014.09.038.

[2] M.C. Riddell, I.W. Gallen, C.E. Smart, et al., Exercise management in type 1 diabetes: a consensus statement, Lancet Diabetes Endocrinol. 5 (5) (2017) 377–390, https://doi.org/10.1016/S2213-8587(17)30014-1.

[3] J. Wadén, C. Forsblom, L.M. Thorn, et al., Physical activity and diabetes complications in patients with type 1 diabetes, Diabetes Care 31 (2) (2008) 230–232, https://doi.org/10.2337/dc07-1238.

[4] F. MacMillan, A. Kirk, N. Mutrie, L. Matthews, K. Robertson, D.H. Saunders, A systematic review of physical activity and sedentary behavior intervention studies in youth with type 1 diabetes: study characteristics, intervention design, and efficacy: physical activity and type 1 diabetes, Pediatr. Diabetes 15 (3) (2014) 175–189, https://doi.org/10.1111/pedi.12060.

[5] M.C. Riddell, Z. Li, R.W. Beck, et al., More time in glucose range during exercise days than sedentary days in adults living with type 1 diabetes, Diabetes Technol. Therapeut. 23 (5) (2021) 376–383, https://doi.org/10.1089/dia.2020.0495.

[6] S.R. Colberg, R.J. Sigal, J.E. Yardley, et al., Physical activity/exercise and diabetes: a position statement of the American diabetes association, Diabetes Care 39 (11) (2016) 2065–2079, https://doi.org/10.2337/dc16-1728.

[7] V.A. Bussau, L.D. Ferreira, T.W. Jones, P.A. Fournier, The 10-s maximal sprint, Diabetes Care 29 (3) (2006) 601–606, https://doi.org/10.2337/diacare.29.03.06.dc05-1764.

[8] J.E. Yardley, G.P. Kenny, B.A. Perkins, et al., Resistance versus aerobic exercise, Diabetes Care 36 (3) (2013) 537–542, https://doi.org/10.2337/dc12-0963.

[9] N. Lascar, A. Kennedy, B. Hancock, et al., Attitudes and barriers to exercise in adults with type 1 diabetes (T1DM) and how best to address them: a qualitative study, in: I. Petersen (Ed.), PLoS ONE, 2014, e108019, https://doi.org/10.1371/journal.pone.0108019, 9(9.

[10] M.C. Riddell, R. Pooni, F.Y. Fontana, S.N. Scott, Diabetes technology and exercise, Endocrinol Metab. Clin. N. Am. 49 (1) (2020) 109–125, https://doi.org/10.1016/j.ecl.2019.10.011.

[11] N.S. Tyler, C. Mosquera-Lopez, G.M. Young, J. El Youssef, J.R. Castle, P.G. Jacobs, Quantifying the impact of physical activity on future glucose trends using machine learning, iScience 25 (3) (2022), 103888, https://doi.org/10.1016/j.isci.2022.103888.

[12] A. Maran, P. Pavan, B. Bonsembiante, et al., Continuous glucose monitoring reveals delayed nocturnal hypoglycemia after intermittent high-intensity exercise in nontrained patients with type 1 diabetes, Diabetes Technol. Therapeut. 12 (10) (2010) 763–768, https://doi.org/10.1089/dia.2010.0038.

[13] M.C. Riddell, J. Milliken, Preventing exercise-induced hypoglycemia in type 1 diabetes using real-time continuous glucose monitoring and a new carbohydrate intake algorithm: an observational field study, Diabetes Technol. Therapeut. 13 (8) (2011) 819–825, https://doi.org/10.1089/dia.2011.0052.

[14] R. Basu, M.L. Johnson, Y.C. Kudva, A. Basu, Exercise, hypoglycemia, and type 1 diabetes, Diabetes Technol. Therapeut. 16 (6) (2014) 331–337, https://doi.org/10.1089/dia.2014.0097.

[15] J.E. Yardley, R.J. Sigal, Exercise strategies for hypoglycemia prevention in individuals with type 1 diabetes, Diabetes Spectr Publ Am Diabetes Assoc 28 (1) (2015) 32–38, https://doi.org/10.2337/diaspect.28.1.32.

[16] T.T.P. Nguyen, P.G. Jacobs, J.R. Castle, et al., Separating insulin-mediated and non-insulin-mediated glucose uptake during and after aerobic exercise in type 1 diabetes, Am. J. Physiol. Endocrinol. Metab. 320 (3) (2021) E425–E437, https://doi.org/10.1152/ajpendo.00534.2020.

