Computers in Biology and Medicine 171 (2024) 108154

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

An ensemble machine learning approach for the detection of unannounced meals to enhance postprandial glucose control

Muhammad Ibrahim [a], Aleix Beneyto [a], Ivan Contreras [a], Josep Vehi [a,b,*]

[a] Modeling, Identification and Control Engineering Laboratory (MICELab), Institut d'Informàtica i Aplicacions, Universitat de Girona, Girona, Spain
[b] Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), Madrid, Spain

ARTICLE INFO

Keywords:
Artificial pancreas
Type 1 diabetes
Meal detection
Ensemble technique
Neural network
Machine learning

ABSTRACT

Background: Hybrid automated insulin delivery systems enhance postprandial glucose control in type 1 diabetes, however, meal announcements are burdensome. To overcome this, we propose a machine learning-based automated meal detection approach;
Methods: A heterogeneous ensemble method combining an artificial neural network, random forest, and logistic regression was employed. Trained and tested on data from two in-silico cohorts comprising 20 and 47 patients. It accounted for various meal sizes (moderate to high) and glucose appearance rates (slow and rapid absorbing). To produce an optimal prediction model, three ensemble configurations were used: logical AND, majority voting, and logical OR. In addition to the in-silico data, the proposed meal detector was also trained and tested using the OhioT1DM dataset. Finally, the meal detector is combined with a bolus insulin compensation scheme;
Results: The ensemble majority voting obtained the best meal detector results for both the in-silico and OhioT1DM cohorts with a sensitivity of 77%, 94%, 61%, precision of 96%, 89%, 72%, F1-score of 85%, 91%, 66%, and with false positives per day values of 0.05, 0.19, 0.17, respectively. Automatic meal detection with insulin compensation has been performed in open-loop insulin therapy using the AND ensemble, chosen for its lower false positive rate. Time-in-range has significantly increased 10.48% and 16.03%, time above range was reduced by 5.16% and 11.85%, with a minimal time below range increase of 0.35% and 2.69% for both in-silico cohorts, respectively, compared to the results without a meal detector;
Conclusion: To increase the overall accuracy and robustness of the predictions, this ensemble methodology aims to take advantage of each base model's strengths. All of the results point to the potential application of the proposed meal detector as a separate module for the detection of meals in automated insulin delivery systems to achieve improved glycemic control.

1. Introduction

Despite technological advancements, achieving fully automatic insulin delivery systems (AID) remains a challenge due to difficulties in controlling postprandial blood glucose (BG) levels in people with type 1 diabetes (T1D). Currently, available AID systems are considered hybrid because they require patients to announce their meals and estimate their carbohydrate (CHO) intake before mealtime [1]. The announcement of the meal is used to trigger the feedforward control action, compensating for the postprandial glucose rise. However, people with T1D often struggle to make meal announcements, which results in missing out on the associated meal insulin bolus.

Unannounced meals (UAM) and subsequent missed insulin bolus injections prior to meals are associated with an increase in glucose level [2,3]. Studies show that the amount of missed meal announcements is directly linked with the level of HbA1c and constantly being at a higher level of HbA1c could have a negative impact on life over time [4,5]. Factors such as stress, diabetes distress, or any other illness could be attributed to forgetting the announcement of meals [6,7]. In particular, the ratio of missed meals is higher in adolescents compared to other age groups [8,9]. In one study, it has been observed that one-third of the adolescent population misses the insulin bolus of meals for more than 15% of their meals [10]. Taking into account all these facts, the need for automated meal detector (MD)-equipped AID systems is adequately justified to detect UAM and regulate postprandial BG.

Multiple studies have shown that the performance of AID systems deteriorates in terms of glycemic control when the announcements of

* Corresponding author at: Modeling, Identification and Control Engineering Laboratory (MICELab), Institut d'Informàtica i Aplicacions, Universitat de Girona, Girona, Spain.
E-mail address: josep.vehi@udg.edu (J. Vehi).

https://doi.org/10.1016/j.compbiomed.2024.108154
Received 5 September 2023; Received in revised form 2 February 2024; Accepted 12 February 2024
Available online 19 February 2024

larger meals (greater than 30 g) are missed [11–13]. The performance of the AID system in managing glucose levels is directly proportional to the size of the UAM, and larger meals result in poorer control. In fact, the glucose rate of appearance (RA) of UAM is also vital for the better performance of the AID system. As the meal appears in the blood much faster than the pharmaco-kinetics and dynamics of current insulin analogues.

Several methods have been used for meal detection in recent years. These include heuristic threshold-based, machine learning (ML)-based, and control systems theory-based approaches. In subsequent paragraphs, most of the meal detection methods are discussed.

A majority voting ensemble approach has been used by [14], combining four methods (Kalman filter (KF), backward difference, second derivative of the BG, and KF+backward difference), the controller receives a meal flag if at least two methods detect a meal. An unscented KF is used by [15], which is an extension of Bergman's minimal model that tracks glucose level, rate of change, and disturbance parameters. A meal detection plus CHO estimation method has been proposed by [16,17] using filtered continuous glucose monitoring (CGM) sensor signal derivatives and a fuzzy logic system. However, relying solely on the rate of change and fixed thresholds for meal detection may increase false positives (FPs) due to glucose level variability. Filtering methods can reduce FPs but also introduce detection delay and reduce sensitivity. Thresholds, based on specific parameters of the training dataset, such as sensor noise and insulin sensitivity, may lead to subpar performance if conditions change during deployment.

Some model-based techniques search for outliers in the glucose curve to detect meals. A probabilistic approach proposed by [18], compares the observed glucose signal with the expected signal for meal detection. It also estimates meal shape and glucose RA. Variable state dimension with an extended KF is utilised by [19], to forecast the glucose curve and estimate meal size using a least-squares approach once the upper threshold is exceeded. Similarly, [20] employs an unscented KF with two CGM sensors, detecting a meal when both signals exceed the 95% prediction interval. A study by [21], compares trajectories from the minimal model and the CGM sensor, raising a meal flag if the difference exceeds a set threshold. An extended super-twisting-based is used in [22], replacing the Euler approximation of the glucose derivative with the KF to reduce FPs by minimising oscillations and providing a smoother estimation.

Data-driven techniques are also used for meal detection. The method proposed by [23] applies logistic regression to estimate UAM probability and determine insulin bolus based on total daily insulin. A deep neural network with long short-term memory (LSTM) and quantile regression are used for UAM detection and estimation in [24]. The [25] uses LSTM-based recurrent neural networks to detect habitual disturbances like meals and exercise. [26] introduces a multioutput neural network with branches for binary meal detection and multiclass meal estimation. Supervised learning models (naive Bayes, decision trees, k-nearest neighbour, linear support vector machine (SVM), discriminant analysis, SVM-Gaussian) detect meals while forecasting models (neural network and SVM) estimate meals in [27]. Whereas, [28] employs unsupervised anomaly detection (isolation forest) to trigger a meal alert when the anomaly score exceeds the threshold.

Some of the limitations that are observed in the literature in the domain of meal detection include high detection times, and in most cases, *in-silico* data has been used with minimal meal models. Although the MD should be able to detect the meal as soon as possible to avoid post-prandial hypoglycemia, the basic meal model cannot represent the challenging real-life scenario. In addition, the methods in the literature have not taken the glucose RA of the UAM into consideration, which we think is an essential factor while designing a method for meal detection. This could be elaborated with the example of two meals with the same amount of CHO but different glucose RAs; both would have different dynamics. The rapid-absorbing meal excursion rate in BG would be higher than the slow-absorbing meal. The MD should be able to work equally well for those various kinds of meals. The proposed MD aims to detect UAM as early as possible after the commencement of the meal, also considering the different compositions of the UAM.

In the proposed work, we aim to employ an ML-based ensemble technique for UAM detection. The key contributions involve introducing realistic scenarios with various forms of variability, including circadian variability, meal randomness, and diverse meal compositions, to mimic real-life conditions. The ensemble approach shows a significant improvement compared to individual models. The proposed model is validated using two in-silico and one real patient cohorts, demonstrating its generalisability across different populations. Additionally, an *in-silico* testing is conducted to assess the performance of the MD in a real-time simulation environment.

The paper is organised as follows: Section 2 provides a detailed description of the proposed methodology for the detection of UAM. All the results acquired are summarised in Section 3. Section 4 starts with a detailed discussion on the comparison of our proposed approach with the most recent state-of-the-art methods in the literature. Finally, a concluding remark and future directions are mentioned in Section 5.

2. Materials and methods

2.1. Problem statement

This study aims to detect UAM and control the postprandial BG with an automated feed-forward action. In this work, an UAM is defined as a meal (of CHO amount equal to or greater than 30 g) that is consumed without estimating the amount of CHO and without informing the system within a 30-min window.

Framing the problem of detecting UAM as a classification problem using supervised learning approaches requires, first of all, properly labelled data. Where each UAM should be labelled at the exact location. A real patient dataset with such an attribute is quite rare to acquire due to the fact that pre-meal boluses are missed, which could put a patient in danger. The best alternative is to generate such a dataset using a T1D simulator, where each meal could be accurately labelled. In this work, the UAM are labelled for two hours of postprandial period.

To get the required data with more realistic scenarios, our previously presented mixed meal library is used in this work [29,30]. The library contains 7 h (with a sampling time of 5 min) of the postprandial glucose RA profiles for each meal. The library has a total of 54 meals that are classified into four classes by Kolmogorov–Smirnov distance: small, medium, large, and fast.

However, for this study, the area under the curve (AUC) for each meal in the library was analysed. A fixed threshold for the AUC to reach 50% of the total AUC was set. The meals taking less than 2 h to get to the 50% of the AUC were classified as rapid-absorbing meals, whereas the meals taking more than 2 h were classified as slow-absorbing meals. The slow-absorbing and rapid-absorbing meals were further sub-classified into three categories based on the amount of CHO. The amount of CHO greater than 10 g and less than 30 g is considered small size; 30 g $\leq$ CHO $<$ 60 g is considered moderate size; and 60 g $\leq$ CHO $<$ 130 g is classified as high size. Based on these classifications of meals, six possible combinations can be formed, which include small size & rapid-absorbing, small size & slow-absorbing, moderate size & rapid-absorbing, moderate size & slow-absorbing, high size & rapid-absorbing, and high CHO & slow-absorbing. Datasets for training and testing were created based on these combinations, which are discussed in the subsequent Section 2.2.

