Learning Difference Equations With Structured Grammatical Evolution for Postprandial Glycaemia Prediction

Daniel Parra, David Joedicke, J. Manuel Velasco, Gabriel Kronberger, and J. Ignacio Hidalgo

***Abstract*—People with diabetes must carefully monitor their blood glucose levels, especially after eating. Blood glucose management requires a proper combination of food intake and insulin boluses. Glucose prediction is vital to avoid dangerous post-meal complications in treating individuals with diabetes. Although traditional methods, and also artificial neural networks, have shown high accuracy rates, sometimes they are not suitable for developing personalised treatments by physicians due to their lack of interpretability. This study proposes a novel glucose prediction method emphasising interpretability: Interpretable Sparse Identification by Grammatical Evolution. Combined with a previous clustering stage, our approach provides finite difference equations to predict postprandial glucose levels up to two hours after meals. We divide the dataset into four-hour segments and perform clustering based on blood glucose values for the two-hour window before the meal. Prediction models are trained for each cluster for the two-hour windows after meals, allowing predictions in 15-minute steps, yielding up to eight predictions at different time horizons. Prediction safety was evaluated based on Parkes Error Grid regions. Our technique produces safe predictions through explainable expressions, avoiding zones D (0.2% average) and E (0%) and reducing predictions on zone C (6.2%). In addition, our proposal has slightly better accuracy than other techniques, including sparse identification of non-linear dynamics and artificial neural networks. The results demonstrate that our proposal provides interpretable solutions without sacrificing prediction accuracy, offering a promising approach to glucose prediction in diabetes management that balances accuracy, interpretability, and computational efficiency.**

***Index Terms*—Diabetes, machine learning, system dynamics, symbolic regression, evolutionary computation, neural networks.**

I. Introduction

More than 450 million people have diabetes. There are two main types of diabetes: type I is an autoimmune disease that causes the destruction of insulin-producing cells (beta cells) of the pancreas, while type II appears when there is resistance to insulin action. Insulin-dependent diabetes patients need to estimate, or even better, to predict their blood glucose levels in the near term to manage their condition and prevent complications. Predicting glucose levels can help individuals make informed decisions about their diet, exercise, and especially the insulin and medication they use to maintain their blood glucose within a healthy range. The last decade has seen the rapid spread of new and reliable continuous glucose monitoring systems (CGM) that provide real-time glucose readings and trend data to help individuals adjust their treatment accordingly.

The availability of data from CGMs has led the research in the field. Great efforts have been made in the search for accurate glucose prediction models. Some of them are black-box models [1], others are based on analytical models [2], and most of them provide predictions for a time horizon from 15 to 120 minutes [3]. Among them, symbolic regression (SR) [4] techniques and artificial neural networks (ANNs) obtained very good performance [5]. One of the key challenges in using ANNs for glucose prediction is the lack of interpretability of the solutions they provide, limiting their usefulness in clinical practice. This problem of interpretability has led researchers to explore alternative explainable AI techniques to provide both accurate forecasts and insight for clinicians.

This work aims to explore techniques for deriving finite difference equations that accurately represent the dynamics of blood glucose levels, with the benefit of obtaining interpretable models that can aid the work not only of people with diabetes, but also of diabetes clinicians. To achieve this interpretability, we have developed a variant of grammatical evolution (GE) that produces sparse solutions: Interpretable Sparse Identification by Grammatical Evolution (ISIGE). This technique seeks to integrate the good results in the last years of Grammatical Evolution for blood glucose prediction, [6], [7] and the advantages of a recently proposed method: sparse identification of non-linear

Manuscript received 21 July 2023; revised 5 December 2023 and 25 January 2024; accepted 20 February 2024. Date of publication 28 February 2024; date of current version 7 May 2024. This work was supported by the Spanish Ministerio de Innovación Ciencia y Universidad under Grant PID2021-125549OB-I00, Grant PDC2022-133429-I00, and Grant RTI2018-095180-B-I00. *(Corresponding author: Daniel Parra.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Principe de Asturias Hospital of Alcalá de Henares, Madrid, Spain, under Application No. JAMA 310(20), 2191 (2013) 2013, and performed in line with the World Medical Association of Helsinki.

Daniel Parra, J. Manuel Velasco, and J. Ignacio Hidalgo are with the Department of Computer Architecture and Automatics, Universidad Complutense de Madrid, 28040 Madrid, Spain (e-mail: dparra02@ucm.es; mvelascc@ucm.es; hidalgo@ucm.es).

David Joedicke and Gabriel Kronberger are with the University of Applied Sciences of Upper Austria, 4232 Hagenberg, Austria (e-mail: david.joedicke@fh-hagenberg.at; gabriel.kronberger@fh-hagenberg.at).

Digital Object Identifier 10.1109/JBHI.2024.3371108

dynamics (SINDy). Sparse identification of non-linear dynamics solely from data can also be helpful for finding models for complex biological and physiological phenomena and, additionally, many more scientific domains, for example, in physics and engineering.

SINDy is an optimization-based method that constructs a sparse model of the dynamics of a system as a linear model with a set of non-linear base functions. To achieve sparsity, SINDy adds a sparsity-promoting regularization term which penalizes models with too many non-zero coefficients, forcing the algorithm to select a smaller subset of variables and interactions most relevant to the dynamics of the system. This helps to avoid overfitting, improve the interpretability of the model, and reduce computational complexity [8]. Conceptually, the approach is similar to fast function extraction (FFX) [9] but focuses on system dynamics.

In addition, to refine the models and predictions, we employ a clustering approach to group the glucose time series before meals which give us several scenarios for prediction. Due to the complexity of this situation and the need for a technique with good performance characteristics, we have implemented ISIGE using Dynamic Structured GE (DSGE). To reduce the risk of premature convergence and to improve the diversity of the generated solutions, we have applied the ϵ-*lexicase* selection technique [10] in particular its dynamic version.

In this paper, we describe our technique and analyze the experimental results against SINDy and ANNs, in terms of precision, suitability for diabetes care and interpretability.

The results of this study are significant in several ways:

- The results demonstrate that the proposed ISIGE approach, together with SINDy, outperforms well tested ANNs in terms of prediction accuracy.
- ISIGE is the technique that provides the most accurate predictions. This is particularly important for glucose prediction in diabetes treatment, where accurate and safe predictions are crucial for clinical decision-making.
- The fact that ISIGE and SINDy show similar interpretability suggests that ISIGE can provide interpretable solutions without sacrificing prediction performance and safety.

These results offer a promising new approach to glucose prediction in diabetes management that balances accuracy, interpretability and computational efficiency.

The rest of the paper is organized as follows. Section II presents an overview of the existing literature. In Section III-B, our approach is explained, including the grammar of DSGE, and the iterative numerical evaluation algorithm. Section IV outlines the workflow employed in this research. The experimental setup, data utilized, and the results are detailed in Section V. Finally, Section VI provides the concluding remarks of this study.

