

Received November 16, 2020, accepted November 30, 2020, date of publication December 4, 2020, date of current version December 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3042483

# Machine Learning Approach for Postprandial Blood Glucose Prediction in Gestational Diabetes Mellitus

EVGENII A. PUSTOZEROV<sup>1,2</sup>, (Member, IEEE), ALEKSANDRA S. TKACHUK<sup>2</sup>, ELENA A. VASUKOVA<sup>2</sup>, ANNA D. ANOPOVA<sup>2</sup>, MARIA A. KOKINA<sup>2</sup>, INGA V. GORELOVA<sup>3</sup>, TATIANA M. PERVUNINA<sup>3</sup>, ELENA N. GRINEVA<sup>2</sup>, AND POLINA V. POPOVA<sup>2,4</sup>

<sup>1</sup>Department of Biomedical Engineering, Saint Petersburg State Electrotechnical University, 197341 Saint Petersburg, Russia

<sup>2</sup>Institute of Endocrinology, Almazov National Medical Research Center, 194156 Saint Petersburg, Russia

<sup>3</sup>Institute of Perinatology and Pediatrics, Almazov National Medical Research Center, 197341 Saint Petersburg, Russia

<sup>4</sup>Department of Internal Diseases and Endocrinology, Pavlov First Saint Petersburg State Medical University, 197022 Saint Petersburg, Russia

Corresponding author: Evgenii A. Pustozarov (pustozarov.e@gmail.com)

This work was supported by the Russian Science Foundation “The Study of Predictors of Postprandial Glycemic Response and the Need for Insulin Therapy in Women with Gestational Diabetes Mellitus in Order to Personalize Approaches to its Treatment” under Grant 18-75-10042.

**ABSTRACT** Postprandial blood glucose prediction is a crucial part of diabetes management. Recently, this topic has been of great interest, resulting in many research projects and published papers. Although different input parameters that might be beneficial for blood glucose prediction models were comprehensively discussed, specific data preprocessing, feature engineering and model tuning steps were not explained in detail in many of these papers. In this work, we developed and comprehensively described a data-driven blood glucose model based on a decision tree gradient boosting algorithm to predict different characteristics of postprandial glycemic responses; the model utilized meal-related data derived from a mobile app diary (including information on the glycemic index), food context (information on previous meals), characteristics of the individual patients and patient behavioral questionnaires. A set of rules was defined and implemented to detect incorrect meal records and to filter faulty data, and analyses were conducted on the overall food diary data and in particular, the data on the current meal for which the postprandial blood glucose response was calculated. Different gradient boosting models were trained and evaluated with parameters selected via random search cross-validation. The best models for the prediction of the incremental area under the blood glucose curve two hours after food intake had the following characteristics:  $R = 0.631$ ,  $MAE = 0.373$  mmol/L\*h for the model not using data on current blood glucose;  $R = 0.644$ ,  $MAE = 0.371$  mmol/L\*h for the model using data on the current blood glucose levels; and  $R = 0.704$ ,  $MAE = 0.341$  mmol/L\*h for the model utilizing data on the continuous blood glucose trends before the meal. The impact of features was evaluated using Shapley values. The meal glycemic load, amount of carbohydrates in the meal, type of meal (e.g., breakfast), amount of starch and amount of food consumed 6 hours before the current meal were the most important contributors in the models.

**INDEX TERMS** Gestational diabetes mellitus, glucose prediction, gradient boosting, machine learning.

## I. INTRODUCTION

The postprandial glycemic response (PPGR) is an important characteristic of blood glucose (BG) control effectiveness and glucose metabolism in patients with all types of diabetes. Clinical trials have shown the importance of

controlling one's blood glucose level after meals within the normal range [1], [2]. Diabetic pregnancy, despite the improved metabolic control, is still a strong risk factor for alterations in fetal development and keeping fasting glucose levels in range can contribute to decreasing number of fetal malformations [3]. A considerable number of papers on blood glucose prediction for type 1 diabetes were published recently [4], and these studies utilized different machine

The associate editor coordinating the review of this manuscript and approving it for publication was Donato Impedovo<sup>1</sup>.

learning algorithms and different sets of input data. Feed-forward neural networks, combinations of physiology-based models and machine learning techniques, recurrent neural networks and support vector machines appear to be the most frequently used algorithms for blood glucose prediction [5]. With the same data, in a direct comparison with other models, gradient boosting tends to show the most precise results [6]. Although different input parameters that might be beneficial for blood glucose prediction models were comprehensively discussed, specific data preprocessing, feature engineering and model tuning steps were not explained in detail in many of these papers. There is also a lack of studies on gestational diabetes mellitus (GDM) and pregnant women in general. The aim of this study was to develop a PPGR prediction model based on data collected from GDM patients that can also be utilized as a main component of a mobile-based recommender system.

## II. MATERIALS AND METHODS

### A. RESEARCH METHODOLOGY

The CGM data were collected in a clinical trial held by the authors at Almazov National Medical Research Centre (St. Petersburg, Russia). Patients with GDM and healthy pregnant women (controls) who participated in the GEM-GDM study were included in the present study. The design of the parent GEM-GDM study is described elsewhere [7]. In brief, the women with GDM were randomized into two groups according to their glycemic goals: the first group had strict glycemic goals (fasting BG <5.1 mmol/L and <7.0 mmol/L BG two hours after meals), and the second group had less strict glycemic goals (fasting BG <5.3 mmol/L and <7.8 mmol/L BG two hours after meals). The GEM-GDM trial was registered at the ClinicalTrials.gov (Identifier: NCT03610178). Altogether, 235 participants took part in the study (97 from the first group, 101 from the second group, and 37 from the control group). The participant characteristics are presented in Table 1.

The participants were invited to participate in a one-week CGM recording session, in which they tracked information on meal consumption in the mobile app, as described elsewhere [8]–[11]. The CGM and meal-related data were matched using the same algorithm that was described by the authors in a previous paper [12].

The data were processed with Python 3.7 programming language and the scikit-learn library [13] for core machine learning procedures.

### B. EVALUATED FEATURES

The set of features used to train the model was created with a combination of the following subsets: meal-related data ( $n = 28$ ), meal context data ( $n = 25$ ), patients' individual characteristics ( $n = 34$ ), patient survey data ( $n = 44$ ), and CGM data on BG trends before the meal ( $n = 21$ ). The complete set of features is presented in the Appendix. Together with the amount of macro- and micronutrients consumed, the meal-related data included the glycemic index and glycemic

**TABLE 1. Characteristics of healthy pregnant women and GDM patients.**

Parameter	Group 1 (n=109)	Group 2 (n=113)	Control group (n=45)	P*
Age, years	32.5±4.4	33.0±4.5	31.5±4.6	0.163
BMI before pregnancy, kg/m <sup>2</sup>	25.0±5.8	25.0±5.3	23.1±4.0	0.085
Pregnancy week on the day of inclusion	25.8±5.1	26.2±4.3	27.2±2.6	0.202
Systolic BP, mmHg	121.5±11.8	121.2±10.2	116.4±9.9	0.021 0.973**
Diastolic BP, mmHg	76.0±10.4	75.6±9.6	72.3±8.1	0.078

The data are presented as the means ± standard deviations;

\* – ANOVA P-value;

\*\* – P-value when groups 1 and 2 were compared with Student's coefficient;

BMI – body mass index, GDM – gestational diabetes mellitus, BP – blood pressure.

load, which were assigned to food included in the database using the algorithm described by the authors in another paper [14].

The output PPGR characteristics included the peak BG level after the meal start (BGMax, mmol/L), incremental area under the glycemic curve 120 minutes after meal start (iAUC120, mmol/L\*h), BG rise from the meal start to the BG peak (BGRise, mmol/L), and BG 60 minutes after the meal start (BG60, mmol/L), which are all numeric values. These characteristics were evaluated in three scenarios:

- no data on the preprandial BG levels were included in the prediction;
- a single BG level measurement at the start of the meal was included as an input feature (BG0, mmol/L);
- the data on the CGM trends before the meal were included as input features.

These scenarios describe situations in which no glucose measurements are made before meals (a), a single measurement is made via a glucometer or flash CGM (b), and real-time CGM trends are used in the prediction algorithm (c). The main focus was on the iAUC120 measure, which is also often referred to as a measure of the postprandial glycemic response in the literature.

### C. FLAWED RECORD DETECTION ALGORITHM

Wrong meal information recorded in the diary due to a lack of motivation or deliberate misreporting from the patient might be the key problem in developing data-driven BG prediction models [14].

A set of rules was formulated by a group of Almazov Centre endocrinologists working with meal diaries to detect and remove flawed meal records in patients' diaries to improve models. These included features were derived

from the analysis of the overall meal diary data during the CGM recording period and the analysis of the characteristics of particular meals for which the PPGR was evaluated.

The following rules were formulated and implemented to filter negligently filled-in and misreported information:

1 – more than 50% of the meals recorded in the diary consisted of a single dish or a single dish with a single beverage (negligent recording);

2 – the average daily consumed energy according to the information in the diary was less than 1000 kcal (evidence of food consumption underreporting);

3 – the weight of more than 50% of the dishes included in the diary were rounded to the hundreds, excluding beverages (rough rounding);

4 – more than 20% of the postprandial BG glucometer estimations were distinguished from the appropriate CGM measurements by 1 mmol/L or more (potential misreporting);

5 – the amount of snacks recorded in the diary was less than 10% of all meal records (meal underreporting).