An ensemble technique is used in this work by combining the prediction of three ML models, which include, artificial neural network (ANN), random forest (RF), and logistic regression (LR). Each step involved in problem framing, a brief introduction to ensemble learning, and the utilisation of the proposed algorithm in a simulation environment is extensively discussed in the following sub-sections,

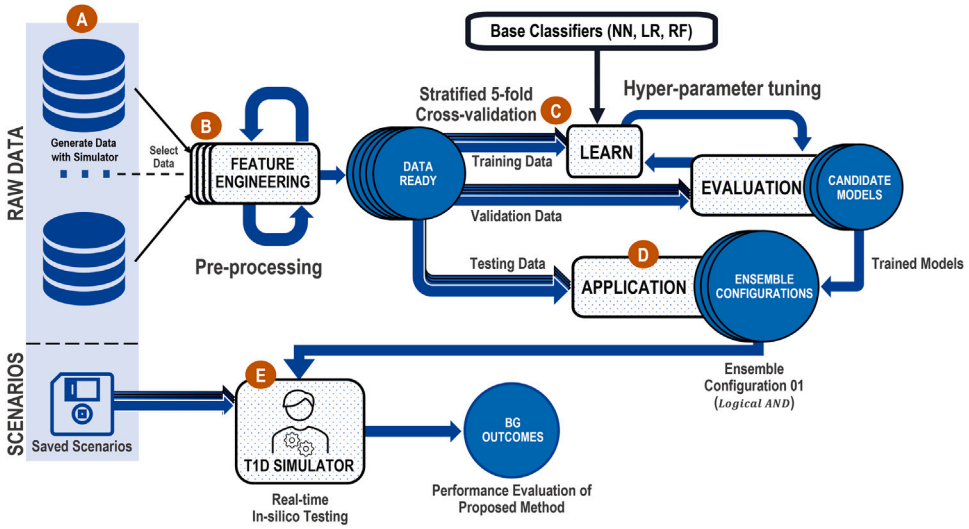

Fig. 1. General schematic diagram of the proposed method implementation: (A) simulated data are generated, (B) feature engineering is performed to extract meaning features from the raw data, (C) the training datasets were then used for training and hyperparameter tuning of the models, (D) the trained models' individual and ensemble performance are then evaluated using unseen testing datasets, (E) the best-performing ensemble configuration is then selected for the real-time in-silico testing and finally, the performance of the proposed method is observed in terms of BG outcomes.

ranging from datasets (training and testing), feature extraction, and *in-silico* testing. The general flow of the implementation of the proposed method is presented in Fig. 1. First, raw datasets were generated, which were then subjected to pre-processing to extract meaningful features. Next, training and tuning of the three models (ANN, RF, and LR) were performed and their individual and ensemble performances were evaluated. Finally, we selected one of the ensemble configuration for real-time in-silico testing.

2.2. Data and preprocessing

The in-silico datasets were obtained from simulations performed using two T1D simulators: (1) a customised simulator based on the Dalla Man model [31]; and (2) a novel simulator based on the Hovorka model [30]. In addition, the proposed approach has also been tested on a real patients' dataset (OhioT1DM) comprising retrospective data from 12 patients [32]. The description of the testing datasets is presented in Supplementary Table 1, while the details of the in-silico datasets as well as real patients' data are provided in subsequent subsections.

2.2.1. Virtual patients

Two in-silico cohorts were utilised in this study. Simulation cohort 1 (SC1) consisted of 20 patients (10 adolescents and 10 adults) using a modified simulator based on [31], while simulation cohort 2 (SC2) comprised 47 patients using a novel simulator based on the Hovorka model [30]. Various variability factors were incorporated into the dataset generation scenarios, including meal randomness (with a meal-time standard deviation of 20 min and meal-content (CV = 10%), respectively), meals from a mixed meal library [29], and intra-patient variability on insulin absorption and insulin sensitivity [30,33]. In addition, to mimic the real-life scenario, CHO counting uncertainty was incorporated with a normal distribution (±40%), and it is important to acknowledge that not all meals would be considered unannounced; thus, the scenarios for generating training and testing datasets contain randomly selected meals that were set as unannounced (no pre-meal insulin bolus were administered for those meals).

Training datasets: The scenario for the training dataset includes four meals per day with breakfast at 07:00 h containing 30–50 g of CHO, lunch at 14:00 h (60–90 g), snack of 15–20 g at 17:00 h, and dinner at 21:00 h (40–70 g). The training dataset provides per-patient data for 120 days with 192 UAM (no bolus was injected for those meals). The datasets were generated using open-loop (OL) insulin therapy with a constant yet patient-specific basal insulin rate.

Only moderate and high-size meals (slow and rapid absorbing) were randomly selected to be unannounced because the unannounced small-size meals do not deteriorate glycemic control [11]. It was ensured that each type of meal should have an equal number of UAM, i.e., 48 UAM each. For the classification task, the data need to be labelled. Thus, the whole data is divided into two classes (class 0 and class 1). Class 0 consists of the data samples where no meal has been consumed or where the meals consumed are announced. Class 1 consists of the data samples of a 2-h post-prandial period of UAM (meal has been consumed without announcing it to the system).

Testing datasets: The testing datasets were generated to evaluate the performance of the proposed method. In testing datasets, each patient contained data for 30 days and OL insulin therapy was used with a constant (patient-specific) basal insulin rate. The scenario included: four meals per day with varying CHO amounts and meal times. The coefficient of variance for CHO size was ±10%, and the standard deviation of meal time was ±20 min. The first meal (breakfast) of 30–50 g starts at 07:00 h, lunch of 60–90 g at 14:00 h, snack of 15–25 g at 17:00 h, and dinner of 40–70 g at 21:00 h. The meal protocol included a total of 120 meals per patient for 30 days scenario. Of these, 48 meals were UAM.

The testing datasets were categorised, considering the size and rate of absorption of the UAM. Each of the testing datasets included a total of 2400 and 5640 meals, including 960 and 2256 UAM of various sizes and rates of absorption for both virtual cohorts, respectively. **Testing dataset 01**, in particular, contained UAM with moderate-sized and slow-absorbing glucose responses. **Testing dataset 02** consisted of moderate-sized rapid-absorbing UAM. **Testing dataset 03** had high-sized rapid-absorbing UAM. Lastly, **testing dataset 04**, included high-sized slow-absorbing UAM.

2.2.2. Real patients

Data from the OhioT1DM dataset has also been used [32]. All the 12 patients in the OhioT1DM dataset were on insulin pump therapy, wore Medtronic 530G or 630G insulin pumps, and used Medtronic Enlite CGM sensors. Each patient's data includes all the information required for this study, such as CGM glucose level with a sampling time equal to 5 min; insulin doses, both bolus and basal insulin; and self-reported meal times with CHO estimation.

For this real dataset, we performed some data pre-processing including filling of missing values and identification or labelling UAM events. First, a linear interpolation is used only for glucose gaps of equal to or less than 20 min, as per the recommendation of the recently published consensus guidelines for ML practitioners in the field of diabetes [34]. After interpolation, each day's data was checked if any missing values still existed, eliminating days that still had more than 10% of missing values. The remaining dataset included a cumulative sum of 349 complete days of insulin and BG data for the whole cohort. Next up, the self-reported meals were analysed, if there was any bolus insulin ±30 min from the meal time was considered as an announced meal or else labelled as an UAM. On average, 23% UAM were observed in the whole dataset.

2.2.3. Feature engineering

After obtaining the required data, the feature engineering module is used to calculate features from CGM sensor readings and insulin data. To describe the underlying informative content of the raw signal concisely, various features are introduced. A total of 14 features were extracted, of which 12 are based on CGM readings and the remaining 2 are based on insulin data. A sliding window method that takes into account the previous 45 min of data is used to extract features at each time interval. This sliding window can be thought of as a dynamic segment that advances along the timeline while continually retaining the last 45 min of data. By using this technique, the relevant information within each particular time period is recorded and processed to calculate features, enabling the detection of patterns and trends across time. All the calculated features are mentioned below:

(a) *Current CGM value:* As features are calculated at each timestamp, the first feature is the CGM value at the current timestamp.

(b) *CGM mean:* The mean is calculated using 45 min (min) of previous data, i.e., the mean of the last 9 CGM samples.

$$f_{mean} = \frac{(G_{i-1} + G_{i-2} + \cdots + G_{i-9})}{9} \tag{1}$$

(c) *Rate of change of CGM:* The Next six features are based on the rate of change in the CGM readings for the last 30 min. The rate of change is calculated at each sample for the previous 30 min of data.

$$f_{ROC_i} = \frac{(G_i - G_{i-1})}{t_s} \tag{2}$$

where $i = 1, 2, 3, 4, 5, 6$ and each value of i represents the following rate of change:

ROC_1: Rate of change in CGM from last 30-25 min
ROC_2: Rate of change in CGM from last 25-20 min
ROC_3: Rate of change in CGM from last 20-15 min
ROC_4: Rate of change in CGM from last 15-10 min
ROC_5: Rate of change in CGM from last 10-05 min
ROC_6: Rate of change in CGM from last 05-00 min

(d) *CGM excursion:* The excursion is determined by finding the difference between the current CGM value and the value 30 min prior to that.

$$f_{exc} = (G_i - G_{i-6}) \tag{3}$$

(e) *Area under the curve (AUC):* The AUC of CGM is defined as

$$f_{AUC} = 70t - \int_{i-p}^{i} vec\_under70(t)dt \tag{4}$$

where $p = 9$ samples or 45 min and

$$vec\_under70(t) = \begin{cases} 70, & \text{if } CGM(t) \geq 70 \\ CGM(t), & \text{otherwise} \end{cases}$$

(f) *Risk index:* The risk indexes; low BG index (LBGI) and high BG index (HBGI) are also computed. These metrics are useful to assess the risk of hypoglycemia and hyperglycemia. The formula for each risk index is mentioned below:

i. LBGI

$$f_{LBGI} = \frac{1}{n} \times \sum (10 \times (fBG_i)^2) \tag{5}$$
where, $fBG_i = min(0, 1.509 \times (log(BG_i)^{1.084} - 5.381))$

ii. HBGI

$$f_{HBGI} = \frac{1}{n} \times \sum (10 \times (fBG_i)^2) \tag{6}$$
where, $fBG_i = max(0, 1.509 \times (log(BG_i)^{1.084} - 5.381))$

(g) *Pre-meal insulin bolus:* Insulin meal bolus is also included in the feature space. For an announced meal, this feature has some value depending on the amount of CHO announced to the system. While for a UAM and for the rest of the samples the value of this feature is zero.