II. Related Work

Dynamic models for glucose prediction are mathematical models that can simulate and predict the behavior of blood glucose levels over time. They have been used in a variety of applications, such as personalized glucose control algorithms, prediction of hypoglycemia events, and optimization of insulin therapy. There are three types of dynamic models for glucose prediction:

- *Physiological models [2], [11]:* These models use a set of differential equations to simulate glucose and insulin dynamics in the body. Physiological models are based on our current understanding of the complex interactions between glucose and insulin in the body.
- *Data-driven models [7], [12], [13], [14], [15], [16]:* These are mathematical models that are constructed based on data obtained from individuals with diabetes. These models use statistical and machine-learning techniques to analyze the data and identify patterns that can be used to predict future glucose levels. Different models have been developed in conjunction with CGM systems to make short-term or long-term predictions.
- *Hybrid models [17], [18]:* These models combine physiological and data-driven models to take advantage of the strengths of both types of models to improve the accuracy of glucose predictions.

An important line of research has focused on using adaptive and recursive techniques [19] for continuous glucose control. In [20], the authors explore the use of adaptive model predictive control algorithms for improving artificial pancreas systems. Their algorithms, which continuously update the glucose prediction model, have been improved by adding explicit consideration of prior knowledge related to the effects of meals and physical activity on blood glucose levels in the paper [21]. This approach has been tested using simulated subjects in a multivariable glucose-insulin-physiological variables simulator. The inclusion of physiological variables such as heart rate in this model has been studied in the work of Hobbs et al. [22]. In this paper, the authors conclude that heart rate is a beneficial input for glucose control in the context of intense physical exercise.

The primary reference for GE applied to glucose forecasting is the book chapter from Hidalgo et al. [6], where a new approach for identifying mathematical models that can predict blood glucose levels in people with diabetes using GE is proposed. The approach was evaluated using data from the OhioT1DM dataset [23] containing glucose and physiological data from people with diabetes. The results showed that the proposed system outperformed several baseline approaches regarding accuracy and interpretability, including linear regression, ARIMA models, and ANNs.

Blood glucose prediction is a very active field of research, and although significant progress has been made, several challenges remain to be overcome, including:

- *Individual differences:* the metabolism of each person is unique, and blood glucose dynamics can vary significantly from individual to individual and from the situation of each person (e.g., changes during pregnancy, illness requiring medication administration, etc.).
- *Delayed response:* There is always a variable delayed response between insulin administration or carbohydrate intake and blood glucose value.
- *Sensor limitations:* The accuracy and reliability of glucose sensors can vary over their lifetime, so that the resulting

TABLE I
INPUT VARIABLES IN THIS STUDY

Variable	Description
B_I	Basal insulin
I_B	Insulin bolus.
F_{ch}	Carbohydrate intake.
HR	Heart rate
G	Glucose concentration.
C	Calories burned
S	Steps.

noise can affect the data quality used for training and model prediction.

- *Complex dynamics:* Blood glucose dynamics are influenced by many factors that can interact in complex ways, such as food intake, physical activity, medication, sleep, and stress.

This paper addresses these challenges using two techniques that we consider particularly suitable for glucose prediction: structured grammatical evolution (SGE) and SINDy. These two approaches can handle large and complex datasets with noisy and incomplete data, automatically extract the most relevant inputs, and identify the main mathematical functions to describe glucose dynamics. Additionally, they generate interpretable models that provide the opportunity to be analyzed, allowing researchers better to understand the underlying dynamics of blood glucose managment.

III. METHODS AND TECHNIQUES

A finite difference equation (FDE) is a mathematical expression that describes the difference between a variable y at two discrete time points [24]. An FDE takes the form of (1), where $\vec{x}$ represents a set of input variables involved in the equation, y represents the target variable, and θ are optional calibration parameters.

$$\Delta y(t) = y(t+\Delta t) - y(t) = f(\vec{x}(t), y(t), \theta) \tag{1}$$

FDEs are classified based on their properties, such as linearity, nonlinearity, and order, which is the highest difference in time steps that explicitly appear in the equation. For example, (1) is a first-order equation, while (2) is a second-order equation as it involves the difference between $y(t+2)$ and $y(t)$.

$$y(t+2\Delta t) - y(t) = f(\vec{x}(t)) \tag{2}$$

Due to their simplicity, finite difference equations are handy for modeling dynamic processes which are measured with constant frequency (equidistant time steps), such as blood glucose levels measured by CGM. Equation (3), express the dynamics of glucose values as a finite difference equation problem. The description of the input variables is given in Table I.

$$\begin{aligned}\Delta G(t) &= G(t+\Delta_t) - G(t) \\ &= f(G(t), B_I(t), I_B(t), F_{ch}(t), HR(t), C(t), S(t))\end{aligned} \tag{3}$$

As is often the case in physical systems, we assume that in the glucose system, only a few terms are relevant in defining its dynamics. Therefore, we can consider the governing equations

Algorithm 1: SINDy Algorithm for Finite Difference Equations.

1: **procedure** SINDy($\mathbf{X}, \Delta(t), \lambda$)
2: $\mathbf{X}, \dot{\mathbf{X}} \leftarrow$ History States Variables
3: $\boldsymbol{\Theta}(\mathbf{X}) \leftarrow$ candidate nonlinear functions library matrix
4: $\boldsymbol{\Theta}(\mathbf{x}^T) \leftarrow$ vector of symbolic functions
5: $\mathbf{x}^{P_2} \leftarrow$ quadratic non-linearities in the state x
6: $\boldsymbol{\Theta}\lambda \leftarrow$ add regularization term to $\boldsymbol{\Theta}$ (LASSO)
7: $\boldsymbol{\Xi} \leftarrow \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\xi}_3, \ldots$ sparse vectors of coefficients
8: $\boldsymbol{\xi}k \leftarrow \dot{\mathbf{X}} = \boldsymbol{\Theta}(\mathbf{X})\boldsymbol{\Xi}$, solve sparse regression
9: Return $\dot{\mathbf{x}}_k = \mathbf{f}_k(\mathbf{x}) = \boldsymbol{\Theta}(\mathbf{x}^{P_2})\boldsymbol{\xi}_k$

are scattered in a high-dimensional non-linear function space. In this way, we aim to create a simplified model that accurately captures the essential dynamics of the system while minimising complexity. The SINDy algorithm is particularly suitable for this purpose. In Section III-A, we briefly explain how SINDy works. Then, in Section III-B, we present our proposal, ISIGE, which seeks to incorporate the benefits of SINDy within the paradigm of GE.

A. Sparse Identification of Nonlinear Dynamics (SINDy)

SINDy is a data-driven method for discovering the governing equations of a dynamical system from time series data [8], [25]. In Algorithm 1, we have summarized the main steps for SINDy.