[17] J. Wahren, Glucose turnover during exercise in man, Ann. N. Y. Acad. Sci. 301 (1 The Marathon) (1977) 45–55, https://doi.org/10.1111/j.1749-6632.1977.tb38184.x.

[18] S.K. McMahon, L.D. Ferreira, N. Ratnam, et al., Glucose requirements to maintain euglycemia after moderate-intensity afternoon exercise in adolescents with type 1

diabetes are increased in a biphasic manner, J. Clin. Endocrinol. Metab. 92 (3) (2007) 963–968, https://doi.org/10.1210/jc.2006-2263.
[19] C. Dalla Man, M.D. Breton, C. Cobelli, Physical activity into the meal glucose—insulin model of type 1 diabetes: *in silico* studies, J. Diabetes Sci. Technol. 3 (1) (2009) 56–67, https://doi.org/10.1177/193229680900300107.
[20] A.S. Brazeau, R. Rabasa-Lhoret, I. Strychar, H. Mircescu, Barriers to physical activity among patients with type 1 diabetes, Diabetes Care 31 (11) (2008) 2108–2109, https://doi.org/10.2337/dc08-0720.
[21] A.S. Brazeau, V. Gingras, C. Leroux, et al., A pilot program for physical exercise promotion in adults with type 1 diabetes: the PEP-1 program, Appl. Physiol. Nutr. Metabol. 39 (4) (2014) 465–471, https://doi.org/10.1139/apnm-2013-0287.
[22] A. Kennedy, P. Narendran, R.C. Andrews, A. Daley, S.M. Greenfield, Attitudes and barriers to exercise in adults with a recent diagnosis of type 1 diabetes: a qualitative study of participants in the Exercise for Type 1 Diabetes (EXTOD) study, BMJ Open 8 (1) (2018), e017813, https://doi.org/10.1136/bmjopen-2017-017813.
[23] L.M. Wilson, N. Tyler, P.G. Jacobs, et al., Patient input for design of a decision support smartphone application for type 1 diabetes, J. Diabetes Sci. Technol. 14 (6) (2020) 1081–1087, https://doi.org/10.1177/1932296819870231.
[24] L. Fried, T. Chetty, D. Cross, et al., The challenges of being physically active: a qualitative study of young people with type 1 diabetes and their parents, Can. J. Diabetes 45 (5) (2021) 421–427, https://doi.org/10.1016/j.jcjd.2020.09.010.
[25] O. Moser, M.C. Riddell, M.L. Eckstein, et al., Glucose management for exercise using continuous glucose monitoring (CGM) and intermittently scanned CGM (isCGM) systems in type 1 diabetes: position statement of the European Association for the Study of Diabetes (EASD) and of the International Society for Pediatric and Adolescent Diabetes (ISPAD) endorsed by JDRF and supported by the American Diabetes Association (ADA), Diabetologia 63 (12) (2020) 2501–2520, https://doi.org/10.1007/s00125-020-05263-9.
[26] R. Reddy, N. Resalat, L.M. Wilson, J.R. Castle, J. El Youssef, P.G. Jacobs, Prediction of hypoglycemia during aerobic exercise in adults with type 1 diabetes, J. Diabetes Sci. Technol. 13 (5) (2019) 919–927, https://doi.org/10.1177/1932296818823792.
[27] M. Cigrovski Berkovic, I. Bilic-Curcic, L. La Grasta Sabolic, A. Mrzljak, V. Cigrovski, Fear of hypoglycemia, a game changer during physical activity in type 1 diabetes mellitus patients, World J. Diabetes 12 (5) (2021) 569–577, https://doi.org/10.4239/wjd.v12.i5.569.
[28] C. Zecchin, A. Facchinetti, G. Sparacino, C. Cobelli, How much is short-term glucose prediction in type 1 diabetes improved by adding insulin delivery and meal content information to CGM data? A proof-of-concept study, J. Diabetes Sci. Technol. 10 (5) (2016) 1149–1160, https://doi.org/10.1177/1932296816654161.
[29] G. Sparacino, F. Zanderigo, S. Corazza, A. Maran, A. Facchinetti, C. Cobelli, Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series, IEEE Trans. Biomed. Eng. 54 (5) (2007) 931–937, https://doi.org/10.1109/TBME.2006.889774.
[30] K. Turksoy, J. Kilkus, I. Hajizadeh, et al., Hypoglycemia detection and carbohydrate suggestion in an artificial pancreas, J. Diabetes Sci. Technol. 10 (6) (2016) 1236–1244, https://doi.org/10.1177/1932296816658666.