(h) *Insulin-on-board (IOB):* The amount of injected insulin that is still present in the body is known as IOB. It is based on the two-compartment dynamical model [35,36].

$$\begin{aligned} C_1(t) &= u(t) - k_{DIA}C_1(t) \\ C_2(t) &= k_{DIA}(C_1(t) - C_2(t)) \\ \widehat{IOB}(t) &= C_1(t) + C_2(t) \end{aligned} \tag{7}$$

where $u(t)$ is the insulin delivery, $C1$ and $C2$ are the two compartments, and $k_{DIA} = 0.013$ is a populational constant representing the duration of insulin action of a 6 h curve.

2.3. Ensemble learning

Ensemble learning, a prominent technique in ML, harnesses the collective intelligence of multiple classifiers to effectively improve the overall performance and accuracy of diverse learning tasks. By consolidating the individual predictions of these classifiers, it enables the generation of robust and reliable prediction [37]. The individual classifiers employed in the ensemble method are referred to as base classifiers. Depending on the selection of these base classifiers, the ensemble method framework can be broadly categorised into two types: homogeneous and heterogeneous [38,39]. In the case of homogeneous ensemble methods, the same type of base classifiers is used, while the heterogeneous ensemble method employs different types of base classifiers.

The second step after selecting the base learners is defining the method to combine the classifiers. Three major aggregating methods are mentioned here: bagging, boosting, and stacking. In bagging, the base classifiers are trained independently from each other in a parallel manner, and then the predictions of each classifier are combined to get the final prediction [40]. The aggregation of the output of each model depends on the type of problem dealt with; for regression problems, the outputs can be simply averaged, while for classification problems, each model output can be considered a vote, and then majority voting is used for class prediction. On the other end, a sequential training technique is used in the case of boosting, where each subsequent classifier depends on the weights of the previously trained classifier and the predictions of all the base classifiers are combined by weighted majority voting [41]. In the stacking method, the base classifiers are trained in parallel and then combined by training the meta-model to output the final prediction [42,43].

This study adopts a heterogeneous approach by utilising three base classifiers: ANN, RF, and LR. The training of these base classifiers is conducted independently and concurrently using a bagging strategy. To generate the final prediction, the outputs of each classifier are combined through three distinct configurations, which will be further elaborated in subsequent subsections.

2.3.1. Base models hyperparameters

The three selected ML algorithms ANN, RF, and LR, all of which were chosen from the Python (*version* 3.10) library Sklearn (*version* 1.1.2) [44]. The Sklearn ensemble module provided the "RandomForestClassifier", followed by the selection of "MLPClassifier" from the Sklearn neural network module, and finally, "LogisticRegression" from the linear_model module of Sklearn. To optimise classifier performance, hyper-parameter tuning was conducted using Sklearn's "RandomisedSearchCV".

Starting from the LR model, the best possible parameters selected after tuning are a "liblinear" solver, which usually performs better for high-dimensional datasets or solving the problems of large-scale classification [45]. To avoid the problem of over-fitting, the weights need to be penalised to prevent any particular weight from growing too high; this approach is known as regularisation. In our case, the "*l2*" regularisation (*reg*) approach is used, which penalises logistic regression by minimising the cost function ($J(w)$) presented in Eq. (8). In the equation, the first part represents the *reg* term, and the subsequent term is the log loss (*log_loss*). In which the weight vector is denoted by "w" and a small "c" term represents the bias added to the model, whereas, "y_i" and X_i are the true label and the input features vector for the ith example. The hyperparameter "C", known as the inverse of the regularisation strength, is selected as equal to 1.6237 from a numpy logspace ranging from −4 to 4 by using RamdomisedSearchCV. The maximum number of iterations "max_iter" required for the solver to converge was set to 5000.

$$J(w) = min_{w,c}\frac{1}{2}w^T w + C\sum_{i=1}^{n} log(exp(-y_i(X_i^T w + c)) + 1) \tag{8}$$

In this work, an ANN model with a five-layer architecture is employed. The hidden layers are comprised of 50, 100, and 50 neurons, respectively, utilising the hyperbolic tangent ($f(x) = tanh(x)$) activation function. The model implements an adaptive learning rate mechanism, it keeps the learning rate constant to the initial rate as long as the training loss keeps decreasing [44]. The maximum number of iterations is set to 3500. The Adam optimiser algorithm is used for the optimisation process which works well for larger datasets. While the model incorporates a regularisation term with a value of 0.05, denoted as "alpha". The Sklearn "MLPClassifier" uses a cross-entropy loss function (CE), presented in Eq. (9). In the equation, the sum is over all training inputs denoted as "x", with "y" representing the corresponding desired output. Here, "n" refers to the total number of training data items and "a" is the output from the neuron.

$$CE = -\frac{1}{n}\sum_{x}[y\ln a + (1-y)\ln(1-a)] \tag{9}$$

Lastly, this work employs the Sklearn ensemble RF model, which combines multiple decision trees to enhance the accuracy and robustness of predictions. The optimal parameter values for the RF model, determined by the RandomSearchCV algorithm, are as follows: n_estimators which represent the number of trees in the forest was set to 48 decision trees; min_samples_split of 2, which requires a minimum of two data points for node splitting; and min_samples_leaf of 2, which limits the minimum number of data points allowed on the leaf node. The max_features is set to auto, which is the maximum number of features considered for splitting the node, and the maximum depth of each decision tree is limited to 4. Furthermore, the bootstrap method is enabled, which involves sampling data points with replacement.

2.3.2. Training and testing of base models

For *in-silico* data, we employ the ensemble configuration by selecting the best-performing models: the models having high true positives (TPs), fewer FPs, and less detection time. The base models are independently trained and tuned (as discussed in Section 2.3.1) on the training dataset. A stratified fivefold cross-validation strategy is used during the training of the model. The trained models underwent testing on aggregated testing datasets. Each testing dataset for SC1 and SC2 contained data from 20 and 47 patients, spanning over 600 and 1410 days, and had a total of 2400 and 5640 meals, of which 960 and 2256 were UAM, respectively.

In the case of the OhioT1DM dataset, we adopted a personalised approach by training individualised models for each patient. Due to the limited amount of data and to effectively use it for both training and testing, we employed a "leave-one-patient-out" approach during each training iteration. This way, each of the three models (ANN, RF, and LR) was trained 12 times by excluding one patient's data during each training instant. In total, 36 models (12 ANN, 12 RF, and 12 LR) were obtained, with each patient having their own set of models from each group.Subsequently, these personalised models were used in an ensemble fashion, treating each patient individually.

2.3.3. Ensemble configurations

In the proposed methodology, a heterogeneous ensemble approach is used. Each base model (ANN, RF, LR) was trained in parallel on the training dataset (following the bagging approach), and then the final prediction was acquired by combining the predictions of base models using three different configurations: logical AND approach, majority voting approach, and logical OR approach. The architecture of these three configurations of an ensemble setup is presented in Fig. 2-A. All the base models are binary classifiers, meaning that a prediction of one indicates the existence of a meal and a prediction of zero indicates the absence of a meal.

Configuration 1: the output of each base model (P_1, P_2, and P_3) is provided to the logical AND gate for the final prediction (P_f). The P_f will be one only if P_1, P_2, and P_3 are equal to one. The first configuration is more robust, as all the base models are confident about the final output.

$$P_f = \begin{cases} 1, & \text{if } P_1 = P_2 = P_3 = 1 \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

Configuration 2: uses the majority voting approach as presented in Eq. (11), meaning if at least two out of three models predict class 1 (meal), then the P_f will be class 1. With this configuration, the TPs could be increased but at the cost of an increase in FPs.

$$P_f = \begin{cases} 1, & \text{if } \sum_{i=1}^{3} P_i \geq 2 \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

Configuration 3: uses logical OR, meaning if one or more of the three models predicts class 1, then the output P_f will be class 1, else it will be class 0 (no meal). The last configuration could help decrease the detection time but at the cost of an increase in FPs.

$$P_f = \begin{cases} 1, & \text{if } P_1 = 1 \vee P_2 = 1 \vee P_3 = 1 \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

2.4. Meal detection and compensation

When the base models are trained and assembled into three ensemble configurations, as detailed in Section 2.3.3, they serve for meal detection. A counter is employed to increase the accuracy of the meal's detection. As a result, for the retrospective testing, if P_f persists for three consecutive samples, the meal is detected; otherwise, no meal is output as the final prediction. This is demonstrated in Fig. 2-B. This strategy improves the robustness as well as the dependability of the suggested meal detection system. The performance of the MD is evaluated on two in-silico cohorts as well as on real patients' data.

Afterwards, the ensemble configuration with fewer false positives was chosen and implemented alongside the bolus insulin compensation scheme to assess its real-time efficacy. In this study, we use the proposed MD within OL insulin therapy alongside a standard bolus calculator (SBC) for both in-silico cohorts (SC1 and SC2).

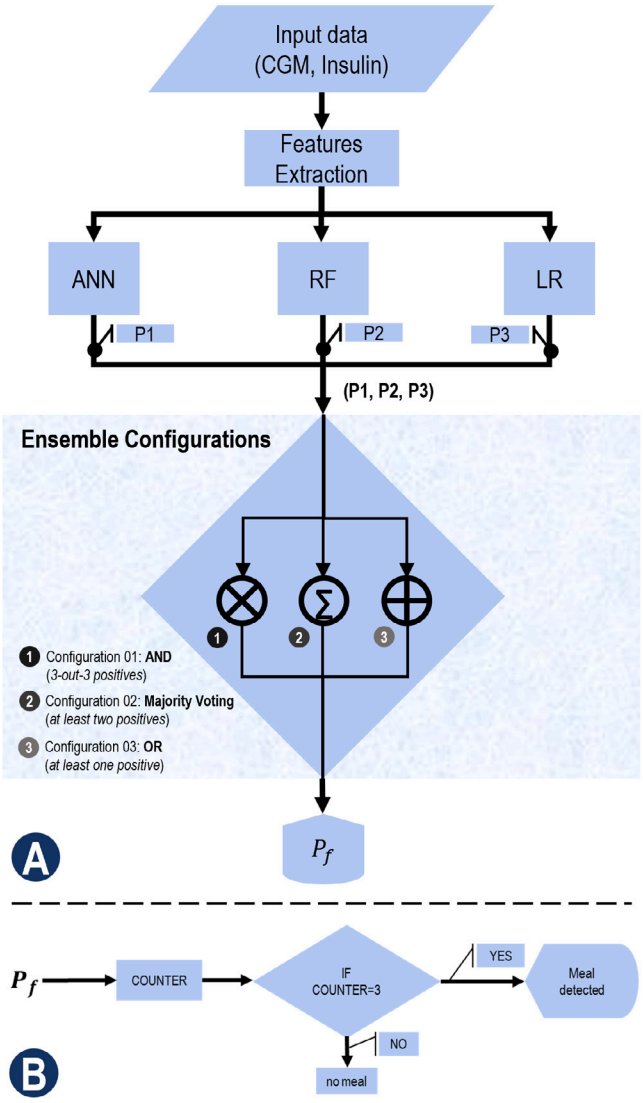

Fig. 2. Heterogeneous ensemble method. (A) Three different ensemble configurations. (B) Integration of a counter to enhance the reliability of prediction.