- Data gathering (line 2): The first step is to collect data from the system that we want to model. The data are time-series measurements of the state variables of the system (Table I).
- Construct a library of candidate functions (line 3): The next step is to construct a library of candidate functions that could potentially be part of the governing equations. In the general SINDy algorithm, these functions could include polynomials, trigonometric functions, exponentials, etc, depending on the system being modeled. In this work and based in our prior knowledge of the system, we have chosen up to X^{P_2}, quadratic nonlinearities of X (4).

$$\Theta(X) = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ 1 & X & \vdots & X^{P_2} \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \tag{4}$$

- The development of all the linear combinations (line 5), are grouped in (5) shown at the bottom of the next page, where t_0 represents the first time step, t_1 the subsequent time step, and so on.
- Add regularization (line 6): SINDy uses sequential threshold ridge regression as a regularization. The objective function $||\Theta\xi - \dot{X}_t||_2^2 + \lambda||\xi||_2^2$ is minimized by iteratively performing least squares and masking out elements of the weight array ξ that are below a given threshold λ. In this paper we used $\lambda = 0.5$ which tests have shown to produce the best results.

- Solving the optimization problem (lines 7-8): through a sparse regression algorithm, we identify the combination of terms that best describes the observed dynamics of the system.
- Once the sparse coefficients have been found (line 9), we have the governing equations of the system.

B. Interpretable Sparse Identification by Grammatical Evolution

This study introduces a novel approach, ISIGE, to model blood glucose dynamics using FDEs. Specifically, we utilize DSGE, a variant of GE, to obtain $\hat{G}(t + \Delta(t))$ as expressed in (3).

1) Evolutionary Algorithm: Algorithm 2 outlines the process of obtaining difference equations using ISIGE. The core is an evolutionary algorithm (EA) defined by its parameters: population size (l), maximum number of generations (N), crossover probability (p_c), and mutation probability (p_m). The EA is DSGE and follows several steps. First, the initial population (P) is generated. Then, the solutions are decoded through the grammar (G), obtaining the equation (phenotype). After that, ϵ-Lexicase selection is applied to the population (Algorithm 3) evaluating the individuals on the training dataset D_T. After selection, crossover and mutation are performed to obtain the new population of solutions. This process is repeated N times (generations). We will select the individual whose mean fitness among all training cases is lower to obtain a solution among the final population. ISIGE has three important features that make it different and interesting from other approaches: DSGE, lexicase selection, and Iterative Numerical Evaluation (INE). We explain all these aspects below.

One of the benefits of using DSGE is the ability to obtain explainable expressions, which enables us to analyze the impact and significance of different components in the system. DSGE addresses two limitations of GE. First, it overcomes the low locality problem by ensuring that a slight change in the genotype results in a corresponding small change in the phenotype, thus ensuring high locality. Second, it eliminates redundancy, which occurs when different genotypes produce the same phenotype, by creating a one-to-one mapping between the genotype and the non-terminal. DSGE employs variable-size lists instead of the fixed-size lists used in GE to achieve this.

2) ϵ-Lexicase Selection: We have grouped the data into clusters comprising multiple 4-hour glucose time series segments. The behavior of these segments is highly varied (Fig. 5). If the selection mechanism of the evolutionary algorithm fails to consider this, individuals that do not achieve high fitness values in all the test cases, would be discarded. Because of this, we selected lexicase (and, more specifically, ϵ-Lexicase) as a selection mechanism.

Algorithm 2: ISIGE.

```
1:  Procedure ISIGE( grammar, datasets, GE-properties)
2:    l ← Population size
3:    N ← Max number of generations
4:    p_c ← Probability of crossover
5:    p_m ← Probability of mutation
6:    w ← Max number of wraps
7:    D_T ← Training Dataset
8:    P ← Generate population(l)
9:    for i = 0 to (N - 1) do
10:     P ← ε-Lexicase selection (D_T, P, G, l)
11:     P ← Crossover (P, p_c)
12:     P ← Mutation (P, p_m c)
13:   Best ← p ∈ (P)/min(1/|T| Σ_{t∈T} INE(p, t))
14:   return Best
```

Lexicase selection is a powerful search strategy used in evolutionary computation to overcome the problem of premature convergence. It works by selecting individuals that perform well in a randomly chosen subset of the test cases and repeating the process with the remaining test cases until only one individual is left. However, this method may become too selective, leading to limited search space exploration and the loss of diversity. To address this issue, ϵ-Lexicase [10] selection was proposed, introducing a tolerance parameter, ϵ. This parameter selects individuals that perform well within a certain threshold of the best-performing individuals in each subset. This approach allows for a more comprehensive search space exploration while preserving good individuals.

Algorithm 3 shows how Dynamic ϵ-Lexicase selection works. The objective is to obtain a set of l parents for the following generation. The function *GetParent* is repeated l times. It starts by dividing the training dataset (D_T) into a set of t training cases d_{T_i}. We work with a copy (D'_T) of the training cases and a copy of the population (P'). The process is as follows:

- Choose a dataset d_T randomly from D'_T.
- Evaluate P' by applying INE (see explanation below).
- Assign to $elite$ the value of the best fitness.
- Calculate ϵ_t as the median absolute deviation function of F.
- Remove from P' all the individuals whose fitness value is higher than ($elite$ + ϵ_t)
- Remove d_T from D'_T.
- while there is at least one training set in D'_T and more than one individual in P', repeat the process.

Once the process is finished, an individual is randomly returned from the set P', if there are more than one solution in it.

3) Grammar: Creating a grammar to define the search space is essential in DSGE. This involves establishing how to generate

$$X^{P_2} = \begin{bmatrix} G(t_0) \cdot B_I(t_0) & G(t_0) \cdot I_B(t_0) & G(t_0) \cdot F_C H(t_0) & \cdots & F_C H^2(t_0) \\ G(t_1) \cdot B_I(t_1) & G(t_1) \cdot I_B(t_1) & G(t_1) \cdot F_C H(t_1) & \cdots & F_C H^2(t_1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ G(t_n) \cdot B_I(t_n) & G(t_n) \cdot I_B(t_n) & G(t_n) \cdot F_C H(t_n) & \cdots & F_C H^2(t_n) \end{bmatrix} \quad (5)$$

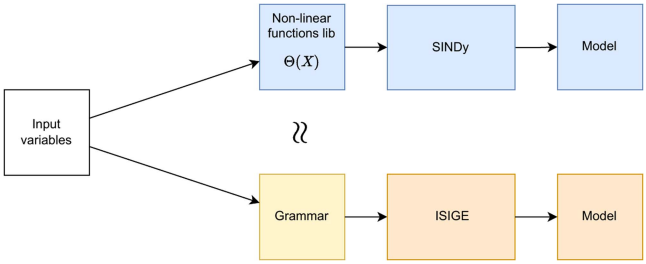


Fig. 1. Common steps between SINDy and ISIGE, our proposal.