Before the meal and corresponding PPGR were added to the dataset for model training, a set of rules was used to filter the meals (mainly to check whether the time recorded in the food diary was correctly recorded):

1 – an insulin injection was performed less than 300 minutes before meal start (we considered only meals without prior insulin injections);

2 – the meal was accompanied by a subsequent meal occurring less than 60 minutes after its start;

3 – the meal was on the CGM peak: the BG level at the meal start was more than 1 mmol/L higher than that at an hour before the meal;

4 – the meal was on the falling edge of the CGM peak: the BG level at half an hour before the meal was at least 0.4 mmol/L higher than that at the meal start, while the BG level at half an hour after the meal was at least 0.4 mmol/L lower than that at the meal start;

5 – the meal had an inadequately low PPGR to a considerable amount of carbohydrates: a meal with more than 40 g of carbohydrates with a subsequent incremental area under the CGM curve at 120 minutes after meal start (iAUC120) of less than 0.3 mmol/L\*h.

#### D. GRADIENT BOOSTING REGRESSION

A gradient tree boosting model [15] was chosen as the model to predict the PPGR due to its high prediction accuracy with heterogeneous datasets and ability to work with missing data. A detailed description of the algorithm can be found in the original paper [16]. In brief, this algorithm creates a prediction model in the form of an ensemble of weak prediction models. It builds the model in a stagewise fashion and generalizes them by allowing the optimization of a loss function, which in our case, was the mean square error (MSE).

#### E. COMPARISON OF ALGORITHM REALIZATIONS

There are three commonly used realizations of gradient boosting algorithms: xgboost [17], catboost [18] and lightgbm [19]. We trained and evaluated three algorithms under the same conditions to compare them in terms of precision. The appropriate R, MAE, MSE and RMSE were chosen as the metrics for comparison. The resulting precision did not differ significantly between the models of different realizations (up to the second value after the dot for the above listed characteristics). While the training time differed vastly between libraries (with lightgbm requiring approximately 9 times less time to train than xgboost and 39 times less time than catboost with an 8-core CPU), the training time itself was not chosen as an important parameter, as after initial training, the model will be used in mobile apps for predictions only. The xgboost model was chosen for further analysis; the model had comparable precision and was implemented with the greatest ease in Android apps, as it has the officially supported Java package XGBoost4J, which had previously been used in a number of projects.

#### F. PARAMETER TUNING AND MODEL SELECTION

Model hyperparameters were tuned via 300 rounds of random grid searching with 10-fold grouped cross-validation with the coefficient of determination as a scoring value. The groups were separated in such a way that the records from the same participant did not appear in either the training or validation set. The set of hyperparameters included the following: the subsample, which was the subsample ratio of the training instances; n\_estimators, which was the amount of trees to be constructed; min\_child\_weight, which was the minimum sum of instance weight (Hessian) needed in a child node for additional partitioning; max\_depth, which was the maximum depth of a tree; learning\_rate (eta), which was the step size shrinkage used in an update to prevent overfitting; gamma, which was the minimum loss reduction required to make another partition on a leaf node of the tree; colsample\_bytree, which was the subsample ratio of columns used when constructing each tree; reg\_alpha, which was the L1 regularization term for the weights; and reg\_lambda, which was the L2 regularization term for the weights. This covers all the tunable parameters of the extreme gradient boosting machine [17].

The ranges of the hyperparameters specified in the random search are shown below:

```
'max_depth': range(1, 13, 1),
'min_child_weight': range(1, 8, 1),
'gamma': [i / 10.0 for i in range(0, 7)],
'subsample': [i / 10.0 for i in range(4, 10)],
'colsample_bytree': [i / 10.0 for i in range(4, 10)],
'reg_alpha': [0, 1e-5, 1e-2, 0.1, 1, 10, 100, 1000],
'reg_lambda': [0, 1e-5, 1e-2, 0.1, 1, 10, 100, 1000],
'n_estimators': [100, 500, 1000, 1500, 2000, 5000],
'learning_rate': [0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.3].
```

**TABLE 2. Comparison of food consumption between groups.**

Parameter	Group 1 (N=57)	Group 2 (N=64)	Control (N=25)	P
Carbohydrates, grams/day	142.8*±40.9	134.8*±42.7	194.5±65.5	<.001
Proteins, grams/day	77.4±18,7	74.6±22.9	80.4±21.9	.484
Fats, grams/day	71.4±19.9	68.7*±22.8	85.7±29.1	.008
Calories, kcal/day	1561.6*±349.8	1484.3*±413.1	1901.5±532.8	<.001
Glycemic load, Units/day	77.5*±26.4	71.6*±24.6	110.0±43.7	<.001

The data are presented as the means ± standard deviations.

\* – p<0.05 compared with the control group (via the Games-Howell test).

The presented above range of hyperparameters covers every possible feasible solution for the extreme gradient boosting algorithm [17].

**G. MODEL EVALUATION**

Pearson’s coefficient of correlation R and mean absolute error (MAE) were chosen as the key metrics for model evaluation with the test set. The final accuracy was evaluated with the data from new patients (25% of all patients). Feature importance and its effect on the output variable in predicted models was evaluated with the Shapley additive explanations method [20].

**III. RESULTS**

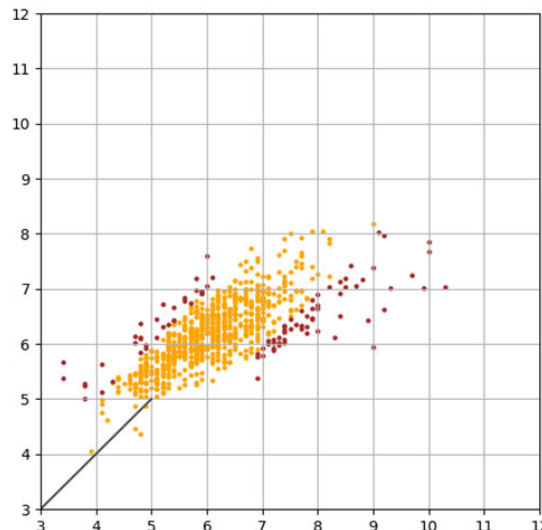
**A. PARTICIPANTS DATA**

The final dataset included information on 3240 records of meals and corresponding PPGRs from patients. The data were divided into a training and test set at a ratio of 75%/25% so that the data from the same participant was included only in the training and/or the test set.

**B. DATA FILTERING**

After the filtering process, the CGM and diaries from 144 participants were selected by the algorithm for model construction (57 from the first group, 64 from the second group, 25 from the control group). The food diary data were analyzed and discussed with the patients by endocrinologists during in-clinic visits and via online consultations. The analysis of the food diary data showed that the women from the control group consumed significantly more calories, carbohydrates and fats per day and had a higher daily glycemic load, while there was no difference between the GDM patients in the groups with different glycemic goals (Table 2).

Such a considerable difference in daily calorie intake and carbohydrates consumed between the GDM and control



**FIGURE 1. The results of peak postprandial blood glucose prediction (BGMax) with the test set. X scale - real values, Y scale – predicted values. The yellow dots correspond to evaluations with mistakes of magnitude less than 1 mmol/L, and the red dots correspond to mistakes of magnitude greater than 1 mmol/L.**

groups might be due to not only dieting behavior but also the underreporting of meal-related data by the GDM patients, even though all the patients were informed about the recommended daily consumption of calories and macronutrients and potential adverse effects of severe diet restrictions on the fetus. However, the patients tended to adhere to strict diets to avoid the need for insulin therapy.