Fig. 3 shows the framework for adding the proposed MD in an OL architecture with an automatic bolus insulin compensation scheme. At every timestamp, CGM values along with the bolus insulin vector are provided to the proposed MD. The proposed detector has been only used during the daytime, from 6 a.m. to 11 p.m.; this time constraint is employed since individuals typically do not consume meals during nighttime hours. By limiting the meal detection algorithm to the active period of the day, the potential for falsely detecting meals and subsequently triggering unintended insulin administration, which may lead to nocturnal hypoglycemia, is effectively mitigated. Normal therapy has been used for glycemic control during the night. At every timestamp, the 14 features are computed and then fed to the trained ML models to make predictions. The final output is then obtained by combining the predictions of the three classifiers via the most reliable ensemble configuration out of the three available configurations. The ensemble configuration for the real-time *in-silico* testing was carefully chosen with care to prioritise the reduction of FPs, resulting in a more accurate and dependable ensemble model.

Moreover, upon detecting a meal, the system examines the insulin bolus administered within the previous 3 h. If any bolus has been injected during this timeframe, the corresponding detection is disregarded. This precautionary measure aims to prevent unnecessary or excessive bolus insulin injections in two key scenarios: (a) as the MD predicts at each timestamp, it is possible for a single meal to be detected multiple times within a few consecutive timestamps, and (b) in cases of erroneous predictions (FPs). By discarding meal detections associated with recent bolus administration, the system mitigates the risk of undesired insulin administration and ensures more accurate meal detection outcomes.

2.4.1. CHO estimation and bolus calculation

The SBC, presented in Eq. (13), is used for meal bolus calculation. The CHO-to-insulin ratio (CR), a correction factor (CF), current BG (BG_c), reference BG (BG_{ref}), the grams of $\widehat{CHO}$ (which could be the actual amount of CHO for an announced meal (CHO_{AM}) or an estimated CHO (CHO_{UAM}) for an UAM), and estimated insulin-on-board $\widehat{IOB}$ are used to compute the meal bolus. When using the MD, we do not have the amount of CHO. Therefore, if a meal is detected and no insulin bolus has been injected within the past 3 h, the CHO estimation block uses the BG data from the past hour to approximate the amount of CHO for the detected meal based on the trend or slope of the BG curve.

$$u_{SBC} = \frac{\widehat{CHO}}{CR} + \frac{(BG_c - BG_{ref})}{CF} - \widehat{IOB} \tag{13}$$

$$\begin{aligned} BG_{base} &= PBG_{min} + \frac{(PBG_{max} - PBG_{min})}{2} \\ BG_{slope} &= \frac{(PBG_{max} - PBG_{min})}{(PT_{max} - PT_{min})} \\ CHO_{UAM} &= \frac{BG_{last} - BG_{base}}{BG_{slope}} \times \alpha \end{aligned} \tag{14}$$

Eq. (14) elaborates the CHO estimation. Where PBG represents the BG values in the past 1 h, PT is the indexes for that one-hour vector, and α is a correction factor constant value equal to 10. BG_{base} locates the BG value from which the upward trend starts; BG_{last} is the BG value at which the meal is detected; and BG_{slope} finds the slope of that trend. Using all these values, the estimated CHO for the detected unannounced meal is calculated, which is then used for the calculation of the bolus for that detected meal.

2.5. Performance metrics

Various metrics are employed to assess the performance of each ML model as well as the proposed ensemble model. The metrics included are sensitivity (SE), precision (PR), F1-score (F1), detection time delay (DT), and FP rate (FP/Day).

Sensitivity calculates how many of the actual positives our model captures while precision talks about how accurate the model is out of all the predicted positives (how many of the predicted positives are the actual positives). The criteria for true positive (TP), FP, and false negative (FN) are mentioned below. However, the calculation of true negative (TN) for this problem is of limited interest since we are dealing with a highly unbalanced dataset: only a few events (UAM) compared to the rest of the data.

$$Recall = \frac{TP}{TP + FN} \tag{15}$$

$$Precision = \frac{TP}{TP + FP} \tag{16}$$

$$F_1 = 2 \times \left(\frac{Recall \times Precision}{Recall + Precision} \right) \tag{17}$$

To get a robust result, strict criteria are set for the correct meal detection: A 3-h postprandial period is considered for meal detection, along with the condition of three consecutive predictions. Correct detection, or TP, occurs when there is a UAM and the model has predicted it within the 3-h postprandial period and the prediction has persisted for three consecutive samples. On the other hand, if there is no UAM in the ground truth but the model has predicted it as a meal for at

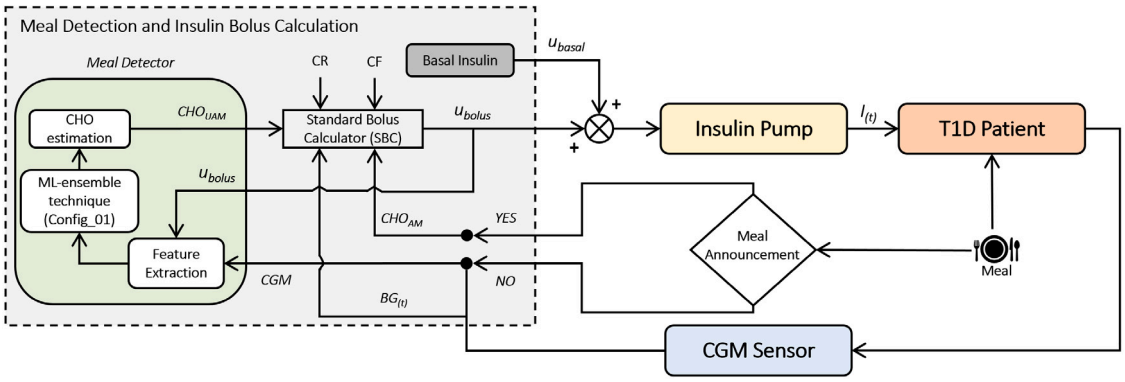

Fig. 3. Implementation of the proposed method in simulation environment.

least three consecutive samples, then that particular prediction is FP. Whereas, if there is a UAM but the model has not flagged or predicted it within the 3-h postprandial window, that particular meal instance is considered FN, and DT is the delay in detecting the meal; it starts with the commencement of the meal and ends when the TP flag is raised. Supplementary Figure 1 illustrates the DT, TP, FP, and FN.

To evaluate glycemic control performance, several commonly used metrics are reported [46]. These include the percentage of time spent in the target range (%TIR), the percentage of time spent above the target range (%TAR), and the percentage of time spent below the target range (%TBR). %TIR refers to the time when BG levels remain within the euglycemic range, which is between 70 mg/dL and 180 mg/dL. %TAR and %TBR are further divided into two levels: %TBR-Level-1 is when BG levels remain between 54 mg/dL and 70 mg/dL, while %TBR-Level-2 is when BG levels fall below 54 mg/dL. %TAR-Level-1 and %TAR-Level-2 represent the percentage of time spent with BG levels between 180 mg/dL and 250 mg/dL and above 250 mg/dL, respectively. In addition, insulin metrics such as basal per day, bolus per day, and total daily insulin (TDI) are also reported.

3. Results

3.1. Meal detection results

The ensemble meal detector has been evaluated both in the described in-silico and OhioT1DM datasets. On one hand, Table 1 shows the overall detection results of the ensemble configurations for the real patients' data. On the other hand, Supplementary Table 2 and Supplementary Table 3 present the results for the in-silico data, indicating a significant improvement compared to the results of individual models. A similar trend in performance has been observed for both in-silico and OhioT1DM datasets. Refer to Supplementary Figure 2 for an overall visualisation of detection times. The FP/Day rate is very low for configuration 1, while it is relatively high for configuration 3. A more in-depth analysis of the results for each configuration reveals the following:

1. **Ensemble configuration 1:** This configuration can be referred to as a robust, low-SE, and low-FP setup, as it has demonstrated better results in terms of FP/Day in comparison to the other two configurations. This is due to the fact that with this configuration, a meal is only detected if all the models detect it and the detection persists for three samples. As a result, the FP rate decreases, but at the expense of an increase in FN and a decrease in SE since some meals may not be detected by all models simultaneously for three consecutive samples. Additionally, the DT in this configuration is higher than in the others, which was expected.
2. **Ensemble configuration 2:** This configuration strikes a balance between FP, FN, and DT. The key factor behind the decrease in DT is that only two out of three models need to be confident about detecting a meal. This, in turn, has increased the number of TP as well as FP compared to the previous configuration. The rise in TP reduced FN, which led to an improvement in SE.
3. **Ensemble configuration 3:** This configuration has the highest SE and lowest DT among the three configurations, but it comes at a significant cost of a higher FP rate. This is because if any one of the three models predicts a meal, it is considered a positive detection. Thus, there could be false detections due to the variability in the rate of change of glucose levels. However, this less-strict criteria result in the highest rate of TP.

Furthermore, the performance of each base model was evaluated individually for all the testing datasets. Supplementary Table 2 shows that all three models show good results for the high-size rapid-absorbing meals; however, for the moderate and high-size meals with slow absorption rates, the FPs and DT increased. In the case of moderate-size meals for SC1, it is hard to detect them as they have a lesser impact on BG. The FNs are thus very high, along with the decrease in sensitivity and F1-score. For SC2, the sensitivity is high, however, the precision is low due to the high number of false positives compared to SC1.