Algorithm 3: Dynamic ϵ-Lexicase Selection. λ is the Median Absolute Deviation Function.

function ϵ-LEXICASE SELECTION($\mathbf{D_T}, \mathbf{P}, \mathbf{G}, \mathbf{l}$)
  $\mathbf{P_{aux}} \leftarrow \emptyset$
  for $i = 1$ to $\mathbf{l}$ **do**
    $\mathbf{p_{aux_i}} \leftarrow$ GETPARENT($\mathbf{P}, \mathbf{D_T}$)
    $\mathbf{P_{aux}} \leftarrow \mathbf{P_{aux}} \cup \mathbf{p_{aux_i}}$
  $\mathbf{P} \leftarrow \mathbf{P_{aux}}$
  return $\mathbf{P}$
function GETPARENT($\mathbf{P}, \mathbf{D_T}$)
  $\mathbf{D_T}' \leftarrow \mathbf{D_T}$
  $\mathbf{P}' \leftarrow \mathbf{P}$
  $\mathbf{s} \leftarrow \mathbf{l}$
  while $|\mathbf{D'_T}| \neq \emptyset$ and $|\mathbf{P}'| > 1$ **do**
    $\mathbf{d_T} \leftarrow$ random choice from $\mathbf{D_T}'$
    $\mathbf{F} \leftarrow \emptyset$
    for $i = 1$ to $\mathbf{s}$ **do**
      $\mathbf{g_i} \leftarrow$ decode(grammar, $\mathbf{p'_i}$)
      $\mathbf{f_i} \leftarrow$ INE($\mathbf{g_i}, \mathbf{d_T}$)
      $\mathbf{F} \leftarrow \mathbf{F} \cup \mathbf{f_i}$
    $elite \leftarrow \min(\mathbf{F})$
    $\epsilon_t \leftarrow \lambda(F)$
    for $i = 1$ to $\mathbf{s}$ **do**
      if $\mathbf{f_i} > elite + \epsilon_t$ **then**
        $\mathbf{P}' \leftarrow \mathbf{P}' \setminus \{\mathbf{p'_i}\}$
        $\mathbf{s} = \mathbf{s} - \mathbf{1}$
      $\mathbf{D'_T} \leftarrow \mathbf{D'_T} \setminus \{\mathbf{d_t}\}$
  return random choice from P'

constants, which operations to use, which structures to generate expressions and which variables to incorporate. In Fig. 1, we can see that the set of base functions used in the SINDy algorithm (Section III-A) is equivalent to the ISIGE grammar, but there is an important consideration here. Due to the method used to evaluate the numerical expressions, it is only possible to calculate the values of the variables for one time step at each iteration. In Fig. 2, we present the grammar used in this work in Backus Naur Form (BNF), which is crucial for obtaining the desired FDEs. One of the essential elements of this grammar is the initial definition of the <func> expression, which establishes that all expressions must follow the form G + <expr>, ensuring that the resulting expressions are in the FDE form. Another critical element is the <var> field, which includes individual variables (excluding G, which has already been included),

```
# Model expression
<func> ::= G + <expr>
<expr> ::= (<expr> <op> <expr>) | (<cte> <op> <var>  <op> <
    expr>)
          | <var> | (-<var>)
          | pow(<var>,<sign><exponent>)
          | (-pow(<var>,<sign><exponent>))
<var> ::= B_I|I_B|F_ch|HR|C|S|
          # First order
          G · B_I|G · I_B|G · F_ch|G · HR|G · C|G · S|
          B_I · I_B|B_I · F_ch|B_I · HR|B_I · C|B_I · S|
          I_B · F_ch|I_B · HR|I_B · C|I_B · S|
          F_ch · HR|F_ch · C|F_ch · S|
          HR · C|HR · S|
          C · S|
          # Second order
          G · G|B_I · B_I|I_B · I_B|F_ch · F_ch|HR · HR|C · C|S · S
<op> ::= +|-|·
<cte> ::= <base> · pow(10,<sign><exponent>)
<base> ::= 1|2|3|4| ··· |99
<exponent> ::= 1|2|3|4|5|6|8|9
<sign> ::= +|-
```

Fig. 2. Grammar used for ISIGE in BNF.

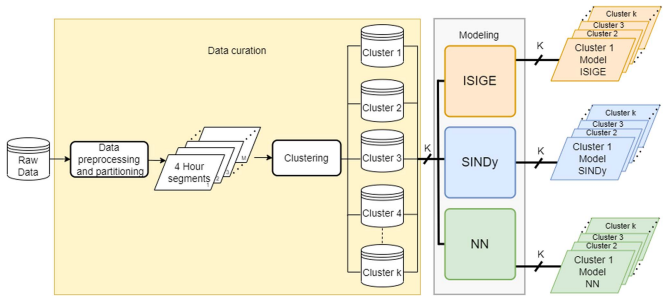


Fig. 3. Workflow diagram of our methodology. The stages represented are: data acquisition, data preprocessing, data partitioning, clustering, modelling.

first-order nonlinearities, and second-order nonlinearities. This approach allows us to mimic the set of base functions used in the SINDy algorithm. Using this grammar, we can effectively generate diverse FDEs that describe the system behaviour and make accurate predictions.

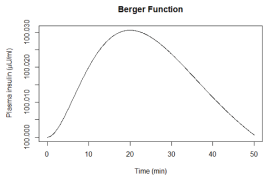
Evolutionary algorithms (EA) do not obtain a single solution because they work with a set of solutions that reach the end of the search or optimization process. Moreover, a pseudo-random process needs several runs to obtain a robust solution. The ability of the EA to obtain a unique solution depends on its ability to obtain the global optimum, which is only sometimes possible. The mechanism of EA to avoid local optima is the mutation operator.

IV. WORKFLOW

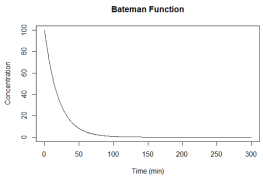
Fig. 3 summarizes the workflow of the methodology applied in this work. It consists on several steps, further described in this section: data acquisition, data preprocessing, data partitioning, clustering, modelling, testing and results comparison.

A. Data Acquisition

For this study, we used data from 24 participants of the Hospital Universitario Príncipe de Asturias, Madrid, Spain. Two different devices were used to obtain the raw data: A continuous



(a) Possible plasma insulin evolution based on Berger equation



(b) Possible insulin concentration evolution based on Bateman Function

Fig. 4. Representation of Berger and Bateman function.

glucose monitor (CGM) system (Free Style Libre sensor) and an activity monitoring wristband (Fitbit Ionic). The CGM measures interstitial glucose levels every 15 minutes, while the wristband records data on calories, steps and heart rate at different time frequencies.

In addition, information about insulin and carbohydrate intakes was recorded by two different methods depending on the insulin administration mode. Participants wearing an automatic insulin continuous infusion system (insulin pumps) obtain this information directly from the device (Medtronic or Roche systems). Participants under multiple doses of insulin (MDI) therapy recorded the information about basal insulin, insulin boluses, and carbohydrate intakes using a mobile application.