**C. PRECISION OF EVALUATED MODELS**

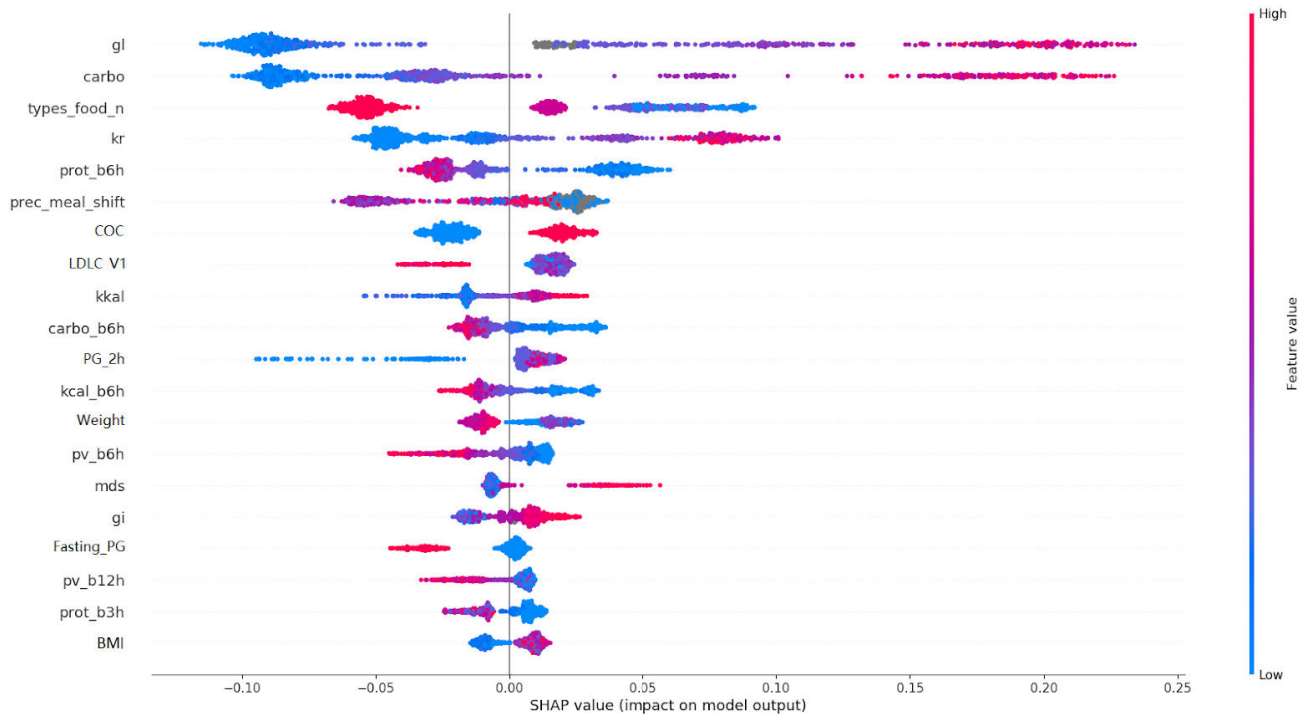
Models were evaluated for each of the following three scenarios: without BG data, with data on the BG level only at meal start and with data on the BG trends derived from CGM. The resulting precision metrics for each of the evaluated PPGR characteristics are presented in Table 3.

The mean absolute error (MAE) for the BGMax model when the CGM trends were used was 0.528 mmol/L, with a Pearson correlation coefficient for the predicted and real values of R = 0.740. When no CGM trends but BG0 data were used, the values were MAE = 0.556 mmol/L and R = 0.725; when the BG data was not used, the values were MAE = 0.682 mmol/L with R = 0.527. The evaluation results of the precision of the BGmax predictive model are shown in Fig. 1. The complete set of model precision metrics and hyperparameters for the xgboost models are shown in table 3.

**D. FEATURE IMPORTANCE EVALUATION**

The Shapley additive explanations method was implemented by means of the SHAP package. The results of the feature importance evaluation using the Shapley value method are presented in Fig. 2. The graph shows the influence of each of the 20 most significant features (in descending order from top to bottom) on the iAUC120 prediction for a given data point in the test set. The scale shows the influence of the feature





**FIGURE 2.** SHAP value evaluation for the 20 most important features for iAUC120 prediction with the test set. *gl* – glycemic load of the meal; *carbo* – the amount of carbohydrates in the meal, in g; *types\_food\_n* – meal type (1 – breakfast, 2 – lunch, 3 – dinner, 4 – snack, the value was one-hot encoded); *kr* – the amount of starch in the meal, in g; *prot\_b6h* – proteins consumed 6 hours before the meal (grams); *prec\_meal\_shift* – time between the preceding and current meals (minutes); *COC* – combined oral contraceptive use before pregnancy (1 – yes, 0 – no); *LDLC\_V1* – low density lipoprotein cholesterol at the time of inclusion in the study; *kkal* – the energy value of the meal; *carbo\_b6h* – carbohydrates consumed 6 hours before the meal (grams); *PG\_2h* – 2-hour plasma glucose level in OGTT (mmol/L); *kcal\_b6h* – energy value of foods consumed 6 hours before the meal (kcal); *Weight* – prepregnancy weight, kg; *pv\_b6h* – alimentary fibers consumed 6 hours before the meal (grams); *mds* – the amount of monosaccharides and disaccharide in the meal, in g; *gi* – glycemic index of the meal; *Fasting\_PG* – fasting plasma glucose at the time of inclusion in the study (mmol/L); *pv\_b12h* – alimentary fibers consumed 12 hours before the meal (grams); *prot\_b3h* – proteins consumed 3 hours before the meal (grams); *BMI* – prepregnancy body mass index, kg/m<sup>2</sup>.

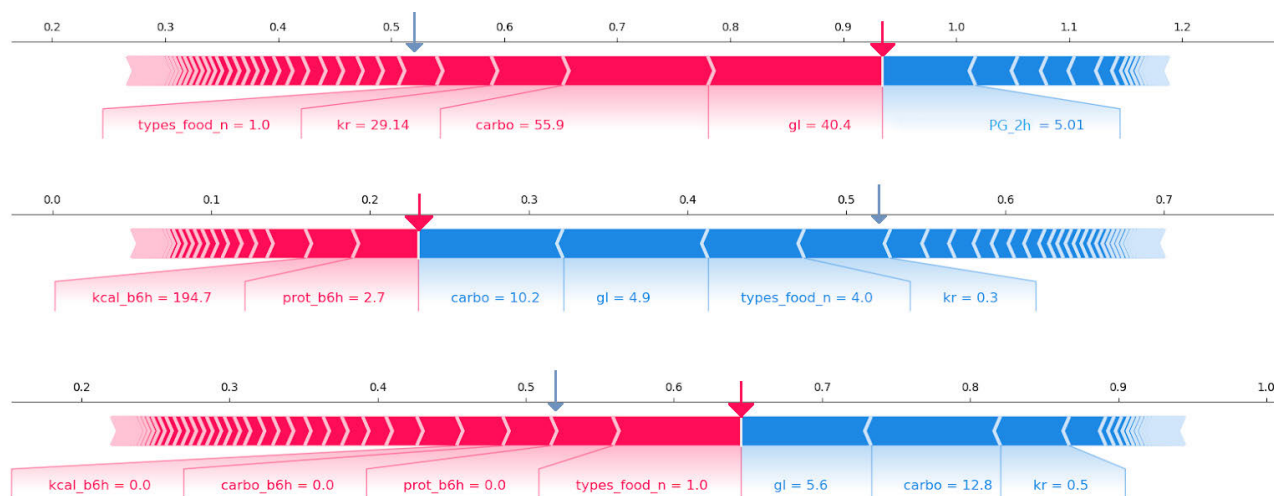
on each prediction (Shapley value). The farther the point lies from zero (shown as a gray vertical line), the stronger the impact of this feature on the output (e.g., a glycemic load of 20 is lower than average and tends to lead to a lower predicted iAUC120). The colors correspond to the values of the features for each particular point, ranging from below average (blue) to average (purple) to above average (red). For the top two most significant features (*gl* and *carbo*), it can be clearly seen that lower values of glycemic load and carbohydrates result in significantly lower iAUC120 values.

Fig. 3 shows three particular predictions of iAUC120 for three meals corresponding to different patients included in the test set. The features that increased the predicted values are shown in red, and those that decreased the values are shown in blue. The cyan vertical arrows at the top of the three graphs show the mean iAUC120 value (0.52 mmol/L\*h, same value on all three graphs), and the currently predicted iAUC120 value is shown with larger red vertical arrows (for the upper graph, it is equal to 0.94 mmol/L\*h; for the middle graph, it is equal to 0.22 mmol/L\*h; and for the bottom graph, it is equal to 0.65 mmol/L\*h). The length of each line corresponding to a different feature is proportional to

the magnitude by which the feature influenced the predicted value. These lines are shown in descending order from the vertical line corresponding to the predicted value (e.g., for the prediction on the top graph values *gl*=40.4, *carbo*=55.9, *kr*=29.14, and *types\_food\_n*=1 are the top four values that increase the prediction values, while *PG\_2h*=5.01 is the top value that decreases the prediction values). As seen from the second example, when we have a glycemic load and an amount of carbohydrates that are significantly lower than average, the model tends to predict a very low PPGR. In contrast, in the third situation, the model predicts a PPGR above the average, despite the low glycemic load and small amount of carbohydrates due to the type of meal (breakfast) and the absence of food being consumed 6 hours before the meal (*kcal\_b6h* = 0), which might also be evidence of a morning meal or snack.

**IV. DISCUSSION**

Effective strategies are required to reduce the immense global burden of GDM on maternal and offspring health outcomes. Diet is a fundamental component of the treatment of GDM. The current nutritional guidelines are based on population averages.



**FIGURE 3.** Sample evaluations of iAUC120 for three different meals consumed by various patients. Upper situation – the meal contains a considerable amount of carbohydrates, middle – the meal contains a small amount of carbohydrates, bottom – the meal consumed in the morning contains a small amount of carbohydrates. The large red vertical arrow shows the place of the predicted iAUC120 on the scale, while the cyan smaller vertical arrow shows the average iAUC120 of the dataset. The red horizontal arrows show the increasing effect of variables, while the blue color shows the decreasing effect of the variables. types\_food – meal type (1 – breakfast, 2 – lunch, 3 – dinner, 4 – snack, the value was one-hot encoded); kr – the amount of starch in the meal, in g; carbo – the amount of carbohydrates in the meal, in g; gi – glycemic load of the meal; PG\_2h – 2-hour plasma glucose level in OGTT (mmol/L); kcal\_b6h – energy value of foods consumed 6 hours before the meal (kcal); prot\_b6h – proteins consumed 6 hours before the meal (grams); carbo\_b6h – carbohydrates consumed 6 hours before the meal (grams).