3.2. Basal-bolus glycemic control with automatic meal compensation

Ensemble configuration 1 is employed within an OL insulin therapy with automated UAM compensation, see Fig. 3, to assess its performance and effectiveness in managing BG levels. *In-silico* testing scenarios were categorised similarly to testing datasets, based on the sizes and glucose RAs of UAM. These scenarios are for 30 days in which each patient comprised 24 UAM. The UAM included moderate-sized & slow-absorbing, moderate-sized & rapid-absorbing, high-sized & rapid-absorbing, and high-sized & slow-absorbing. The meal protocol includes 3 meals per day (breakfast (45 g - 8 AM), lunch (90 g - 2 PM), and dinner (70 g - 8 PM). The addition of various factors such as meals from the mixed meal library, meal randomness (mealtime ($\sigma_T = 20$ min), and meal size variability (CV = 10%), and additional variability on insulin absorption and insulin sensitivity were also incorporated to make the simulation environment mimic real-life conditions. The insulin absorption variation was set to ±30%, and the insulin sensitivity varied in a sinusoidal manner. The CHO counting misestimation was also included in the scenarios equal to ±40% (normal distribution), as explained in recent studies [47,48]. This section contains the results obtained during the real-time *in-silico* testing.

The basal-bolus glycemic control performance is observed in three different settings (baseline, without MD, with MD) for four different types of UAM (moderate-size slow-absorbing, moderate-size rapid-absorbing, high-size rapid-absorbing, high-size slow-absorbing). The three settings are (1) the baseline, in which all the meals are announced

Table 1
Evaluation of ensemble configurations on OhioT1DM data.

Metrics	Ensemble configuration 01	Ensemble configuration 02	Ensemble configuration 03
TP	0.47 (121)	0.61 (158)	0.72 (184)
FN	0.53 (136)	0.39 (99)	0.28 (73)
FP	24	60	112
DT	43	41	33
FP/Day	0.07	0.17	0.32
SE	0.47	0.61	0.71
PR	0.83	0.72	0.62
F1Score	0.6	0.66	0.66

The total number of unannounced meal instances in the OhioT1DM data was 257 for all the patients; TP: true positives; FN: false negatives; FP: false positives; DT: detection time in minutes; FP/Day: false positive per day; SE: sensitivity; PR: precision. The SE, PR, and f1score are presented in percentage; TP and FN in percentage (instances); FP are the instances; FP/Day are the total FP divided by the total number of days; DT is presented in minutes as the median of all the detected meals. TP are the correct detection of unannounced meals; FN is the unannounced meals which are not detected; and FP is the incorrect detection of meal instances.

with ±40% of CHO misestimation and the insulin bolus is injected exactly at mealtime, (2) basal-bolus control without MD is the setting in which >25% of meals are randomly set unannounced meaning misses the meal bolus insulin, whereas, (3) the basal-bolus control with MD setting is exactly same as previous one but this time the proposed MD is employed which automatically detect the UAM and then injected the estimated insulin meal bolus to cope the postprandial excursion. The SBC is used to calculate meal boluses. Supplementary Tables 4, 5, 6 and 7 present the performance of the proposed MD in terms of CGM and insulin metrics. Whereas, Fig. 4 presents a populational comparison of CHO estimation in terms of mean absolute percentage error (MAPE) along with the distribution of the estimation errors (defined as the difference between actual CHO and estimated CHO). Patient-wise comparison is presented in Supplementary Figure 4. The MAPE for each patient is calculated as:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{ActualCHO_i - EstimatedCHO_i}{ActualCHO_i}\right| \times 100 \quad (18)$$

where n represents the total number of detected UAM for each patient. The average MAPE recorded for all patients across the four scenarios is equal to 30%, where the error in the case of smaller-size meals is lower compared to high-size meals. All the metrics are reported as median (interquartile range, 25%–75%). It clearly shows that using the proposed MD can improve the TIR and decrease the TAR without increasing the percentage of time in hypoglycemia. We have achieved an overall better performance during *in-silico* testing, with an average +10.48% and +16.03% improvement in TIR, a reduction of −5.16% and −11.85% in TAR, and only a +0.35% and 2.69% increase in TBR for both in-silico cohorts, respectively.

In addition, the Wilcoxon signed rank test was considered for the calculation of p-values to check the significance of the proposed MD. The asterisk (*) in Supplementary Tables 4, 5, 6 and 7 indicates the intervention (p-value < 0.05) of the proposed MD when compared to the without MD setting, where the null hypothesis was that there is no intervention. Overall, the proposed MD showed good results for all types of UAM; all the metrics (median, mean, max, TIR, and TAR) were improved.

To illustrate the effect of MD in postprandial glucose control compared to the baseline, two days of glucose records along with the CHO amount taken for adult #001 is presented in Fig. 5. Total meals taken in these two days are 6 where dinner on day 1 and lunch on day 2 are UAM. Due to this the excursion in the glucose curve can be seen compared to the baseline. The MD helped in both cases to reduce the postprandial glucose level. A detailed analysis of the postprandial period for adults and adolescents is presented in an Excel file, attached as supplementary material (see mmc1.xlsx). This table shows the aggregated outcomes of postprandial periods (4 h) of unannounced meals for each patient, depending on the meal classification. The postprandial period is defined as a 4-h time window from the real-time of the meal.

In addition, to explicitly represent the percentage improvement in TIR for the adolescent population compared to adults, Fig. 6 presents a brief overview. A clear improvement in TIR can be seen for each population in various scenarios. The detection rate of the MD vs meal onset time for both SC1 and SC2 cohorts as well as for the retrospective testing of real patients' data is presented in Supplementary Figure 2. The supplementary Figure 3 illustrates the best, medium and worst postprandial glucose curve for both in-silico cohorts based on the glycemic risk index (GRI) [49].

4. Discussion

Meal detection can be treated as a classification task where a binary classifier determines the presence or absence of a meal (represented as 1 or 0). However, using a single binary classifier in an AID system can pose significant risks as false detection may cause improper insulin delivery, leading to hypoglycemia. For this reason, we employed a more robust approach with the ensemble learning method to mitigate this risk. Ensemble learning integrates multiple ML models to improve predictive performance beyond that of an individual model [50]. Such an approach has recently been used for hypoglycemia prediction [51] but to the best of our knowledge this is the first time that such kind of approach has been used for UAM detection.

To check the efficacy of the proposed methodology, the evaluation is done on two in-silico cohorts (SC1 and SC2), with SC1 comprising 20 patients, including 10 adolescents and 10 adults. The rationale behind selecting this specific age group was that most of the missed meal boluses are reported in adults and adolescents [8–10]. The study [52] indeed showed a consistently increasing HbA1c pattern, peaking in adolescence and early adulthood. These findings underscore the significance of adolescence and early adulthood as prime stages for enhancing diabetes management. The second cohort contained 47 virtual patients (VP), these VPs are generated using the Hovorka model with physiological variabilities along with challenging and realistic scenarios [30]. To simulate a realistic scenario, insulin sensitivity variability was identified using basal profile patterns from existing literature. Additionally, a library of mixed meals [29] derived from real-world data was incorporated and an OL insulin therapy was used.

Supplementary Table 2 shows that for SC1 the moderate-size UAM are hard to detect compared to the high-sized meal. However, this is not a big concern because the studies show that AID systems can compensate for the postprandial rise due to low-size meals [11,13]. In the case of SC2, the sensitivity is good in each scenario but the FPs are higher compared to SC1. This is may be a consequence of the higher variability of the SC2 VP. From the results presented in Supplementary Table 3, it is quite evident that configuration 2 and configuration 3 have a higher sensitivity but at the expense of higher rates of FP. For an AID system, it is crucial to have an MD with a low FP rate to prevent the injection of an inappropriate dose of insulin that may lead to hypoglycemia. Each FP may trigger a delivery of an inappropriate insulin bolus. Therefore, configuration 1 is the best option to go with as it has very few FP along with considerably good SE, PR, F1-score, and DT. While testing the proposed algorithm on the

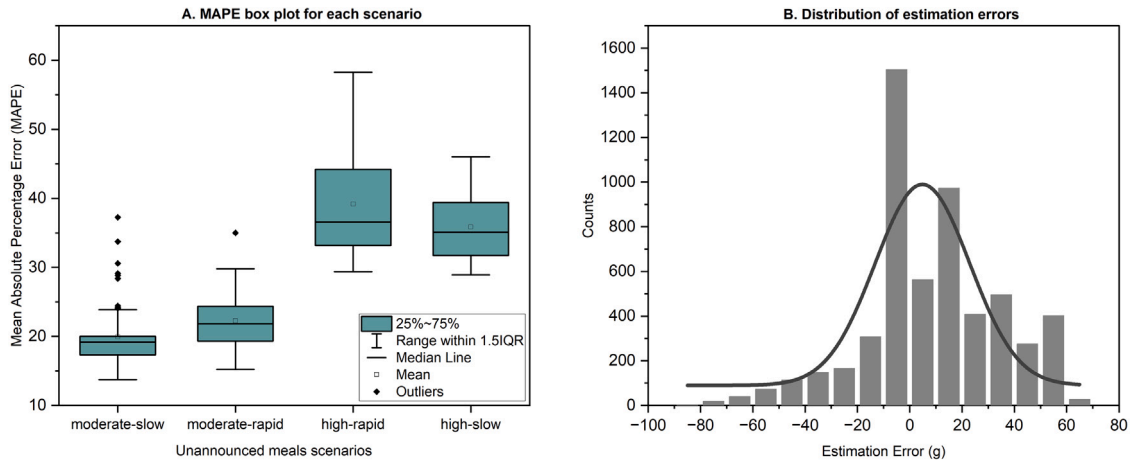

Fig. 4. Populational MAPE and distribution of estimation errors of CHO estimation for both in-silico cohorts. (A) Box plot of the MAPE (as defined in Eq. (18)) for all the T1D patients within the in-silico cohorts across the four scenarios; each point in the box plot for a particular scenario represents the MAPE of each patient (B) Distribution of the estimation errors (difference between actual and estimated CHO) for all the detected UAM across all the scenarios for both in-silico cohorts.