The data was recorded in free-living conditions. After a training session with the clinician staff, the participants estimated the carbohydrate count of meals. We proceeded in this way to replicate the real free-living conditions. The error in the estimation is a limitation of this work. The Fitbit device was used for passive collection of physiological during the daily living of the subjects. Participants were trained to wear the Fitbit device during the exercise, and inconsistencies in the data were curated in the preprocessing step. Heart Rate is collected every minute using the activity smartwatch. Heart Rate data is the average of the 15 minutes. Hence we can capture the HR variations due to exercise as an increment in this variable. Calories and steps were also recorded every minute. We reconciled them with CGM data by adding the values every 15 minutes. The participants provided their written informed consent to participate in this study. The study was aproved by the ethical comitee of Alcalá de Henares Hospital, Madrid, Spain (Hospital Universitario Príncipe de Asturias de Madrid), Protocol Number: EC/11/2018, Date of approval: December 12th 2018.

B. Data Preprocessing

Several preprocessing steps had to be performed to use the recorded data to model the blood glucose level. This includes both cleaning as well as feature engineering.

1) Features for the Absorption of Carbohydrates and Insulin: Since the dissolution of substances in the body occurs gradually, we preprocessed both the reported insulin bolus and carbohydrate values using two functions to spread the uptake over multiple observations: the Berger function ((7) and Fig. 4(a)) [26] and the Bateman function ((8) and Fig. 4(b)) [27].

The Berger function is

$$\frac{dA}{dt} = \frac{s \cdot t^s \cdot (a \cdot D \cdot b)^s \cdot D}{t \cdot ((a \cdot D \cdot b)^s + t^s)^2} - A \tag{6}$$

where A is the plasma insulin, D the dose, t the time after the consumption, s the mean absorption rate and a and b are parameters to characterize dependency on D. We use the parameter values $s = 1.6$, $a = 5.2$ and $b = 41$ [26].

The Bateman function is

$$c(t) = f \cdot \frac{D}{V} \cdot \frac{k_a}{k_a - k_e} \cdot (e^{-k_a \cdot t} - e^{-k_e \cdot t}) \tag{7}$$

where k_a is the absorption rate, k_e the elimination rate, D the dose, t the time since consumption, V the volume of distribution and f the bioavailabilty. We used the parameter values $k_a = 0.1$, $k_e = 0.2$, $V = 0.5$ and $f = 0.5$ [27].

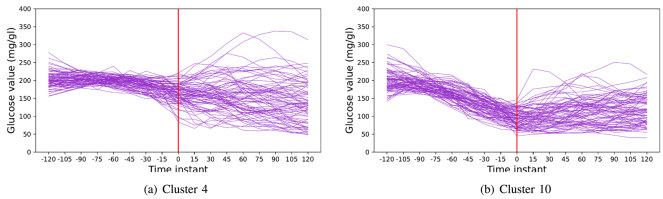
Both functions convert an instant input such as the insulin bolus as well as reported meals that the uptake of the substance into the body does not take place abruptly but is distributed over a longer period of time. It is difficult to estimate the errors caused by the use of these equations. Several factors affect the variability in the absorption and effect of insulin, representing also a source of glucose variability and a challenge in insulin therapy. There is "within-subject variability", since differences exist from one injection to another in the same person. This variability is the sum of a pharmacokinetic component, determined by parameters of equations 8 and 9 and a pharmacodynamic component, determined by insulin's metabolic effects. Moreover, the pharmacokinetic component is also affected by the injection site, and the exercise [28]. We selected the parameters of Humalog Insulin that is still in the market and is the insulin used by most of the participants of our cohort.

2) Time-Shifts: We added time-shifted features for the carbohydrates and the bolus values. The observations have a time shift of thirty minutes to the past, and the data points are recorded at fifteen-minute intervals.

3) Mean Values: Since the heart rate and steps measurements every 15 minutes can change quickly, we smoothed them using a moving average with a window size of 30 minutes (two observations).

4) Creation of Segments: We define a segment as the period of 2 hours before and after a reported meal (i.e. carbohydrate input). Although prediction horizons of 4 and 8 hours are indeed challenging, we are working on a postprandial scenario in this paper. Predictions can be less accurate as the time horizon advances. For postprandial prediction, 2 hours is the usual time to evaluate whether the person needs a correction bolus or not. Regarding the inputs of the models, in this work, we are obtaining *What if models* that are useful for bolus calculations and recommendations. If the future values of carbohydrates and insulin dosing are not known, we would be in an *agnostic* model which is more useful for night and no-prandial situations. Segments may overlap if two meals are taken within a time span of two hours. For each of those segments we check for the following constraints:

1) Each of the parameter values for each time-step has to be greater than 0.



(a) Cluster 4 (b) Cluster 10

Fig. 5. Representation of glucose over time for clusters 4 and 10.

2) We only allow interpolated data up to one hour.
3) A change between time steps must not be more than 25% of the previous value.

The amount of used segments per patient after the preprocessing differs from 2 to 726, with a total of 1444 segments. Each segment contains 17 data points for the data measured from 2 hours before to 2 hours after each meal.

C. Clustering

We used k-means clustering to cluster segments based on the 2-hour window of glucose measurements before the meal. We tried $k \in 3, 5, 7, 9, 11, 13, 15, 17, 19$ and selected $k = 15$ clusters where the intra-distance leveled off, and improved only slightly for higher values of k. 100 random restarts where executed and the cluster assignment with the best intra-distance was used. To give an example of the clusters obtained, Fig. 5 shows the blood glucose concentration for the segments in Clusters 4 and 10. Each graph shows the glucose value over time. We can see how the behavior of the segments is similar for the first two hours between the segments of the respective clusters, and it is not until food intake, the red vertical line, that they begin to vary to a greater extent.

D. Training, Validation and Test Split

For each one of the k clusters, a first division is made into two groups, test with one-third of the cases and, on the other hand, validation-training with the remaining two-thirds. This division will be performed only once, and the same for all the techniques in this article. The validation-training set is then divided into validation, with one-third of the cases in the subgroup and the rest for training. There are sufficient data to have a separate representative test set and we therefore did not use cross-validation.

In Table II, we show the total number of cases for each cluster and the division for training and testing.

E. Modeling

The process for obtaining the results is similar to the other two techniques. In Fig. 6, cluster N is divided into subsets to obtain the results for ISIGE. The process consists of three stages: training, validation, and testing. During the training phase, 30 runs are performed, and the best model for each run is selected based on the mean RMSE (MRMSE) for different segments. In the validation stage, a single model is chosen, and in the testing phase, we obtain the results of the technique. for cluster N. This process is repeated for all clusters to obtain a model for each cluster.