[31] E.I. Georga, V.C. Protopappas, D. Polyzos, D.I. Fotiadis, A predictive model of subcutaneous glucose concentration in type 1 diabetes based on Random Forests, Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf. 2012 (2012) 2889–2892, https://doi.org/10.1109/EMBC.2012.6346567.
[32] E.I. Georga, V.C. Protopappas, D. Ardigo, et al., Multivariate prediction of subcutaneous glucose concentration in type 1 diabetes patients based on support vector regression, IEEE J. Biomed. Health Inform. 17 (1) (2013) 71–81, https://doi.org/10.1109/TITB.2012.2219876.
[33] M. Sevil, M. Rashid, I. Hajizadeh, M. Park, L. Quinn, A. Cinar, Physical activity and psychological stress detection and assessment of their effects on glucose concentration predictions in diabetes management, IEEE Trans. Biomed. Eng. 68 (7) (2021) 2251–2260, https://doi.org/10.1109/TBME.2020.3049109.
[34] K. Li, J. Daniels, C. Liu, P. Herrero, P. Georgiou, Convolutional recurrent neural networks for glucose prediction, IEEE J. Biomed. Health Inform. 24 (2) (2020) 603–613, https://doi.org/10.1109/JBHI.2019.2908488.
[35] C. Zecchin, A. Facchinetti, G. Sparacino, C. Cobelli, Jump neural network for real-time prediction of glucose concentration, Methods Mol. Biol. Clifton NJ 1260 (2015) 245–259, https://doi.org/10.1007/978-1-4939-2239-0_15.
[36] C. Zecchin, A. Facchinetti, G. Sparacino, G. De Nicolao, C. Cobelli, Neural network incorporating meal information improves accuracy of short-time prediction of glucose concentration, IEEE Trans. Biomed. Eng. 59 (6) (2012) 1550–1560, https://doi.org/10.1109/TBME.2012.2188893.
[37] S.M. Pappada, B.D. Cameron, P.M. Rosman, et al., Neural network-based real-time prediction of glucose in patients with insulin-dependent diabetes, Diabetes Technol. Therapeut. 13 (2) (2011) 135–141, https://doi.org/10.1089/dia.2010.0104.
[38] Y. Amar, S. Shilo, T. Oron, E. Amar, M. Phillip, E. Segal, Clinically accurate prediction of glucose levels in patients with type 1 diabetes, Diabetes Technol. Therapeut. 22 (8) (2020) 562–569, https://doi.org/10.1089/dia.2019.0435.
[39] C. Mosquera-Lopez, P.G. Jacobs, Incorporating glucose variability into glucose forecasting accuracy assessment using the new glucose variability impact index and the prediction consistency index: an LSTM case example, J. Diabetes Sci. Technol. 16 (1) (2022) 7–18, https://doi.org/10.1177/19322968211042621.
[40] T. Zhu, C. Uduku, K. Li, P. Herrero, N. Oliver, P. Georgiou, Enhancing self-management in type 1 diabetes with wearables and deep learning, Npj Digit. Med. 5 (1) (2022) 1–11, https://doi.org/10.1038/s41746-022-00626-5.
[41] M.Z. Wadghiri, A. Idri, T. El Idrissi, H. Hakkoum, Ensemble blood glucose prediction in diabetes mellitus: a review, Comput. Biol. Med. 147 (2022), 105674, https://doi.org/10.1016/j.compbiomed.2022.105674.
[42] N. Ben Brahim, J. Place, E. Renard, M.D. Breton, Identification of main factors explaining glucose dynamics during and immediately after moderate exercise in patients with type 1 diabetes, J. Diabetes Sci. Technol. 9 (6) (2015) 1185–1191, https://doi.org/10.1177/1932296815607864.
[43] N. Hobbs, I. Hajizadeh, M. Rashid, K. Turksoy, M. Breton, A. Cinar, Improving glucose prediction accuracy in physically active adolescents with type 1 diabetes, J. Diabetes Sci. Technol. 13 (4) (2019) 718–727, https://doi.org/10.1177/1932296818820550.
[44] H.M. Romero-Ugalde, M. Garnotel, M. Doron, et al., ARX model for interstitial glucose prediction during and after physical activities, Control Eng. Pract. 90 (2019) 321–330, https://doi.org/10.1016/j.conengprac.2019.07.013.
[45] B. De Paoli, F. D'Antoni, M. Merone, S. Pieralice, V. Piemonte, P. Pozzilli, Blood glucose level forecasting on type-1-diabetes subjects during physical activity: a comparative analysis of different learning techniques, Bioeng Basel Switz 8 (6) (2021) 72, https://doi.org/10.3390/bioengineering8060072.