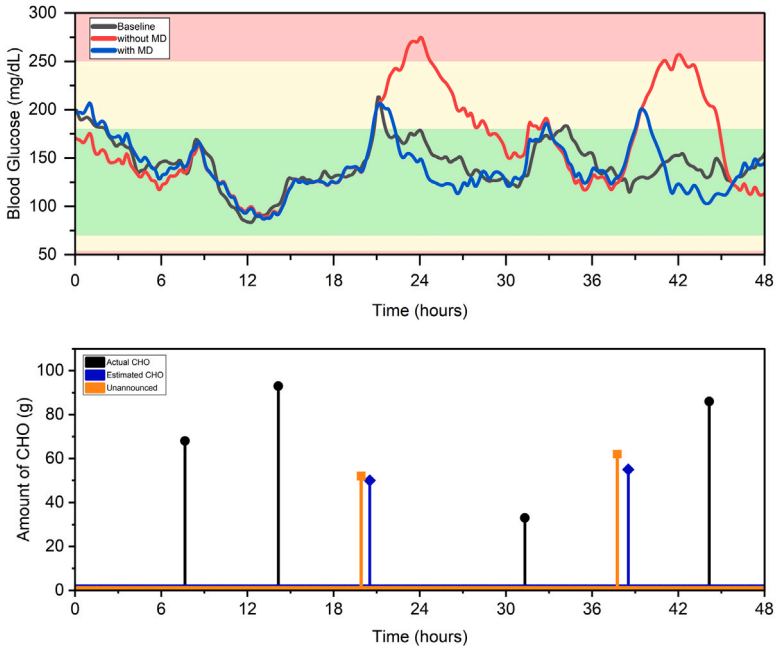

Fig. 5. Representative BG trajectories for the three analysed architectures for SC1 adult #001.

OhioT1DM dataset, the meal detection performance was similar to what was observed with the in-silico data. Before evaluating the performance of the meal detector, the OhioT1DM was preprocessed and left with a cumulative sum of 349 days of insulin and BG data for the 12 patients. When comparing the results of each ensemble configuration, the PR is higher in ensemble configuration 1 due to a lower rate of FP, however, at the cost of lower SE. The DT is also higher in ensemble configuration 1 compared to other ensemble configurations, the same as recorded in in-silico cohorts. The reason for having a lower SE for OhioT1DM compared to in-silico cohorts may be caused by the limited amount of data. Hence, a "leave-one-patient-out" approach was used to effectively use the OhioT1DM for both training and testing.

An extensive analysis was performed using the OhioT1DM dataset comprising of real patients' data and two different testing cohorts of virtual patients were considered for evaluation of the MD. Whereas, during the 30-day *in-silico* trial, the proposed MD with insulin compensation overall improved the TIR and decreased the TAR, as shown in Supplementary Tables 4, 5, 6 and 7. A slight increase in hypoglycemia was also observed in all scenarios. Various challenging scenarios were created and different variabilities, such as randomness in meals, various rates of absorption of meals (using the mixed-meal library), and circadian variability in insulin sensitivity were incorporated into the simulator to mimic the real-life conditions. We acknowledge that one of the limitations of the study is using in-silico data, even though we have tried to minimise this issue by introducing the Ohio T1DM cohort. The meal detector (without the meal compensation feature) has been tested on the OhioT1DM dataset. Since the OhioT1DM dataset is retrospective data, the performance of the meal compensation bolus of the proposed methodology cannot be evaluated.

In consideration of patient safety and to prevent unintended insulin administration, our methodology includes specific constraints. Initially, the meal detection algorithm operates exclusively during daytime hours, specifically from 6 AM to 11 PM, when the likelihood of meals is higher. Additionally, we discard meal detections if a bolus has been administered within the past three hours. This precautionary measure is implemented to mitigate the potential risk of inducing hypoglycemia.

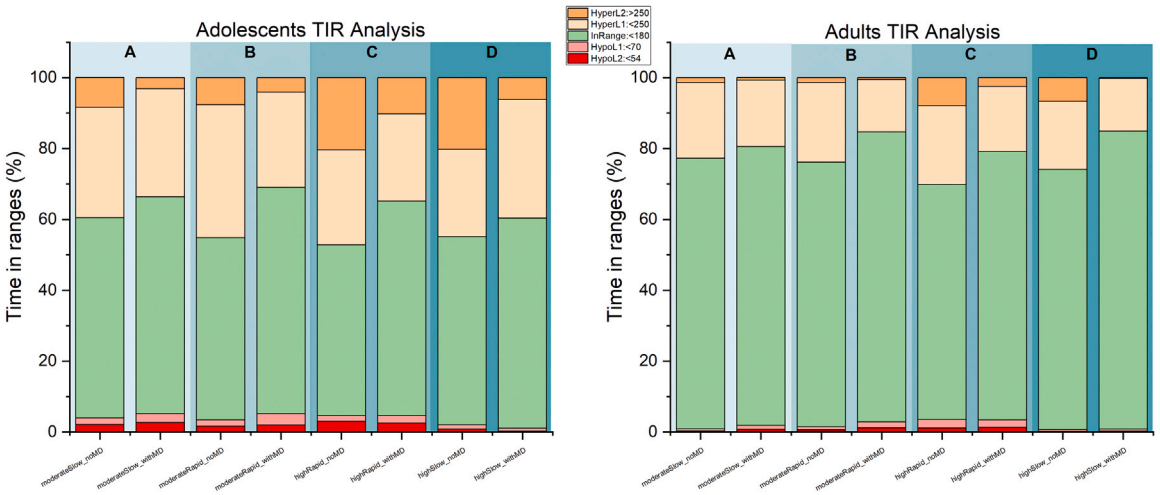

Fig. 6. TIR improvement by incorporating MD algorithm: Adolescents vs. Adults Different shades in the background represent various UAM scenarios. Starting from left to right in each plot: (A) scenario with moderate size & slow-absorbing UAM; (B) moderate size & rapid-absorbing UAM scenario; (C) scenario with high size & rapid-absorbing UAM; (D) high size & slow-absorbing UAM scenarios.

4.1. Comparative analysis

For the detection and estimation of UAM, several approaches have recently been put forth, as highlighted in Section 1. Table 2 presents a comparison of the proposed method performance metrics with recently published research works.

The majority of the approaches are threshold-based employing KF; meals are detected by evaluating the rate of change of glucose concentration levels. Starting with the method proposed by Dassau et al. [14], four different methods were used in an ensemble manner and evaluated on seventeen subjects. On average, 56 g of meal was consumed by the subjects for breakfast. The best result was achieved for the two-out-of-three voting scheme, in which the meals were detected within 30 min on average and more than 90% meals were detected before the difference in glucose level exceeded 40 mg/dL from the baseline. The CGM measurements they used, however, have a sample interval of 1 min, and the stated parameters are distinct, making it difficult to make a meaningful comparison.

The meal detection algorithm proposed by Samadi et al. [16] was evaluated on *in silico* UVA/Padova cohort of 30 patients: 10 adults, 10 adolescents, and 10 children. This scenario that they used is more comparable to ours; the basal insulin and carbs-to-insulin ratio were set to the default values for each patient, and the time to consume a meal was considered to be 15 min for all the meals. The overall performance reported in terms of PE and SE for this method is 91.7% and 91.3% respectively. However, the performance reported for the adult patients is 79% and 87%, while the detection time of the meal is not presented in this study. A subsequent study by the same group achieved PE and SE scores of 79% and 93.5%, respectively, for 11 adult patients' clinical data, and the mean detection time recorded was 35 min. The detection algorithm proposed by Zheng et al. [21] was evaluated on *in silico* data comprised of 100 patients. The performance presented includes a detection delay time equal to 26 min along with PE and SE equal to 93.3% and 88%, respectively. A study by Ramkissoon et al. [15] also used simulated data (10 UVA/Padova adult patients) for the evaluation of their proposed methodology. Three different setups have been implemented, and the trade-off (trade-off between FP and SE) setup demonstrates the best results. The FP rate recorded for this setup was 0.2 per day, whereas the SE of 82% was achieved, and 38 min of detection delay time was also reported. From the available data, a PE is calculated to be equal to 92.5%.

The problem with the rate of change-based methods is that the meal cannot be detected until it appears in the bloodstream with a significant difference from the preprandial glucose. This inevitably induces a delay in detecting the meal and in this case, the bolus for the detected meal could cause postprandial hypoglycemia and ultimately results in poor glycemic control. We tried in the proposed methodology to minimise the detection delay as much as possible and deliver the insulin earlier to avoid higher glucose excursion for longer periods and to avoid postprandial hypoglycemia. For example, if we examine the DT for the publications stated in the preceding paragraph, we see that [16] did not record the DT, whereas [15,21] have DT of 38 min and 26, respectively. In our investigation, the best DT we have is 21 min. Moreover, ML methods, in contrast to traditional model-based approaches, offer numerous benefits, such as data-driven decision-making based on learned patterns, the capability to handle complex relationships between variables, scalability to handle large datasets, the ability to generalise well to unseen data, adaptability to changing

Table 2
Comparison of our proposed MD with the method reported in the literature.

MD[a] algorithms	Performance metrics				
	SE (%)	PR (%)	F1 (%)	DT (MIN)	FP/Day
Dassau et al. [14]	–	–	–	30	–
Samadi et al. [16]	87	79	86	–	–
Xie and Wang [19]	76	84	80	45	–
Mahmoudi et al. [20]	99.5	–	–	58	–
Samadi et al. [17]	93.5	79	86	35	–
Ramkissoon et al. [15]	82	92.5	87	38	0.2
Zheng et al. [21]	88	93	91	26	–
Daniels et al. [24]	76	93	84	38	–
Rodriguez et al. [27]	–	–	–	35	–
Idi et al. [28]	78	–	–	34.8	0.05
Lopez et al. [26]	83	–	–	26	0.16
Cohorts[b]	Ensemble configuration 1				
SC1	62	98	74	31	0.02
SC2	87	94	90	31	0.09
OhioT1DM	47	83	60	43	0.07
	Ensemble configuration 2				
SC1	77	96	85	25	0.05
SC2	94	89	91	25	0.19
OhioT1DM	61	72	66	41	0.17
	Ensemble configuration 3				
SC1	93	88	90	21	0.2
SC2	97	79	87	21	0.42
OhioT1DM	71	62	66	33	0.32

[a] Meal detector.

[b] Mean values are presented for each configuration.

environments by incorporating new information, and the potential to leverage ensemble learning techniques to enhance prediction accuracy and robustness. These advantages make machine learning methods a powerful and versatile tool for various applications, surpassing the limitations of traditional model-based approaches that rely on predefined thresholds or rules.