However, the variability in the success of diet and lifestyle programs as well as the increasing evidence of high inter-personal variability in PPGRs supports the concept that one size does not fit all in terms of nutritional recommendations. To address this issue, several studies on the development of PPGR prediction algorithms in healthy adults have been recently published [21]–[23].

In this study, we derived algorithms that predict PPGRs to specific foods in pregnant women with and without GDM and evaluated the influence (input) of factors explaining PPGRs.

Tuning hyperparameters using a randomized grid search with a big set of iterations on a wide range of hyperparameters with the 10-fold grouped cross validation and the following evaluation on the data from new patients reassures the top features listed in Fig. 2 and the set of hyperparameters presented in Table 3 should remain the same to fit the data of new patients. However, there might be fluctuations due to variance in the data.

The comparison of the results acquired in the study with results from recent papers shows a similar level of precision, although the analysis methods and input features are not directly comparable. For instance,  $R = 0.70$  for the model predicting iAUC120 (which reflects the PPGR) presented in a study by Zeevi et al. [21] was achieved when the model was tested and evaluated in healthy patients with the use of gut microbiota data. Mendes-Soares [22] achieved  $R = 0.62$  while also using data on CGM trends. In our study,  $R = 0.631$  and  $MAE = 0.373$  mmol/L\*h for the model not using blood glucose data,  $R = 0.644$  and  $MAE = 0.371$  mmol/L\*h for the model using data on the current blood glucose levels, and  $R = 0.704$  and  $MAE = 0.341$  mmol/L\*h for the model using

data on the continuous blood glucose trends. To improve the precision of the presented algorithm, gut microbiota data can be included. However, gut microbiota profiling increases the cost and thus may decrease the utility of the algorithm.

There are methods showing good accuracy in predicting BG based on preceding CGM records, such as methods utilizing convolutional neural networks [24], [25], which showed RMSE of 1.21 (for the best patient) and 1.85 mmol/L (in average) respectively for BG prediction 60 minutes ahead, which are good values concerning prediction for type 1 diabetes patients). The comparable result was recently shown utilizing random forest in the same setting [26]. But the requirement of CGM systems to be constantly utilized in order to predict BG is expensive and inapplicable in a wide clinical practice for GDM patients.

Table 4 compares the prediction quality of the developed models to different types of models recently developed and presented in the literature. There are no models developed for GDM and pregnant women in the literature, so we compared our model with those for healthy people and type 1 diabetes mellitus patients. All the models exhibit adequate accuracy that allows them to be used in patient assistance. The developed model in comparison to others does not require microbiome data as models by Zeevi et al. [21] and Mendes-Soares et al. [22] or continuous blood glucose measurements on the time of prediction as models by Li et al. [24], Zhu et al. [25] and Rodriguez-Rodriguez et al. [26], which makes it much more accessible for clinical practice.

In this study, we demonstrated the significant importance of meal characteristics, food context and some individual

**TABLE 3.** Resulted precision metrics and hyperparameters for the best models.

Y	BG0	CGM	R <sup>a</sup>	MAE	Hyperparameters for the best model selected in random search
BGRise	no	no	0.622	0.540	subsample: 0.5, reg_lambda: 0, reg_alpha: 10, n_estimators: 5000, min_child_weight: 6, max_depth: 5, learning_rate: 0.001, gamma: 0.1, colsample_bytree: 0.4
BGRise	yes	no	0.638	0.529	subsample: 0.6, reg_lambda: 1, reg_alpha: 0.1, n_estimators: 1000, min_child_weight: 5, max_depth: 2, learning_rate: 0.005, gamma: 0.4, colsample_bytree: 0.4
BGRise	yes	yes	0.673	0.496	subsample: 0.9, reg_lambda: 1, reg_alpha: 1, n_estimators: 500, min_child_weight: 1, max_depth: 6, learning_rate: 0.01, gamma: 0.5, colsample_bytree: 0.7
iAUC120	no	no	0.629	0.378	subsample: 0.6, reg_lambda: 10, reg_alpha: 10, n_estimators: 500, min_child_weight: 1, max_depth: 12, learning_rate: 0.01, gamma: 0.6, colsample_bytree: 0.5
iAUC120	yes	no	0.644	0.377	subsample: 0.7, reg_lambda: 0.1, reg_alpha: 10, n_estimators: 500, min_child_weight: 2, max_depth: 11, learning_rate: 0.01, gamma: 0.1, colsample_bytree: 0.5
iAUC120	yes	yes	0.704	0.341	subsample: 0.8, reg_lambda: 1, reg_alpha: 1e-05, n_estimators: 1000, min_child_weight: 5, max_depth: 4, learning_rate: 0.005, gamma: 0, colsample_bytree: 0.5
BGmax	no	no	0.527	0.682	subsample: 0.8, reg_lambda: 0.1, reg_alpha: 1e-05, n_estimators: 500, min_child_weight: 5, max_depth: 2, learning_rate: 0.01, gamma: 0.5, colsample_bytree: 0.9
BGmax	yes	no	0.725	0.556	subsample: 0.6, reg_lambda: 1e-05, reg_alpha: 0.1, n_estimators: 5000, min_child_weight: 6, max_depth: 12, learning_rate: 0.001, gamma: 0.1, colsample_bytree: 0.5
BGmax	yes	yes	0.740	0.528	subsample: 0.5, reg_lambda: 10, reg_alpha: 0.1, n_estimators: 1500, min_child_weight: 2, max_depth: 6, learning_rate: 0.01, gamma: 0.6, colsample_bytree: 0.7
BG60	no	no	0.462	0.712	subsample: 0.7, reg_lambda: 1000, reg_alpha: 0.01, n_estimators: 5000, min_child_weight: 6, max_depth: 10, learning_rate: 0.005, gamma: 0.2, colsample_bytree: 0.4
BG60	yes	no	0.688	0.556	subsample: 0.7, reg_lambda: 0.1, reg_alpha: 1, n_estimators: 100, min_child_weight: 4, max_depth: 12, learning_rate: 0.05, gamma: 0.5, colsample_bytree: 0.9
BG60	yes	yes	0.718	0.540	subsample: 0.7, reg_lambda: 0, reg_alpha: 0.1, n_estimators: 5000, min_child_weight: 7, max_depth: 8, learning_rate: 0.005, gamma: 0.1, colsample_bytree: 0.9

BGRise – blood glucose level rise from the meal start to the peak level (mmol/L), iAUC120 – incremental area under the glycemic curve at 120 minutes after meal start (mmol/L\*h), BGMax – peak BG level after the meal start (mmol/L), BG60 – blood glucose level at 60 minutes after the meal start (mmol/L), R – Pearson's correlation, MAE – mean absolute error (mmol/L or mmol/L\*h).

<sup>a</sup>All p-values for correlations are below 0.001.

**TABLE 4. Comparison between the proposed and prior developed models.**

Value	Authors	Model	Precision	Population
iAUC120	Pustozarov et al.	Boosted decision trees	R=0.644, MAE=0.377 mmol/L*h	pregnant and GDM
iAUC120	Zeevi et al. [21]	Boosted decision trees	R=0.70	healthy
iAUC120	Mendes-Soares et al. [22]	Boosted decision trees	R=0.62	healthy
BG60	Pustozarov et al.	Boosted decision trees	R=0.687, MAE=0.556	pregnant and GDM
BG60	Li et al. [23]	Convolutional NN	RMSE=1.21 mmol/L	T1DM
BG60	Zhu et al. [24]	Convolutional NN	RMSE=1.85 mmol/L	T1DM
BG60	Rodriguez-Rodriguez et al. [25]	Random Forest	RMSE=1.23 mmol/L	T1DM

MAE – mean absolute error;  
RMSE – root mean squared error.

characteristics in PPGR prediction. Information on preceding BG measurements also plays a significant role in improving model precision, but in the majority of cases in clinical practice, these data are not available or significantly increase the cost of monitoring and cause an inconvenience for the patient. Therefore, in this paper, we focused on SHAP value evaluation for the most important features for iAUC120 prediction with the algorithm not utilizing the data on BG trends.

The most important input into PPGR prediction was made by GL and the amount of carbohydrates, which is in line with the existing evidence [23], [27]. GL is the result of the amount of carbohydrates in the food consumed multiplied by its GI. The impact of GI itself was much smaller than that of the amount of carbohydrates, as it was ranked 16<sup>th</sup> in the top most important PPGR contributors. This is because the algorithm takes the most important data on GI from GL characteristics and, in the majority of cases, does not require the inclusion of GI itself.

The third most important feature contributing to PPGR was the type of food consumed, with breakfast being the factor that increased the PPGR values. This finding can be explained by the increase in insulin resistance due to the physiological surge in contra-insulin hormones in the morning

hours. We found that meal timing is an important factor influencing the PPGR, which is in line with the recent data obtained by Berry et al. These authors developed a machine-learning model that predicted metabolic responses to food intake based in a large cohort of healthy adults in the United Kingdom and noticed that meal timing had larger effects than anticipated [27].