TABLE II
DATA PARTITIONING OF SEGMENTS PER CLUSTER

Cluster	Total	Training-validation	Test
1	99	66	33
2	92	61	31
3	21	14	7
4	75	50	25
5	143	95	48
6	137	91	46
7	44	29	15
8	162	108	54
9	16	10	6
10	75	50	25
11	23	15	8
12	108	72	36
13	183	122	61
14	171	114	57
15	95	63	32
Total	1444	960	484

In addition to our proposal (ISIGE) and SINDy, we have experimented ANNs. As we have already stated, blood glucose prediction is an important and complex task, and ANNs have been proven valuable tools in the last years. ANNs are among the state of the art solutions for prediction of BG, however ANN have not been approved by regulatory agencies yet for use in automated insulin delivery (AID) systems. In a recent literature review on glucose and hypoglycemia prediction methods using data from real patients [3], the authors found that ANNs, particularly recurrent neural networks (RNN) and hybrid models, performed well in predicting glucose levels. Moreover, many software tools in different languages are available on the market, and their use is increasingly ubiquitous in all kinds of problems, making them a recommendable option for the problem of blood glucose level prediction. In another study [5], the performances of several neural network models were compared for predicting blood glucose in patients with diabetes. Following the findings of the previous study, we have chosen, for this work, three ANNS models: Meijner [29], Mirshekarian [30], and Sun [31]. Based on our experimental findings and comparative analysis, it was determined that the ANN version developed by Mirshekarian exhibited superior performance. Therefore, to clearly represent ANNs, only the outcomes for the model of Mirshekarian will be showcased for comparison with SINDy and ISIGE.

V. EXPERIMENTAL RESULTS

The methods studied in this work are ANNs, SINDy, and ISIGE, and their performance is compared with a simple baseline that is calculated by taking the mean of the glucose values at each time step over the segments of the training set. In terms of training times, SINDy is the fastest of the three. However, due to the numerous tests with different parameters, its time increases enormously. Regarding evaluation times, both SINDy and ISIGE generate mathematical expressions with which large volumes of cases can be evaluated quickly. The average execution time for one training run for ISIGE on an Intel(R) Core(TM) i7-4770 with 32 GB RAM is 1.8 h. In the case of ANN, the average times for Meijner, Mirshekarian, and Sun are 22.35, 15.09 and 37.16 seconds for one run, respectively. For SINDy one complete configuration takes around 12.30 minutes on an Intel(R)

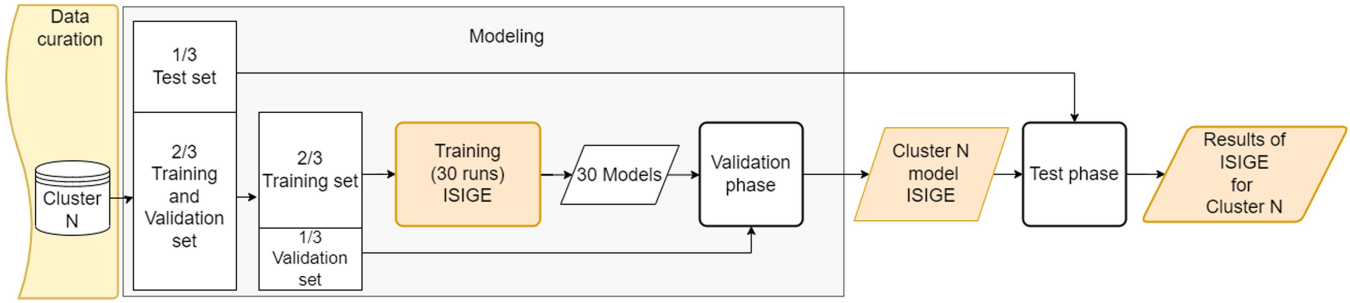


Fig. 6. Generation of the results ϵ-Lexicase ISIGE for cluster N. Division of cluster N into training, validation and test sets used in the homonymous phases.

TABLE III
MRMSE (MG/DL) FOR THE TEST SET FOR ALL METHODS AND CLUSTERS

Cluster ID	# Segments	Mean pred.	ANN	SINDy	ISIGE
1	99	42.14	45.67	52.52	**41.17**
2	92	38.85	43.56	35.45	**32.81**
3	21	**50.25**	52.58	55.65	55.03
4	75	43.63	47.05	44.42	**42.87**
5	143	30.15	32.14	28.73	**26.43**
6	137	24.75	26.96	**22.42**	23.04
7	44	53.84	62.53	**44.52**	48.14
8	162	37.20	38.73	**30.78**	32.44
9	16	71.29	81.12	**44.82**	45.98
10	75	36.34	36.44	30.22	**29.92**
11	23	**43.50**	46.08	45.87	50.48
12	108	53.33	53.17	50.27	**48.65**
13	183	36.17	37.39	34.81	**34.18**
14	171	32.95	34.94	31.50	**30.97**
15	95	27.08	30.34	32.41	**24.72**
Avg.	96.27	41.43	44.58	38.96	37.79

The bold values denote the best results.

TABLE IV
PARKES ERROR GRID ZONES

Zone	Description
A	The actual value and the predicted value are similar.
B	Does not lead to any action by the patient or leads to benign treatment.
C	Overcorrection of acceptable blood glucose levels
D	Dangerous failure to detect and treat errors
E	Erroneous treatment, contradictory to that actually needed

Xeon(R) CPU E5-2670 with 32 GB RAM, where more than 2700 different combinations of parameters are needed.

A. MRMSE Results

Table III shows the MRMSE in the test for each method and all clusters. The MRMSE values of SINDy and SGE are lower than the values for the mean prediction in most cases. The average MRMSE in all clusters is 38.96 [mg/dL] and 37.79 [mg/dL], respectively. The best values for each cluster and the best mean are marked.

B. Parkes Error Grid Analysis

To measure the error of these techniques, we should not only look at the MRMSE. Since we are dealing with a clinical problem, it is necessary to take into account the clinical consequences. An established method for assessing errors made by blood glucose estimation or prediction systems is the Parkes Error Grid (PEG), [32]. In the PEG the actual value of blood glucose and the predicted values are plotted on a grid and associated with risk levels. These levels go from A to E and are described in Table IV.

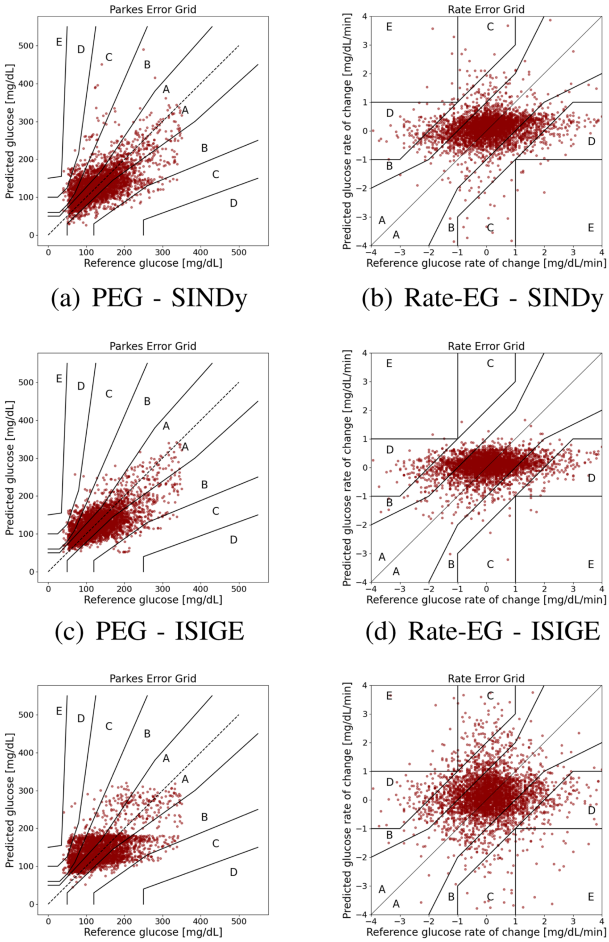
TABLE V
PERCENTAGE OF PREDICTION / MEASUREMENT PAIRS IN THE PARKES ERROR ZONES BY CLUSTER AND METHOD