The methods proposed by Xie and Wang [19] and Mahmoudi et al. [20] detect UAM via outlier detection. Both studies used *in silico* data for evaluation. The algorithm proposed by [20] using 10 adult patients for evaluation, achieved the highest SE of 99.5% but at the cost of a longer delay time which was reported as 58 min. On the other hand, [19] evaluated their model on 30 patients and 76% SE and 84% PE were demonstrated with a detection time equal to 45 min. In addition, both these approaches rely on the KF to estimate state variables and predict future states.

The rest of the methods listed for comparison are data-driven, which is more comparable to our proposed approach. A deep learning approach by Daniels et al. [24] has been tested for different sizes of meals including snacks (mean CHO size of 70 g) and the overall performance reported is 76% SE, 93% PR, and a detection time of 38 min. The usage of meals from the mixed meal library [29] gives our proposed method an edge over [24] and allows us to test its effectiveness on a range of meal sizes and appearance rates. Since the detection is based on both the size and the glucose RA of the UAM, for instance, if we have two meals that are both 70 g in size but appear at different rates, the faster meal will cause a high and abrupt excursion, making it easier to detect than the slower meal. The +3.9% TIR, −4.2% TAR, and +0.1% reported by [24]. The *in silico* performance comparison, however, is unfair because our suggested method has only been validated in an OL therapy, whereas theirs has been validated in a closed-loop therapy. In the future study, we will assess the proposed meal detection module's performance in closed-loop insulin administration.

Some of the recent studies have utilised traditional binary classification models. We believe the results reported by Rodriguez et al. [27] and Idi et al. [28] are the preliminary findings of their work, as these were collected from the e-poster published at [53]. [27] used supervised classification models for meal detection, and the only parameter reported is the detection time, which is 35 min. A fair comparison with their approach is not possible as they have not presented parameters such as SE, PE, or FP rate. Whereas, [28] achieved 78% SE, 35 min DT, and an FP rate of 0.05, and it has a slight edge on our proposed framework as they have used an unsupervised anomaly detection model, i.e., isolation forest, which eliminates the requirement of labelled data for training. Finally, the most recently published work by Lopex et al. [26] employed a multioutput neural network setup and achieved 83% SE, a 26 min detection time, and 16.6% FP in a clinical trial involving just 5 patients. The same algorithm recorded an average SE of up to 90% with 10% FP and DT equal to 27.5 min. This encourages us to evaluate our proposed framework in the future in a clinical trial to check its efficacy in real time. Our results suggest that the ensemble meal detector approach performs as good as most recent published works, but with a significant improvement in the FP daily rate. However, this is just a qualitative analysis because head-to-head comparison with the published works is challenging as not all the parameters of each approach are available.

4.2. Future directions

In addition, the potential risk associated with the meal detection algorithm is the inadvertent detection of meals, which results in the unintended administration of an insulin dose and possibly causes hypoglycemia. One possible such scenario could be the inadvertent detection of rescue CHO intakes as meals. Furthermore, including scenarios with physical activity, psychological stress, pressure induced sensor attenuation faults, and hormonal changes in could significantly improve the robustness of the proposed system. Having said that, there is no mathematical model available in the latest simulators for psychological stress or hormonal mutations for T1D [30,54–57]. Some potential techniques for dealing with these real-world challenges could include: (1) Disable automated MD for rescue CHO intakes in future implementation. (2) Integrate physical activity data to differentiate between meal and exercise-related glucose changes. (3) Use a personalised model to account for interpatient variability. (4) Include a feedback mechanism. The future goal also includes the translation of the proposed MD into a clinical setting.

5. Conclusion

Postprandial glucose control is one of the major challenges in AID systems. Currently, available hybrid closed-loop systems have addressed this problem to some extent by requiring patients to be fed information about the meal to be consumed, such as the amount of CHO. Estimating CHO before each meal is burdensome and a challenge for T1D patients. In this work, we proposed an ensemble approach comprised of three classifiers: ANN, RF, and LR. The proposed approach is validated on OhioT1DM real patients' data and two different VP cohorts: (1) 20 patients (10 adults and 10 adolescents) based on the Dalla Man model, and (2) 47 patients based on the Hovorka model with realistic scenarios. To create realistic scenarios, various variability factors were incorporated into these datasets, which include meal randomness, intra-patient variability, meals of various sizes (moderate, and high), and glucose RAs (slow-appearing and rapid-appearing). The proposed MD is also tested in a real-time simulation environment using it alongside SBC in the OL insulin therapy for both in-silico cohorts. The use of the proposed method significantly improved the TIR (10.48% and 16.03%) and reduced the TAR (5.16% and 11.85%) with minimal increase in the percentage of time in hypoglycemia (0.35% and 2.69%) for both SC1 and SC2 cohorts, respectively. In addition, the results with the real patients' data (OhioT1DM) are also encouraging, the best results achieved were sensitivity, precision, and F1-score of 61%, 72%, and 66%, respectively, with only 0.17 false positives per day. The effect of the proposed method on closed-loop therapy using the advanced control technique is yet to be evaluated, and it will be addressed in a future study. However, this study's findings imply that the proposed MD can serve as a complementary module in an AID system to detect meals without relying on meal announcements, thereby ensuring consistent glucose control.

Funding

This work was partially supported by the Spanish Ministry of Science and Innovation under Grant number PID2019-107722RB-C22 and PDC2021-121470-C22, by the Autonomous Government of Catalonia under Grant number 2021 SGR 01598 and by the program for researchers in training at the University of Girona IFUdG2021 (2021 FI_B 00876). Open Access funding provided thanks to the CRUE-CSIC agreement with Elsevier.

CRediT authorship contribution statement

Muhammad Ibrahim: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. **Aleix Beneyto:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Visualization, Writing – review & editing. **Ivan Contreras:** Funding acquisition, Visualization, Writing – review & editing. **Josep Vehi:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Visualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.compbiomed.2024.108154.

References

[1] H. Blauw, P. Keith-Hynes, R. Koops, J.H. DeVries, A review of safety and design requirements of the artificial pancreas, Ann. Biomed. Eng. 44 (11) (2016) 3158–3172.

[2] J. Burdick, H.P. Chase, R.H. Slover, K. Knievel, L. Scrimgeour, A.K. Maniatis, G.J. Klingensmith, Missed insulin meal boluses and elevated hemoglobin A1c levels in children receiving insulin pump therapy, Pediatrics 113 (2004) e221–e224, http://dx.doi.org/10.1542/peds.113.3.e221.

[3] S.R. Patton, M.A. Clements, A. Fridlington, C. Cohoon, A.L. Turpin, S.A. DeLurgio, Frequency of mealtime insulin bolus as a proxy measure of adherence for children and youths with type 1 diabetes mellitus, Diabetes Technol. Ther. 15 (2013) 124–128, http://dx.doi.org/10.1089/dia.2012.0229.

[4] K.A. Datye, C.T. Boyle, J. Simmons, D.J. Moore, S.S. Jaser, N. Sheanon, J.M. Kittelsrud, S.E. Woerner, K.M. Miller, Timing of meal insulin and its relation to adherence to therapy in type 1 diabetes, J. Diabetes Sci. Technol. 12 (2) (2018) 349–355.

[5] A. Peters, M.A. Van Name, B. Thorsted, J.S. Piltoft, W.V. Tamborlane, Postprandial dosing of bolus insulin in patients with type 1 diabetes: a cross-sectional study using data from the T1D exchange registry, Endocr. Pract. 23 (10) (2017) 1201–1209.

[6] D. Hessler, L. Fisher, W. Polonsky, U. Masharani, L. Strycker, A. Peters, I. Blumer, V. Bowyer, Diabetes distress is linked with worsening diabetes management over time in adults with type 1 diabetes, Diabetic Med. 34 (9) (2017) 1228–1234.

[7] K.L. Joiner, M.L. Holland, M. Grey, Stressful life events in young adults with type 1 diabetes in the US T1D exchange clinic registry, J. Nurs. Scholarsh. 50 (6) (2018) 676–686.

[8] M. O'Connell, S. Donath, F. Cameron, Poor adherence to integral daily tasks limits the efficacy of CSII in youth, Pediatr. Diabetes 12 (6) (2011) 556–559.

[9] S. Robinson, R.S. Newson, B. Liao, T. Kennedy-Martin, T. Battelino, Missed and mistimed insulin doses in people with diabetes: A systematic literature review, Diabetes Technol. Ther. 23 (12) (2021) 844–856.

[10] A.L. Olinder, A. Kernell, B. Smide, Missed bolus doses: devastating for metabolic control in CSII-treated adolescents with type 1 diabetes, Pediatr. Diabetes 10 (2) (2009) 142–148.

[11] R. Shalit, N. Minsky, M. Laron-Hirsh, O. Cohen, N. Kurtz, A. Roy, B. Grosman, A. Benedetti, A. Tirosh, Unannounced meal challenges using an advanced hybrid closed loop system (AHCL), Diabetes Technol. Ther. (ja) (2023).

[12] E. Dassau, H. Zisser, R.A. Harvey, M.W. Percival, B. Grosman, W. Bevier, E. Atlas, S. Miller, R. Nimri, L. Jovanovič, et al., Clinical evaluation of a personalized artificial pancreas, Diabetes Care 36 (4) (2013) 801–809.

[13] M. Reddy, P. Herrero, M.E. Sharkawy, P. Pesl, N. Jugnee, D. Pavitt, I.F. Godsland, G. Alberti, C. Toumazou, D.G. Johnston, et al., Metabolic control with the bio-inspired artificial pancreas in adults with type 1 diabetes: a 24-hour randomized controlled crossover study, J. Diabetes Sci. Technol. 10 (2) (2016) 405–413.

[14] E. Dassau, B.W. Bequette, B.A. Buckingham, F.J. Doyle III, Detection of a meal using continuous glucose monitoring: implications for an artificial β-cell, Diabetes Care 31 (2) (2008) 295–300.

[15] C.M. Ramkissoon, P. Herrero, J. Bondia, J. Vehi, Unannounced meals in the artificial pancreas: detection using continuous glucose monitoring, Sensors 18 (3) (2018) 884.

[16] S. Samadi, K. Turksoy, I. Hajizadeh, J. Feng, M. Sevil, A. Cinar, Meal detection and carbohydrate estimation using continuous glucose sensor data, IEEE J. Biomed. Health Inform. 21 (3) (2017) 619–627.