Another interesting finding is that not only does the composition of meals for which the PPGR is evaluated play a crucial role, but also the data on all food consumed within 6-12 h prior to the meal are important. For example, the fifth most important feature for iAUC120 prediction was the amount of proteins consumed 6 hours before the meal. Numerous studies conducted in rodents and humans have demonstrated that high protein (HP) diets improve glucose homeostasis. Acute short-term HP intake lowers postprandial glucose levels compared to low protein (LP) intake in healthy adults [28], [29] and in individuals with diabetes [30]. It has been postulated that these improvements in glucose control result from a decrease in dietary carbohydrate content; however, the glucoregulatory role of upper small intestinal peptide transporter 1 (PepT1) in the upper small intestine of healthy rats was recently demonstrated by Dranse et al [31], providing evidence that the glucoregulatory influence of acute HP intake results from the presence of protein itself and providing insight into the underlying mechanism.

Among the individual participants' features explaining the glycemic response, the use of combined oral contraceptives (COC) before pregnancy had the highest SHAP values, increasing the predicted PPGR. These data support the conclusions of several studies in which impairment in insulin sensitivity and glucose tolerance had been described with the use of oral contraceptives and evidenced by higher glucose and insulin levels [32], [33].

The limitation of the study is the self-reported nature of the meal-related data derived from the electronic diaries. Unfortunately, the importance of precise evaluation of meal composition in patients' diaries, which is almost impossible for unmotivated patients, plays a key role in the performance of developed PPGR prediction algorithms. In this work, we presented a set of rules that can be used to automatically identify flawed user inputs and filter them to improve model accuracy.

Although we developed PPGR prediction models with precision levels comparable to those reported in other studies, there is room for improvement; for example, the inclusion of microbiome and metabolomics data and detailed assessments of physical activity would increase the costs but may also enhance prediction quality.

## V. CONCLUSION

Gradient boosting models provide an effective solution for postprandial blood glucose prediction. Glycemic load, the amount of carbohydrates and meal type are the most significant features influencing the PPGR (with BG levels much higher than expected after breakfast), while the amount



of food consumed during the 6 hours before the current meal also plays a significant role.

## APPENDIX

The complete set of features used for PPGR prediction.

### Basic Features:

group – group number (1 – GDM, 2 – healthy)  
n\_cgm – the order of CGMS installation (1 – first, 2 – second)  
part\_of\_day – time of the day (1=0-4, 2=4-8, 3=8-12, 4=12-16, 5=16-20, 6=20-24)

preg\_week – gestational age on the day of the meal

time\_of\_day – time of the day (hour)

### Meal Characteristics:

a – the amount of retinol in the meal, in mcg  
b1 – the amount of thiamine in the meal, in mg  
b2 – the amount of riboflavin in the meal, in mg  
c – the amount of ascorbic acid, in mg  
ca – the amount of Ca in the meal, in mg  
carbo – the amount of carbohydrates in the meal, in g  
fat – the amount of fats in the meal, in g  
fe – the amount of iron in the meal, in mg  
gi – glycemic index of the meal  
gl – glycemic load of the meal  
k – the amount of K in the meal, in mg  
kar – the amount of beta-carotene in the meal, in mcg  
kkal – the energy value of the meal  
kr – the amount of starch in the meal, in g  
mds – the amount of monosaccharides and disaccharides in the meal, in g  
mg – the amount of Mg in the meal, in mg  
na – the amount of Na in the meal, in mg  
ne – the amount of niacin equivalent in the meal, in mg  
ok – the amount of organic acids in the meal, in g  
p – the amount of P in the meal, in mg  
prot – the amount of proteins in the meal, in g  
pv – the amount of alimentary fiber in the meal, in g  
re – the amount of retinol equivalent in the meal, in mcg  
types\_food – meal type (1 – breakfast, 2 – lunch, 3 – dinner, 4 – snack)

water – the amount of water in the meal, in g

zola – the amount of ash in the meal, in g

### Meal Context:

carbo\_b3h – carbohydrates consumed 3 hours before the meal (grams)  
carbo\_b6h – carbohydrates consumed 6 hours before the meal (grams)  
carbo\_b12h – carbohydrates consumed 12 hours before the meal (grams)  
fat\_b3h – fats consumed 3 hours before the meal (grams)  
fat\_b6h – fats consumed 6 hours before the meal (grams)  
fat\_b12h – fats consumed 12 hours before the meal (grams)  
gl\_b3h – glycemic load of the foods consumed 3 hours before the meal  
gl\_b6h – glycemic load of the foods consumed 6 hours before the meal

gl\_b12h – glycemic load of the foods consumed 12 hours before the meal

kcal\_b3h – energy value of the foods consumed 3 hours before the meal (kcal)

kcal\_b6h – energy value of the foods consumed 6 hours before the meal (kcal)

kcal\_b12h – energy value of the foods consumed 12 hours before the meal (kcal)

prec\_meal\_gi – glycemic index of the preceding meal

prec\_meal\_gl – glycemic load of the preceding meal

prec\_meal\_carbo – the amount of carbohydrates in the preceding meal (grams)

prec\_meal\_prot – the amount of proteins in the preceding meal (grams)

prec\_meal\_fat – the amount of fats in the preceding meal (grams)

prec\_meal\_pv – the amount of alimentary fiber in the meal (grams)

prec\_meal\_shift – time between preceding and current meals (minutes)

prot\_b3h – proteins consumed 3 hours before the meal (grams)

prot\_b6h – proteins consumed 6 hours before the meal (grams)

prot\_b12h – proteins consumed 12 hours before the meal (grams)

pv\_b3h – alimentary fibers consumed 3 hours before the meal (grams)

pv\_b6h – alimentary fibers consumed 6 hours before the meal (grams)

pv\_b12h – alimentary fibers consumed 12 hours before the meal (grams)

### Participant's Individual Characteristics:

AH – arterial hypertension in history

AI\_V1 – atherogenic index at the time of inclusion in the study (V1)

Age – age, years

beta\_OHB\_V1 – beta-hydroxybutyrate level at the time of inclusion in the study

BMI – prepregnancy body mass index, kg/m<sup>2</sup>

BP\_dyast1 – diastolic blood pressure at the time of inclusion in the study, mm Hg

BP\_syst1 – systolic blood pressure at the time of inclusion in the study, mm Hg

CI – cervical insufficiency (1 – yes, 2 – no)

Chol\_V1 – cholesterol level (mmol/L) at the time of inclusion in the study

COC – combined oral contraceptive use (1 – yes, 0 – no)

DM\_hystory – the presence of diabetes mellitus in the family history (1 – yes, 0 – no)

Diet\_start – gestational age at the time dieting was started

edema1 – edema during pregnancy (0 – no, 1 – yes)

education – level of education (1 – secondary, 2 – higher)

Fasting\_PG – fasting plasma glucose level at the time of inclusion in the study (mmol/L)

FPG\_OGTT – fasting plasma glucose level in OGTT

FR\_V1 – serum fructosamine level (mcmol/l) at the time of inclusion in the study

GA\_sm\_stopped – gestational age when smoking was stopped

GDM\_history – a history of GDM (1 – yes, 2 – no)

gest\_age\_V1 – gestational age at the time of testing V1 (at the time of inclusion in the study)

HbA1C\_V1 – glycosylated hemoglobin level at the time of inclusion in the study

HDLC\_V1 – high-density lipoprotein cholesterol level at the time of inclusion in the study

height – height, cm

insulin\_V1 – plasma insulin level at the time of inclusion in the study

IGT – impaired glucose tolerance before pregnancy

ketones\_V1 – urinary ketones level at the time of inclusion in the study

LDLC\_V1 – low-density lipoprotein cholesterol level at the time of inclusion in the study

leptin\_V1 – serum leptin (ng/ml) level at the time of inclusion in the study

menses – the presence of a regular menstrual cycle (1 – yes, 0 – no)

N\_abortions – the number of abortions in the patient history

N\_deliveries – the number of deliveries in the patient history

N\_pregnancies – the number of pregnancies in the patient history

N\_pregnancy\_loss – the number of pregnancy loss episodes in the patient history

PCOS – the presence of polycystic ovary syndrome (0 – no, 1 – yes)

PG\_1h – 1-hour plasma glucose level in the OGTT (mmol/L)

PG\_2h – 2-hour plasma glucose level in the OGTT (mmol/L)

placenta\_previa – the presence of placenta previa during pregnancy (1 – yes, 0 – no)

prolactin – a history of hyperprolactinemia (1 – yes, 0 – no)

smoking\_duration – smoking duration, years (before pregnancy)

TG\_V1 – serum triglyceride level (mmol/L) at the time of inclusion in the study

threatened\_miscarriage – threatened miscarriage at any time of pregnancy (1 – yes, 0 – no)

VLDLC\_V1 – very low density lipoprotein cholesterol at the time of inclusion in the study