		Mean prediction				Mirshekarian (ANN)				SINDy				ISIGE			
Cluster	Test segments	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
1	33	48.5	42.4	8.1	1	40.5	47.7	10.6	1.1	45.5	39.8	14.8	0	48.5	42.8	8.7	0
2	31	55.2	40.9	2.9	1.1	46.8	48.4	2.8	2.0	58.1	38.3	3.2	0.4	59.7	35.9	4.4	0
3	7	76.2	23.8	0	0	71.4	28.6	0	0	76.8	14.3	8.9	0	60.7	39.3	0	0
4	25	51.6	39.5	8.9	2.7	41.5	45.0	9.5	4.0	43.0	42.5	13.0	1.5	50.5	38.0	10.5	1
5	48	57.2	30.3	12.5	0	51.3	32.8	15.6	0.3	60.7	35.4	3.9	0	64.1	31.0	4.9	0
6	46	61.6	33.1	5.3	0	57.3	33.7	9.0	0	70.1	25.5	4.3	0	71.2	26.9	1.9	0
7	15	67.4	28.1	3.7	0.7	55.8	38.3	5.0	0.8	70.8	25.0	3.3	0.8	66.7	27.5	5.0	0.8
8	54	51	40.1	8.8	0	44.4	44.2	11.1	0.2	59.3	36.3	4.4	0	56.7	36.3	6.9	0
9	6	59.3	27.8	7.4	5.6	52.1	33.3	8.3	6.2	62.5	31.3	2.1	4.2	62.5	31.2	6.2	0
10	25	42.2	45.8	11.1	0.9	42.0	43.5	13.5	1.0	63.5	27.0	8.0	1.5	56.5	35.0	8.5	0
11	8	52.8	38.9	6.9	1.4	51.6	37.5	9.4	1.6	54.7	40.6	4.7	0	50.0	34.4	15.6	0
12	36	46.9	40.1	9.6	3.4	44.8	38.9	12.5	3.8	46.9	43.8	7.3	2.1	48.6	41.7	8.0	1.7
13	61	56.8	34.6	8.6	0	49.8	38.7	11.5	0	55.3	36.1	8.4	0.2	59.2	33.4	7.2	0.2
14	57	59.8	36.1	4.1	0	53.5	41.0	5.3	0.2	63.2	33.3	3.5	0	62.9	34.2	2.9	0
15	32	61.1	30.2	8.7	0	51.6	37.9	10.5	0	55.1	32.0	11.3	1.6	65.2	32.8	2.0	0
Mean	32.3	56.5	35.3	7.1	1.1	50.3	39.3	9.0	1.4	59.0	33.4	6.7	0.8	58.9	34.7	6.2	0.2

For this work we use the version of PEG designed for type-I diabetes [32]. In addition, Prediction-Error Grid Analysis (PRED-EGA) [33] is a valuable tool for evaluating the accuracy of glucose prediction models in the context of managing Type-1 diabetes. This analysis technique is derived from the earlier continuous glucose-EGA [34], which has been demonstrated as a reliable and robust method for assessing the precision of glucose predictions concerning both nominal glucose values and their corresponding derivatives or rates of change. PRED-EGA comprises two essential components:

- *Point-EGA (P-EGA):* Point-EGA focuses on evaluating the accuracy of the model's predictions concerning nominal glucose values. It provides insights into how well the prediction model aligns with the actual glucose measurements at specific points in time. This component helps determine the model's ability to produce predictions consistently close to the actual glucose values.
- *Rate-EGA (R-EGA):* R-EGA is designed to assess the model's capability to characterize the derivative and rate of change of glucose measurement values over time. It is crucial for a better understanding of the rate at which blood glucose levels are changing. Rate-EGA allows us to evaluate how well the model captures and predicts these changes in glucose levels, which is particularly important for proactive interventions and maintaining stable blood glucose levels.

Table V shows the percentage of cases in each of the zones of the PEG, for each method and divided by clusters. Zone E has been omitted, since its value was zero for all methods in all clusters. Analyzing the results, most of the points are in zones A and B in all the methods, with SINDy standing out as the method



(a) PEG - SINDy (b) Rate-EG - SINDy

(c) PEG - ISIGE (d) Rate-EG - ISIGE

(e) PEG - Mirshekarian-ANN (f) Rate-EG - Mirshekarian-ANN

Fig. 7. Plot showing the PEG and Rate-EG for each method and all clusters.

with the highest mean number of points in zone A, 59.0%, and with a total of 92.4% when adding the means of zones A and B. In addition, we must also highlight the results obtained by ISIGE, achieving 93.6% in total for the sum of the averages of zones A and B. As we move away from zones A and B, the predictions made put the patient at greater risk, with C being not very recommendable, D potentially dangerous, and E leading to an erroneous and extremely dangerous treatment. In this aspect, the models have an average percentage below 10% in zone C, with SGE standing out with only 6.2%. In the case of zone D, SINDy and ISIGE, are below 1.0%, with the lowest being ISIGE, with an average of only 0.2% of the points in this zone.

Fig. 7 shows the PEG and Rate-EG error zones for all the clusters and all methods. If we look at the subfigures of PEG, 7(a), (c) and (e), we can see that most of the points are located between zones A and B. To a minor amount, we can find points in zones C and D, particularly in the left zone of the grid. This zone is related to the actual values associated with hypoglycemia (values below 70), those predictions that give a high glucose value being in this zone are potentially dangerous, so it is of great interest to avoid these situations. On the other hand, it is easier to detect hypoglycemic values compared to hyperglycemic ones. Results obtained using Rate-EG are in subfigures, 7(b), (d), and (f). In the case of SINDy and ISIGE, we can see that most of the points are in section -1 1 of the y-axis, so we can interpret that the models obtained generate predictions with a low rate of change, having some difficulty in capturing abrupt changes in measured glucose (zone D of the Rate-Eg). In the case of the ANNs, more significant variability can be observed in their predictions but with a greater volume of points in zones C, D, and E. The interpretability of machine learning models is a topic of significant interest, yet a consensus on its definition, quantification, and measurement remains elusive. Complexity, transparency, and the ability to be simulated are among the key characteristics often associated with interpretability. However, interpretations may vary depending on the perspective of the evaluator, with mathematicians and healthcare professionals likely to have differing opinions on the same model. In this study, focusing on the healthcare domain, we specifically examine the transparency of models as a crucial aspect of interpretability. Based on the work of Lipton (2018) [35] and Belle (2021) [36], three dimensions are proposed to analyze model transparency: simulatability, decomposability, and algorithmic transparency.