[17] S. Samadi, M. Rashid, K. Turksoy, J. Feng, I. Hajizadeh, N. Hobbs, C. Lazaro, M. Sevil, E. Littlejohn, A. Cinar, Automatic detection and estimation of unannounced meals for multivariable artificial pancreas system, Diabetes Technol. Ther. 20 (3) (2018) 235–246.

[18] F. Cameron, G. Niemeyer, B.A. Buckingham, Probabilistic evolving meal detection and estimation of meal total glucose appearance, 2009.

[19] J. Xie, Q. Wang, A variable state dimension approach to meal detection and meal size estimation: in silico evaluation through basal-bolus insulin therapy for type 1 diabetes, IEEE Trans. Biomed. Eng. 64 (6) (2017) 1249–1260.

[20] Z. Mahmoudi, K. Nørgaard, N.K. Poulsen, H. Madsen, J.B. Jørgensen, Fault and meal detection by redundant continuous glucose monitors and the unscented Kalman filter, Biomed. Signal Process. Control 38 (2017) 86–99.

[21] M. Zheng, B. Ni, S. Kleinberg, Automated meal detection from continuous glucose monitor data through simulation and explanation, J. Am. Med. Inform. Assoc. 26 (12) (2019) 1592–1599.

[22] S. Faccioli, I. Sala-Mira, J. Díez, A. Facchinetti, G. Sparacino, S. Del Favero, J. Bondia, Super–twisting-based meal detector for type 1 diabetes management: Improvement and assessment in a real-life scenario, Comput. Methods Programs Biomed. 219 (2022) 106736.

[23] J. Garcia-Tirado, J.L. Diaz, R. Esquivel-Zuniga, C.L. Koravi, J.P. Corbett, M. Dawson, C. Wakeman, C.L. Barnett, M.C. Oliveri, H. Myers, et al., Advanced closed-loop control system improves postprandial glycemic control compared with a hybrid closed-loop system following unannounced meal, Diabetes Care 44 (10) (2021) 2379–2387.

[24] J. Daniels, P. Herrero, P. Georgiou, A deep learning framework for automatic meal detection and estimation in artificial pancreas systems, Sensors 22 (2) (2022) 466.

[25] A. Cinar, M.R. Askari, M. Rashid, M. Abdel-Latif, M. Sevil, A. Shahidehpour, Machine learning to detect meals and physical activities from historical data of people with type 1 diabetes in free living, in: The Official Journal of ATTD Advanced Technologies & Treatments for Diabetes Conference 22-25 February 2023 - Berlin & Online, Diabetes Technology & Therapeutics, 2023, pp. A–85–A–86, URL http://dx.doi.org/10.1089/dia.2023.2525.abstracts.

[26] C. Mosquera-Lopez, L.M. Wilson, J. El Youssef, W. Hilts, J. Leitschuh, D. Branigan, V. Gabo, J.H. Eom, J.R. Castle, P.G. Jacobs, Enabling fully automated insulin delivery through meal detection and size estimation using artificial intelligence, NPJ Digit. Med. 6 (1) (2023) 39.

[27] E. Rodriguez, R.V. Mejia, Meal detection and estimation on type 1 diabetic patients, in: The Official Journal of ATTD Advanced Technologies & Treatments for Diabetes Conference 22-25 February 2023 - Berlin & Online, Diabetes Technology & Therapeutics, 2023, pp. A–92, URL http://dx.doi.org/10.1089/dia.2023.2525.abstracts.

[28] E. Idi, J. Pavan, M. Vettoretti, A. Facchinetti, G. Sparacino, S.D. Favero, Detection of unannounced meals in artificial pancreas systems using isolated forest algorithm and survival analysis technique, in: The Official Journal of ATTD Advanced Technologies & Treatments for Diabetes Conference 22-25 February 2023 - Berlin & Online, Diabetes Technology & Therapeutics, 2023, pp. A–118–A–119, URL http://dx.doi.org/10.1089/dia.2023.2525.abstracts.

[29] F.M.L. Vargas, et al., Design and implementation of a closed-loop blood glucose control system in patients with type 1 diabetes, Univ. Girona (2013) 69–70.

[30] E. Estremera, A. Cabrera, A. Beneyto, J. Vehi, A simulator with realistic and challenging scenarios for virtual T1D patients undergoing CSII and MDI therapy, J. Biomed. Inform. 132 (2022) 104–141.

[31] C. Dalla Man, R.A. Rizza, C. Cobelli, Meal simulation model of the glucose-insulin system, IEEE Trans. Biomed. Eng. 54 (10) (2007) 1740–1749.

[32] C. Marling, R. Bunescu, The OhioT1DM dataset for blood glucose level prediction: Update 2020, in: CEUR Workshop Proceedings, Vol. 2675, NIH Public Access, 2020, p. 71.

[33] F. León-Vargas, F. Garelli, H. De Battista, J. Vehí, Postprandial blood glucose control using a hybrid adaptive PD controller with insulin-on-board limitation, Biomed. Signal Process. Control 8 (6) (2013) 724–732.

[34] P.G. Jacobs, P. Herrero, A. Facchinetti, J. Vehi, B. Kovatchev, M. Breton, A. Cinar, K. Nikita, F. Doyle, J. Bondia, et al., Artificial intelligence and machine learning for improving glycemic control in diabetes: best practices, pitfalls and opportunities, IEEE Rev. Biomed. Eng. (2023).

[35] R. Hu, C. Li, An improved PID algorithm based on insulin-on-board estimate for blood glucose control with type 1 diabetes, Comput. Math. Methods Med. 2015 (2015) 1–8, http://dx.doi.org/10.1155/2015/281589.

[36] A. Beneyto, V. Puig, B.W. Bequette, J. Vehi, A hybrid automata approach for monitoring the patient in the loop in artificial pancreas systems, Sensors 21 (21) (2021) 7117.

[37] Z.-H. Zhou, Ensemble Methods: Foundations and Algorithms, CRC Press, 2012.

[38] H. Parvin, M. MirnabiBaboli, H. Alinejad-Rokny, Proposing a classifier ensemble framework based on classifier selection and decision tree, Eng. Appl. Artif. Intell. 37 (2015) 34–42.

[39] J. Mendes-Moreira, A.M. Jorge, J.F. de Sousa, C. Soares, Improving the accuracy of long-term travel time prediction using heterogeneous ensembles, Neurocomputing 150 (2015) 428–439.

[40] L. Breiman, Bagging Predictors, Technical Report 421, University of California, Berkeley, 1994.

[41] Y. Freund, R.E. Schapire, et al., Experiments with a new boosting algorithm, in: ICML, Vol. 96, Citeseer, 1996, pp. 148–156.

[42] D.H. Wolpert, Stacked generalization, Neural Netw. 5 (2) (1992) 241–259.

[43] W. Sun, Z. Li, Hourly PM2. 5 concentration forecasting based on feature extraction and stacking-driven ensemble model for the winter of the Beijing-Tianjin-Hebei area, Atmos. Pollut. Res. 11 (6) (2020) 110–121.

[44] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[45] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, C.-J. Lin, LIBLINEAR: A library for large linear classification, J. Mach. Learn. Res. 9 (2008) 1871–1874.

[46] D.M. Maahs, B.A. Buckingham, J.R. Castle, A. Cinar, E.R. Damiano, E. Dassau, J.H. DeVries, F.J. Doyle III, S.C. Griffen, A. Haidar, et al., Outcome measures for artificial pancreas clinical trials: a consensus report, Diabetes Care 39 (7) (2016) 1175–1179.

[47] S. Ahmad, A. Beneyto, I. Contreras, J. Vehi, Bolus insulin calculation without meal information. A reinforcement learning approach, Artif. Intell. Med. 134 (2022) 102436.

[48] C. Roversi, M. Vettoretti, S. Del Favero, A. Facchinetti, P. Choudhary, G. Sparacino, Impact of carbohydrate counting error on glycemic control in open-loop management of type 1 diabetes: quantitative assessment through an in silico trial, J. Diabetes Sci. Technol. 16 (6) (2022) 1541–1549.

[49] D. Klonoff, J. Wang, D. Rodbard, et al., A glycemia risk index (GRI) of hypoglycemia and hyperglycemia for continuous glucose monitoring validated by clinician ratings [published online ahead of print march 29, 2022], J. Diabetes Sci. Technol..

[50] X. Dong, Z. Yu, W. Cao, Y. Shi, Q. Ma, A survey on ensemble learning, Front. Comput. Sci. 14 (2020) 241–258, http://dx.doi.org/10.1007/s11704-019-8208-z.

[51] V. Felizardo, N.M. Garcia, I. Megdiche, N. Pombo, M. Sousa, F. Babič, Hypoglycaemia prediction using information fusion and classifiers consensus, Eng. Appl. Artif. Intell. 123 (2023) 106194.

[52] K.M. Miller, R.W. Beck, N.C. Foster, D.M. Maahs, T. Exchange, Hba1c levels in type 1 diabetes from early childhood to older adults: a deeper dive into the influence of technology and socioeconomic status on HbA1c in the T1D exchange clinic registry findings, Diabetes Technol. Ther. 22 (9) (2020) 645–650.

[53] The official journal of ATTD advanced technologies & treatments for diabetes conference 22-25 february 2023 | Berlin & online, in: Advanced Technologies & Treatments for Diabetes Conference, Vol. 25, Diabetes Technology & Therapeutics, 2023, pp. A–1–A–269, http://dx.doi.org/10.1089/dia.2023.2525.abstracts.

[54] C.D. Man, F. Micheletto, D. Lv, M. Breton, B. Kovatchev, C. Cobelli, The UVA/PADOVA type 1 diabetes simulator: new features, J. Diabetes Sci. Technol. 8 (1) (2014) 26–34.

[55] R. Hovorka, V. Canonico, L.J. Chassin, U. Haueter, M. Massi-Benedetti, M.O. Federici, T.R. Pieber, H.C. Schaller, L. Schaupp, T. Vering, et al., Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes, Physiol. Meas. 25 (4) (2004) 905.

[56] M. Rashid, S. Samadi, M. Sevil, I. Hajizadeh, P. Kolodziej, N. Hobbs, Z. Maloney, R. Brandt, J. Feng, M. Park, et al., Simulation software for assessment of nonlinear and adaptive multivariable control algorithms: glucose–insulin dynamics in type 1 diabetes, Comput. Chem. Eng. 130 (2019) 106565.

[57] M.R. Smaoui, R. Rabasa-Lhoret, A. Haidar, Development platform for artificial pancreas algorithms, PLos One 15 (12) (2020) e0243139.