Weight – prepregnancy weight, kg

*Lifestyle Survey:*

alcohol1 – alcohol consumption frequency before pregnancy (1 – did not consume alcohol before pregnancy; 2 – consumed alcohol before pregnancy 0.5 – 2 times/week; 3 – consumed alcohol before pregnancy more than 2 times/week)

alcohol2 – alcohol consumption frequency during pregnancy (1 – did not consume alcohol before pregnancy; 2 – consumed alcohol before pregnancy 0.5 – 2 times/week; 3 – consumed alcohol before pregnancy more than 2 times/week)

bread\_any1 – frequency of eating bread (if any) before pregnancy (1 – less than 6 times per week; 2 – 6-12 times a week; 3 – more than 12 times per week)

bread\_any2 – frequency of eating bread (if any) during pregnancy (1 – less than 6 times per week; 2 – eating 6-12 times a week; 3 – more than 12 times per week)

bread\_whole\_grain\_bread1 – frequency of eating whole grain bread before pregnancy (1 – less than 1 time per week; 2 – less than 1-3 times a week; 3 – more than 3 times a week)

bread\_whole\_grain\_bread2 – frequency of eating whole grain bread during pregnancy (1 – less than 1 time per week; 2 – less than 1-3 times a week; 3 – more than 3 times a week)

cakes1 – frequency of eating cakes before pregnancy (1 – less than 2 times per week; 2 – 2-4 times a week; 3 – more than 4 times a week)

cakes2 – frequency of eating cakes during pregnancy (1 – less than 2 times per week; 2 – 2-4 times a week; 3 – more than 4 times a week)

chocolate1 – frequency of eating chocolate before pregnancy (1 – less than 2 times a week; 2 – 2-4 times a week; 3 – more than 4 times a week)

chocolate2 – frequency of eating chocolate during pregnancy (1 – less than 2 times a week; 2 – 2-4 times a week; 3 – more than 4 times a week)

climbing\_the\_stairs1 – number of flights of stairs climbed before pregnancy (1 – less than 4 flights per day; 2 – 4-16 flights of stairs per day; 3 – more than 16 flights of stairs per day)

climbing\_the\_stairs2 – number of flights of stairs climbed during pregnancy (1 – less than 4 flights per day; 2 – 4-16 flights of stairs per day; 3 – more than 16 flights of stairs per day)

coffe1 – frequency of drinking coffee before pregnancy (1 – 0-1 per day; 2 – 2-3 per day; 3 – more than 3 times per day)

coffe2 – frequency of drinking coffee during pregnancy (1 – 0-1 cup per day; 2 – 2-3 per day; 3 – more than 3 times per day)

dairy\_products1 – frequency of eating dairy products before pregnancy (1 – less than 3 times per week; 2 – 3-6 times a week; 3 – more than 6 times a week)

dairy\_products2 – frequency of eating dairy products during pregnancy (1 – less than 3 per week; 2 – 3-6 times a week; 3 – more than 6 times a week)

dried\_fruits\_1 – frequency of eating dried fruit before pregnancy (1 – 0; 2 – 1-3 times a week; 3 – more than 3 times a week)

dried\_fruits\_2 – frequency of eating dried fruit during pregnancy (1 – 0; 2 – 1-3 times a week; 3 – more than 3 times a week)

fish1 – frequency of fish consumption before pregnancy (1- less than 3 times per week; 2 – 3-6 times a week; 3 – more than 6 times a week)

fish2 – frequency of fish consumption during pregnancy (1 – less than 3 times per week; 2 – 3-6 times a week; 3 – more than 6 times a week)

fruits1 – frequency of eating fruits before pregnancy (1 – less than 6 per week; 2 – 6-12 per week; 3 – more than 12 per week)

fruits2 – frequency of eating fruits during pregnancy (1 – less than 6 per week; 2 – 6-12 per week; 3 – more than 12 per week)

legumes1 – frequency of eating legumes before pregnancy (1 – less than 1 time per week; 2 – 1-3 times a week; 3 – more than 3 times a week)

legumes2 – frequency of eating legumes during pregnancy (1 – less than 1 time per week; 2 – 1-3 times a week; 3 – more than 3 times a week)

meat1 – frequency of eating meat before pregnancy (1 – less than 3 times a week; 2 – 3-6 times a week; 3 – more than 6 times a week)

meat2 – frequency of eating meat during pregnancy (1 – less than 3 times a week; 2 – 3-6 times a week; 3 – more than 6 times a week)

pastries1 – frequency of eating pastries before pregnancy (1 – less than 2 per week; 2 – 2-4 times a week; 3 – more than 4 times a week)

pastries2 – frequency of eating pastries during pregnancy (1 – less than 2 times per week; 2 – 2-4 times a week; 3 – more than 4 times a week)

performing\_sports1 – frequency of performing sports before pregnancy (1 – less than 2 times a week; 2 – 2-3 times a week; 3 – more than 3 times a week)

performing\_sports2 – frequency of performing sports during pregnancy (1 – less than 2 times a week; 2 – 2-3 times a week; 3 – more than 3 times a week)

sauces1 – frequency of using sauces before pregnancy (1 – less than 2 times per week; 2 – 2-4 times a week; 3 – more than 4 times a week)

sauces2 – frequency of using sauces during pregnancy (1 – less than 2 times per week; 2 – 2-4 times a week; 3 – more than 4 times a week)

sausages1 – frequency of consuming sausage products before pregnancy (1 - less than 1 time a week; 2 - 1-3 times a week; 3 -more than 3 times a week)

sausages2 – frequency of consuming sausage products during pregnancy (1 – less than 1 time per week; 2 – 1-3 times a week; 3 – more than 3 times a week)

skimmed\_dairy\_products1 – frequency of eating skimmed dairy foods before pregnancy (1 – less than 3 times per week; 2 – 3-6 times a week; 3 – more than 6 times a week)

skimmed\_dairy\_products2 – frequency of eating skimmed dairy foods during pregnancy (1 – less than 3 times per week; 2 – 3-6 times a week; 3 – more than 6 times a week)

smoking\_1 – smoking before pregnancy (0 – no, 1 – yes)

smoking\_2 – smoking during pregnancy (0 – no, 1 – yes)

sweet\_drinks1 – frequency of drinking sweet drinks before pregnancy (1 – less than 2 times per week; 2 – 2-4 times a week; 3 – more than 4 times a week)

sweet\_drinks2 – frequency of drinking sweet drinks during pregnancy (1 – less than 2 times per week; 2 – 2-4 times a week; 3 – more than 4 times a week)

vegetables1 – frequency of eating vegetables before pregnancy (1 – less than 6 times per week; 2 – 6-12 times a week; 3 – more than 12 times a week)

vegetables1\_raw – frequency of eating raw vegetables before pregnancy (1 – less than 6 per week; 2 – 6-12 times a week; 3 – more than 12 times a week)

vegetables2 – frequency of eating vegetables during pregnancy (1 – less than 6 per week; 2 – 6-12 times a week; 3 – more than 12 times a week)

vegetables2\_raw – frequency of eating raw vegetables during pregnancy (1 – less than 6 per week; 2 – 6-12 times a week; 3 – more than 12 times a week)

walking1 – duration of walking before pregnancy (1 – less than 30 minutes a day; 2 – 30-60 minutes a day; 3 – more than 60 minutes a day)

walking2 – duration of walking during pregnancy (1 – less than 30 minutes a day; 2 – 30-60 minutes a day; 3 – more than 60 minutes a day)

#### *CGM Trend Features That Were Selected:*

BGb240 – BG level 240 minutes before meal start (mmol/L)

BGb120 – BG level 120 minutes before meal start (mmol/L)

BGb60 – BG level 60 minutes before meal start (mmol/L)

BGb50 – BG level 50 minutes before meal start (mmol/L)

BGb40 – BG level 40 minutes before meal start (mmol/L)

BGb30 – BG level 30 minutes before meal start (mmol/L)

BGb25 – BG level 25 minutes before meal start (mmol/L)

BGb20 – BG level 20 minutes before meal start (mmol/L)

BGb15 – BG level 15 minutes before meal start (mmol/L)

BGb10 – BG level 10 minutes before meal start (mmol/L)

BGb5 – BG level 5 minutes before meal start (mmol/L)

BG0 – blood glucose level at the beginning of the meal according to the CGM signal (mmol/L)

BGb60\_to\_mean – BG 60 minutes before meal start, divided by CGM\_mean

BGRiseb240 – BG rise from 240 minutes before the meal to meal start (mmol/L)

BGRiseb120 – BG rise from 120 minutes before the meal to meal start (mmol/L)

BGRiseb60 – BG rise from 60 minutes before the meal to meal start (mmol/L)

BGTrend240 – BG trend 4 hours before the meal start (mmol/L)

BGTrend120 – BG trend 2 hours before the meal start (mmol/L)

BGTrend60 – BG trend 1 hour before the meal start (mmol/L)

iAUCb240 – Incremental AUC 240 minutes before the meal start (mmol/L\*hour)

iAUCb120 – Incremental AUC 120 minutes before the meal start (mmol/L\*hour)

iAUCb60 – Incremental AUC 60 minutes before the meal start (mmol/L\*hour)

## ACKNOWLEDGMENT

The authors would like to thank A. Dronova, E. Vasilieva, E. Vasilieva, and F. Piven, for their help in collecting the clinical data. Special thanks to the Biobanking Research Group of Almazov Medical Research Center for storing the analysis samples used in the study.