[46] M.D. Breton, S.D. Patek, D. Lv, et al., Continuous glucose monitoring and insulin informed advisory system with automated titration and dosing of insulin reduces glucose variability in type 1 diabetes mellitus, Diabetes Technol. Therapeut. 20 (8) (2018) 531–540, https://doi.org/10.1089/dia.2018.0079.
[47] J.R. Castle, J. El Youssef, L.M. Wilson, et al., Randomized outpatient trial of single- and dual-hormone closed-loop systems that adapt to exercise using wearable sensors, Diabetes Care 41 (7) (2018) 1471–1477, https://doi.org/10.2337/dc18-0228.
[48] Blood sugar and exercise | ADA, Accessed, https://diabetes.org/healthy-living/fitness/getting-started-safely/blood-glucose-and-exercise. (Accessed 7 December 2022).
[49] M.B. Gillingham, Z. Li, R.W. Beck, et al., Assessing mealtime macronutrient content: patient perceptions versus expert analyses via a novel phone app, Diabetes Technol. Therapeut. 23 (2) (2021) 85–94, https://doi.org/10.1089/dia.2020.0357.
[50] C. Ngufor, H. Van Houten, B.S. Caffo, N.D. Shah, R.G. McCoy, Mixed effect machine learning: a framework for predicting longitudinal change in hemoglobin A1c, J. Biomed. Inf. 89 (2019) 56–67, https://doi.org/10.1016/j.jbi.2018.09.001.
[51] M. Pellagatti, C. Masci, F. Ieva, A.M. Paganoni, Generalized mixed-effects random forest: a flexible approach to predict university student dropout, Stat. Anal. Data Min. ASA Data Sci. J. 14 (3) (2021) 241–257, https://doi.org/10.1002/sam.11505.
[52] B.P. Kovatchev, Metrics for glycaemic control — from HbA1c to continuous glucose monitoring, Nat. Rev. Endocrinol. 13 (7) (2017) 425–436, https://doi.org/10.1038/nrendo.2017.3.
[53] P.G. Jacobs, J. El Youssef, J. Castle, et al., Automated control of an adaptive bihormonal, dual-sensor artificial pancreas and evaluation during inpatient studies, IEEE Trans. Biomed. Eng. 61 (10) (2014) 2569–2581, https://doi.org/10.1109/TBME.2014.2323248.
[54] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, https://doi.org/10.1023/A:1010933404324.
[55] A. Altmann, L. Toloşi, O. Sander, T. Lengauer, Permutation importance: a corrected feature importance measure, Bioinformatics 26 (10) (2010) 1340–1347, https://doi.org/10.1093/bioinformatics/btq134.
[56] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Stat. 29 (5) (2001) 1189–1232, https://doi.org/10.1214/aos/1013203451.
[57] B. Song, G. Zhang, W. Zhu, Z. Liang, ROC operating point selection for classification of imbalanced data with application to computer-aided polyp detection in CT colonography, Int. J. Comput. Assist. Radiol. Surg. 9 (1) (2014) 79–89, https://doi.org/10.1007/s11548-013-0913-8.
[58] S. Holm, A simple sequentially rejective multiple test procedure, Scand. J. Stat. 6 (2) (1979) 65–70.
[59] T. Chetty, V. Shetty, P.A. Fournier, P. Adolfsson, T.W. Jones, E.A. Davis, Exercise management for young people with type 1 diabetes: a structured approach to the exercise consultation, Front. Endocrinol. 10 (2019). Accessed, https://www.frontiersin.org/articles/10.3389/fendo.2019.00326. (Accessed 7 December 2022).
[60] Bokeh documentation, Accessed, https://docs.bokeh.org/en/latest/index.html. (Accessed 1 July 2022).
[61] G. Valli, D. Minnock, G. Tarantino, R.D. Neville, Delayed effect of different exercise modalities on glycaemic control in type 1 diabetes mellitus: a systematic review and meta-analysis, Nutr. Metabol. Cardiovasc. Dis. 31 (3) (2021) 705–716, https://doi.org/10.1016/j.numecd.2020.12.006.
[62] T. Battelino, C.M. Alexander, S.A. Amiel, et al., Continuous glucose monitoring and metrics for clinical trials: an international consensus statement, Lancet Diabetes Endocrinol. 11 (1) (2023) 42–57, https://doi.org/10.1016/S2213-8587(22)00319-9.
[63] B. Kovatchev, S. Anderson, L. Heinemann, W. Clarke, Comparison of the numerical and clinical accuracy of four continuous glucose monitors, Diabetes Care 31 (6) (2008) 1160–1164, https://doi.org/10.2337/dc07-2401.