They would also like to thank the valuable comments and suggestions of the reviewers.

## REFERENCES

- [1] B. Gallwitz, "Implications of postprandial glucose and weight control in people with type 2 diabetes: Understanding and implementing the international diabetes federation guidelines," *Diabetes Care*, vol. 32, no. 2, pp. S322–S325, Nov. 2009, doi: [10.2337/dc09-S331](https://doi.org/10.2337/dc09-S331).
- [2] P. Popova, K. Castorino, E. N. Grineva, and D. Kerr, "Gestational diabetes mellitus diagnosis and treatment goals: Measurement and measures," *Minerva Endocrinol.*, vol. 41, no. 4, pp. 421–432, Dec. 2016.
- [3] E. Wender-Ozegowska, K. Wróblewska, A. Zawiejska, M. Pietryga, J. Szczapa, and R. Biczysko, "Threshold values of maternal blood glucose in early diabetic pregnancy—prediction of fetal malformations," *Acta Obstetrica et Gynecologica Scandinavica*, vol. 84, no. 1, pp. 17–25, Jan. 2005, doi: [10.1111/j.0001-6349.2005.00606.x](https://doi.org/10.1111/j.0001-6349.2005.00606.x).
- [4] S. Oviedo, J. Vehí, R. Calm, and J. Armengol, "A review of personalized blood glucose prediction strategies for T1DM patients," *Int. J. Numer. Methods Biomed. Eng.*, vol. 33, no. 6, p. e2833, Jun. 2017, doi: [10.1002/cnm.2833](https://doi.org/10.1002/cnm.2833).
- [5] A. Z. Woldaregay, E. Årsand, S. Walderhaug, D. Albers, L. Mamykina, T. Botsis, and G. Hartvigsen, "Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes," *Artif. Intell. Med.*, vol. 98, pp. 109–134, Jul. 2019, doi: [10.1016/j.artmed.2019.07.007](https://doi.org/10.1016/j.artmed.2019.07.007).
- [6] G. Alfian, M. Syafrudin, J. Rhee, M. Anshari, M. Mustakim, and I. Fahrurrozi, "Blood glucose prediction model for type 1 diabetes based on extreme gradient boosting," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 803, May 2020, Art. no. 012012, doi: [10.1088/1757-899X/803/1/012012](https://doi.org/10.1088/1757-899X/803/1/012012).
- [7] P. Popova, L. Vasilyeva, A. Tkachuk, M. Puzanov, A. Golovkin, Y. Bolotko, E. Pustozero, E. Vasilyeva, O. Li, I. Zazerskaya, R. Dmitrieva, A. Kostareva, and E. Grineva, "A randomised, controlled study of different glycaemic targets during gestational diabetes treatment: Effect on the level of adipokines in cord blood and ANGPTL4 expression in human umbilical vein endothelial cells," *Int. J. Endocrinology*, vol. 2018, pp. 1–8, Jan. 2018, doi: [10.1155/2018/6481658](https://doi.org/10.1155/2018/6481658).
- [8] E. Pustozero, P. Popova, A. Tkachuk, Y. Bolotko, Z. Yuldashev, and E. Grineva, "Development and evaluation of a mobile personalized blood glucose prediction system for patients with gestational diabetes mellitus," *JMIR mHealth uHealth*, vol. 6, no. 1, p. e6, Jan. 2018, doi: [10.2196/mhealth.9236](https://doi.org/10.2196/mhealth.9236).
- [9] E. A. Pustozero, V. Y. Chernykh, P. V. Popova, E. A. Vasyukova, A. S. Tkachuk, and Z. M. Yuldashev, "Health monitoring system for patients with gestational diabetes mellitus based on nutrition diaries and fitness bracelets," *Biomed. Eng.*, vol. 53, no. 5, pp. 305–308, Jan. 2020, doi: [10.1007/s10527-020-09931-3](https://doi.org/10.1007/s10527-020-09931-3).
- [10] E. Pustozero and P. Popova, "Mobile-based decision support system for gestational diabetes mellitus," in *Proc. Ural Symp. Biomed. Eng., Radioelectron. Inf. Technol. (USBEREIT)*, May 2018, pp. 45–48, doi: [10.1109/USBEREIT.2018.8384546](https://doi.org/10.1109/USBEREIT.2018.8384546).
- [11] E. A. Pustozero, Z. M. Yuldashev, P. V. Popova, Y. A. Bolotko, and A. S. Tkachuk, "Information support system for patients with gestational diabetes mellitus," *Biomed. Eng.*, vol. 51, no. 6, pp. 407–410, Mar. 2018, doi: [10.1007/s10527-018-9759-2](https://doi.org/10.1007/s10527-018-9759-2).
- [12] E. Pustozero, A. Tkachuk, E. Vasukova, A. Dronova, E. Shilova, A. Anopova, F. Piven, T. Pervunina, E. Vasilieva, E. Grineva, and P. Popova, "The role of glycemic index and glycemic load in the development of real-time postprandial glycemic response prediction models for patients with gestational diabetes," *Nutrients*, vol. 12, no. 2, p. 302, Jan. 2020, doi: [10.3390/nu12020302](https://doi.org/10.3390/nu12020302).
- [13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and J. Vanderplas, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.
- [14] E. Pustozero, P. Popova, Y. Bolotko, A. Tkachuk, A. Vazhenina, and E. Grineva, "Investigating detection strategy for false food consumption reports based on CGM signals in gestational diabetes patients," *Diabetes Technol. Therapeutics*, vol. 21, no. S1, p. 153, Feb. 2019, doi: [10.1089/dia.2019.2525.abstracts](https://doi.org/10.1089/dia.2019.2525.abstracts).
- [15] J. Josse, N. Prost, E. Scornet, and G. Varoquaux, "On the consistency of supervised learning with missing values," Jun. 2020, *arXiv:1902.06931v3*. [Online]. Available: <https://arxiv.org/abs/1902.06931v3>
- [16] J. H. Friedman, "Stochastic gradient boosting," *Comput. Statist. Data Anal.*, vol. 38, no. 4, pp. 367–378, 2002.
- [17] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794, doi: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785).
- [18] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, Montreal, QC, Canada, Dec. 2018 pp. 6639–6649.
- [19] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, Long Beach, CA, USA, Dec. 2017, pp. 3149–3157.
- [20] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, Long Beach, CA, USA, Dec. 2017, pp. 4765–4774.
- [21] D. Zeevi et al., "Personalized nutrition by prediction of glycemic responses," *Cell*, vol. 163, no. 5, pp. 1079–1094, Nov. 2015, doi: [10.1016/j.cell.2015.11.001](https://doi.org/10.1016/j.cell.2015.11.001).
- [22] H. Mendes-Soares, T. Raveh-Sadka, S. Azulay, K. Edens, Y. Ben-Shlomo, Y. Cohen, T. Ofek, D. Bachrach, J. Stevens, D. Colibaseanu, L. Segal, P. Kashyap, and H. Nelson, "Assessment of a personalized approach to predicting postprandial glycemic responses to food among individuals without diabetes," *JAMA Netw. Open*, vol. 2, no. 2, Feb. 2019, Art. no. e188102, doi: [10.1001/jamanetworkopen.2018.8102](https://doi.org/10.1001/jamanetworkopen.2018.8102).
- [23] S. E. Berry et al., "Human postprandial responses to food and potential for precision nutrition," *Nature Med.*, vol. 26, no. 6, pp. 964–973, Jun. 2020, doi: [10.1038/s41591-020-0934-0](https://doi.org/10.1038/s41591-020-0934-0).
- [24] T. Zhu, K. Li, P. Herrero, J. Chen, and P. Georgiou, "A Deep Learning Algorithm For Personalized Blood Glucose Prediction," in *Proc. KHD@IJCAI*, Stockholm, Sweden, 2018, p. 5.
- [25] K. Li, J. Daniels, C. Liu, P. Herrero, and P. Georgiou, "Convolutional recurrent neural networks for glucose prediction," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 603–613, Feb. 2020, doi: [10.1109/JBHI.2019.2908488](https://doi.org/10.1109/JBHI.2019.2908488).
- [26] I. Rodriguez-Rodriguez, J. V. Rodriguez, I. Chatzigiannakis, and M. A. Zamora, "On the possibility of predicting glycaemia 'on the fly' with constrained iot devices in type 1 diabetes mellitus patients," *Sensors*, vol. 19, pp. 4482–4496, Jan. 2019.
- [27] A. B. Evert, J. L. Boucher, M. Cypress, S. A. Dunbar, M. J. Franz, E. J. Mayer-Davis, J. J. Neumiller, R. Nwankwo, C. L. Verdi, P. Urbanski, and W. S. Yancy, "Nutrition therapy recommendations for the management of adults with diabetes," *Diabetes Care*, vol. 37, no. 1, pp. S120–S143, Jan. 2014, doi: [10.2337/dc14-S120](https://doi.org/10.2337/dc14-S120).
- [28] J. A. L. Calbet and D. A. MacLean, "Plasma glucagon and insulin responses depend on the rate of appearance of amino acids after ingestion of different protein solutions in humans," *J. Nutrition*, vol. 132, no. 8, pp. 2174–2182, Aug. 2002, doi: [10.1093/jn/132.8.2174](https://doi.org/10.1093/jn/132.8.2174).
- [29] M. Claessens, W. Calame, A. D. Siemensma, M. A. van Baak, and W. H. M. Saris, "The effect of different protein hydrolysate/carbohydrate mixtures on postprandial glucagon and insulin responses in healthy subjects," *Eur. J. Clin. Nutrition*, vol. 63, no. 1, pp. 48–56, Jan. 2009, doi: [10.1038/sj.ejcn.1602896](https://doi.org/10.1038/sj.ejcn.1602896).
- [30] M. C. Gannon, F. Q. Nuttall, A. Saeed, K. Jordan, and H. Hoover, "An increase in dietary protein improves the blood glucose response in persons with type 2 diabetes," *Amer. J. Clin. Nutrition*, vol. 78, no. 4, pp. 734–741, Oct. 2003, doi: [10.1093/ajcn/78.4.734](https://doi.org/10.1093/ajcn/78.4.734).
- [31] H. J. Dranse, T. M. Z. Waise, S. C. Hamr, P. V. Bauer, M. A. Abraham, B. A. Rasmussen, and T. K. T. Lam, "Physiological and therapeutic regulation of glucose homeostasis by upper small intestinal PepT1-mediated protein sensing," *Nature Commun.*, vol. 9, no. 1, p. 1, Mar. 2018, doi: [10.1038/s41467-018-03490-8](https://doi.org/10.1038/s41467-018-03490-8).



- [32] E. Diamanti-Kandarakis, J.-P. Baillargeon, M. J. Iuorno, D. J. Jakubowicz, and J. E. Nestler, "A modern medical quandary: Polycystic ovary syndrome, insulin resistance, and oral contraceptive pills," *J. Clin. Endocrinology Metabolism*, vol. 88, no. 5, pp. 1927–1932, May 2003, doi: [10.1210/jc.2002-021528](https://doi.org/10.1210/jc.2002-021528).
- [33] A. A. Adeniji, P. A. Essah, J. E. Nestler, and K. I. Cheang, "Metabolic effects of a commonly used combined hormonal oral contraceptive in women with and without polycystic ovary syndrome," *J. Women's Health*, vol. 25, no. 6, pp. 638–645, Jun. 2016, doi: [10.1089/jwh.2015.5418](https://doi.org/10.1089/jwh.2015.5418).



the development of mobile-based data-driven decision support systems for patients with endocrine diseases. He is also a member of IEEE Russia (Northwest) Section.

**EVGENII A. PUSTOZEROV** (Member, IEEE) received the Ph.D. degree in biomedical engineering in Saint Petersburg, in 2017. He is currently working as an Associate Professor with Saint Petersburg State Electrotechnical University and as a Fellow Researcher with the Almazov National Medical Research Center. His professional interests lay on the edge of biomedical engineering, data science, and medical informatics. His current grant project at the Research Center is related to



with the Almazov National Medical Research Center, where she is also involved in the study on gestational diabetes supported by Russian Science Foundation. She has coauthored 38 articles. Her research interests include gestational diabetes mellitus, polycystic ovary syndrome, and thyroid diseases.

**ALEKSANDRA S. TKACHUK** was born in Baltiysk, Russia, in 1988. She graduated the doctor's degree from the First Saint Petersburg State Medical University. She is currently pursuing the master's degree with the Department of Endocrinology, Almazov National Medical Research Center, St. Petersburg, Russia. She is also a Doctor with the Department of Endocrinology, Almazov National Medical Research Center. Since 2014, she has been a Research Associate



gestational diabetes mellitus, polycystic ovary syndrome, thyroid gland, and pregnancy.

**ELENA A. VASUKOVA** was born in Ufa, Russia, in 1993. She graduated from the First Pavlov State Medical University of St. Petersburg, in 2017. From 2017 to 2019, she studied in clinical residency in the specialty "Endocrinology" at the Department of Internal Medicine, Institute of Medical Education, Almazov National Medical Research Center. Since 2019, she has been working with the Laboratory of Endocrine Disorders in pregnant women. Her research interests include



gestational diabetes mellitus, iodine deficiency among children and pregnant women, and polycystic ovary syndrome and obesity.

**ANNA D. ANOPOVA** studied at Saint-Petersburg State Pediatric Medical University. She graduated from the residency of the Almazov National Medical Research Center, in 2020. She received the degree of endocrinologist. She is currently pursuing the Ph.D. degree in endocrinology with the Almazov Medical Research Center under the supervision of P. Popova, preparing articles in international journals and participating in conferences. Since 2015, she broadens her interests in



laboratory Diagnostics. She has also been working with the Laboratory of Endocrine Diseases in Pregnancy since 2019. She is interested in such scientific fields as gestational diabetes, polycystic ovary syndrome, thyroid, and pregnancy.

**MARIA A. KOKINA** was born in Norilsk, in 1990. She graduated from Mechnikov North-West State Medical University, the Faculty of Preventive medicine. She had been passing residency from 2013 to 2015 with a course of clinical laboratory diagnostics at The Federal State Public Enterprise Nikiforov's All-Russian Center for Emergency and Radiation Medicine. Since 2015, she has been working with the Almazov National Medical Research Center as a Doctor of Clinical Laboratory



and pregnancy.

**INGA V. GORELOVA** received the Ph.D. degree in medicine from Mechnikov North-West State Medical University, Russia, in 2010. Since 2017, she has been the Head of the Research Laboratory of Assisted Reproductive Technologies, Almazov National Medical Research Center, St. Petersburg, Russia. Her research interests include the problems of infertility in men and women, gynecological endocrinology, and polycystic ovary syndrome and pregnancy.



serves as the principal investigator in studies on gynaecology and participates in the management of pregnant women with severe comorbidities.

**TATIANA M. PERVUNINA** graduated from Saint Petersburg State Pediatric Medical University. She received the medical sciences and doctor of medical sciences degrees in 2003 and 2019, respectively. She has been working as the Head of the Institute of Perinatology and Pediatrics, Almazov National Medical Research Center, St. Petersburg, Russia, since 2017. Her research interests include pediatric cardiology, orphan disorders, and searching for effective early omic-biomarkers. She also



**ELENA N. GRINEVA** graduated the degree (*summa cum laude*) from the First Leningrad Medical Institute (renamed First Saint-Petersburg State Medical University), in 1983. She received the Ph.D. degree in 1991 with the thesis "The activity of nucleolar organizer thyrocytes in patients with diffuse toxic goiter, chronic autoimmune thyroiditis, and thyroid nodules." She was the first endocrinologist in St. Petersburg to master the fine-needle aspiration biopsy of the thyroid gland and cytolog-

ical diagnosis. She made it a day-to-day procedure having performed about 10 000 thyroid biopsies. She has trained many surgeons and endocrinologists from St. Petersburg and other cities throughout Russia to perform fine-needle aspiration biopsy of the thyroid gland, as well as to diagnose diseases of the thyroid cytology. In 2004, she defended her doctoral thesis on diagnostics and management of thyroid nodules. She is also the coauthor of a number of monographs and textbooks on *Endocrinology*, *The Pituitary*, *The Adrenal Gland*, and *Thyroid Diseases* (published in Russian). She holds an official position of a leading authority in St. Petersburg and Russia on thyroid and pituitary disease. Since 2008, she has been successfully running different research projects on diagnosis and treatment of patients with diabetes mellitus, pituitary and adrenal glands diseases (including pregnant women with endocrine diseases) at the Institute of Endocrinology which is a department of Almazov National Medical Research Center. She is currently a Director of the Institute of Endocrinology, Almazov National Medical Research Center. She has also been a Professor at Saint Petersburg Pavlov State Medical University teaching internal medicine and endocrinology since 1999. She is also a member of Russian Association of Endocrinologists, European Association of Neuroendocrinologists, and European Society of Endocrinology.



**POLINA V. POPOVA** received the M.D. and Ph.D. degrees from Saint Petersburg Pavlov State Medical University, Saint Petersburg, Russia, where she conducted a study «The role of weight reduction and the use of biguanids in correcting the risk factors of cardiovascular diseases and menstrual function in women with polycystic ovary syndrome». Clinical expertise includes diabetes, obesity, polycystic ovary syndrome, and endocrine diseases during pregnancy. She is currently the

Head of the Research Laboratory of Endocrine Diseases in Pregnancy, an Associate Professor with the Department of Endocrinology, Almazov National Medical Research Center, and an Associate Professor at the Department of Faculty Therapy, Saint Petersburg Pavlov State Medical University. She has published over 50 scientific articles in national and international journals. Her research interests include developmental origin of health and disease within the field "offspring from pregnancies with diabetes", the use of telemedicine in GDM patients, prediction of postprandial glycemic response in women with GDM, prediction of GDM, personalized treatment, the impact of diets on microbiome in women with PCOS, and «thyroid and pregnancy». Her lab is participating in several international projects within the above mentioned scope.

• • •