- Simulatability refers to the capacity of the model to be simulated by a human, with simplicity and compactness being key attributes. Simple and compact models are more likely to fall into this category. However, it is essential to note that more than simplicity alone is required, as an excessive number of simple rules can hinder the ability of a human to mentally calculate the decisions of the model. Conversely, some complex models, such as neural networks without hidden layers, may still exhibit simulatability.
- Decomposability entails breaking down the model into interpretable parts, including inputs, parameters, and computations, and subsequently explaining each component. Unfortunately, not all models satisfy this property, making it challenging to explain their inner workings comprehensively.
- Algorithmic transparency focuses on understanding the procedural steps employed by the model to generate its outputs. Models that employ clear procedures, such as similarity-based classifiers like K-nearest neighbours, exhibit algorithmic transparency. On the other hand, complex models like neural networks construct elusive loss functions, and the solution to the training objective often requires approximation. Inspecting the model through mathematical analysis is the primary requirement for it to fall into this category.

Some models inherently possess one or more of these qualities, making them candidates for transparent models. For example, logistic/linear regression models used to predict continuous and categorical targets, where the target is a linear combination of variables, can be considered transparent due to their clear modelling choices. However, to maintain transparency, such models must be limited, and the variables used should be understandable to the intended users. The complexity of the model is directly linked to its transparency. Even if a model satisfies several transparency dimensions, a highly complex procedure

TABLE VI
EXPRESSIONS FOR EACH CLUSTER USING ISIGE

Cluster	Expression
1	$G(t_{n+1}) = G(t_n) - I_B(t_n) \cdot C(t_n) + I_B(t_n) - F_{ch}(t_n) \cdot G(t_n) + 4.6$
2	$G(t_{n+1}) = G(t_n) + I_B(t_n) \cdot \text{HR}(t_n) - G(t_n)^2/4000 + 1/(I_B(t_n) \cdot \text{HR}(t_n))$
3	$G(t_{n+1}) = G(t_n) - 49 \cdot I_B(t_n) \cdot \text{HR}(t_n)/10 + F_{ch}(t_n) \cdot G(t_n) - F_{ch}(t_n) \cdot S(t_n) + 1/(F_{ch}(t_n)^2 \cdot G(t_n)^2)$
4	$G(t_{n+1}) = G(t_n) - B_I(t_n) \cdot C(t_n) - 2 \cdot F_{ch}(t_n) \cdot G(t_n) + F_{ch}(t_n) \cdot \text{HR}(t_n) + 7.8$
5	$G(t_{n+1}) = G(t_n) - B_I(t_n) \cdot G(t_n) + B_I(t_n) \cdot \text{HR}(t_n) - I_B(t_n) \cdot \text{HR}(t_n) + 9.9$
6	$G(t_{n+1}) = G(t_n) + 2 \cdot I_B(t_n) \cdot C(t_n) - I_B(t_n) \cdot G(t_n) - F_{ch}(t_n) \cdot G(t_n) + 8$
7	$G(t_{n+1}) = G(t_n) - I_B(t_n) \cdot S(t_n) + 49/5 - G(t_n)/C(t_n)$
8	$G(t_{n+1}) = G(t_n) - F_{ch}(t_n) \cdot \text{HR}(t_n) + 3.6$
9	$G(t_{n+1}) = G(t_n) - C(t_n) \cdot F_{ch}(t_n) \cdot (-5 \cdot I_B(t_n) \cdot F_{ch}(t_n) \cdot \text{HR}(t_n)^2 + 5 \cdot F_{ch}(t_n) + 91)/5$
10	$G(t_{n+1}) = G(t_n) - I_B(t_n) \cdot G(t_n) + F_{ch}(t_n) \cdot \text{HR}(t_n) + 3.9$
11	$G(t_{n+1}) = G(t_n) + B_I(t_n) \cdot C(t_n) - I_B(t_n) \cdot C(t_n) - I_B(t_n) \cdot G(t_n) - 1/(I_B(t_n) \cdot \text{HR}(t_n))$
12	$G(t_{n+1}) = G(t_n) + B_I(t_n) \cdot C(t_n) - I_B(t_n) \cdot C(t_n) - I_B(t_n) \cdot G(t_n) - F_{ch}(t_n) \cdot G(t_n) + 2.4$
13	$G(t_{n+1}) = G(t_n) - I_B(t_n) \cdot G(t_n) - F_{ch}(t_n) \cdot G(t_n) + F_{ch}(t_n) \cdot \text{HR}(t_n) + 9.49989$
14	$G(t_{n+1}) = G(t_n) - I_B(t_n) \cdot C(t_n) + C(t_n) \cdot F_{ch}(t_n) - F_{ch}(t_n) \cdot G(t_n) + 5.9$
15	$G(t_{n+1}) = G(t_n) - B_I(t_n) \cdot G(t_n) + B_I(t_n) \cdot \text{HR}(t_n) - I_B(t_n) \cdot \text{HR}(t_n) + 7.3027$

or a large number of dimensions can render the model less interpretable.

The definition of transparency by Lipton and Belle aligns with the models obtained by SINDy and ISIGE, suggesting their transparency and explainability. Evaluating the complexity of these models confirms their interpretability, although SINDy models with numerous terms may pose challenges in interpretation due to the increased complexity introduced.

Table VI presents the expressions obtained by ISIGE for the different clusters. Notably, a consistent element observed across all expressions is the inclusion of $G(t_n)$ at the beginning. This common occurrence arises from the grammar utilized during the expression generation process, ensuring that all the expressions obtained conform to the form of FDEs. By imposing this starting point, the resulting expressions adhere to the desired mathematical structure.

To facilitate comprehension of the expressions, they have been simplified using the Python sympy library [37], which enables a more accessible understanding of the obtained mathematical representations. The employment of this library aids in unraveling the complexity of the expressions and enhances their interpretability.

The recurrent occurrence of the variable HR in the expressions is interesting. Previous studies (Rothberg et al., 2016) [38], have extensively investigated the correlation between heart rate and blood glucose levels, establishing a solid connection between these variables. Therefore, the presence of HR in the expressions is justified, considering its established relationship with the target variable. Additionally, it is worth noting that HR often co-occurs with carbohydrate intake in the expressions. This frequent association suggests a potential connection between heart rate, carbohydrate intake, and the target variable. This observation raises intriguing possibilities for future research, as it could provide valuable insights into the relationship among these variables. Further investigations in this direction may be warranted to explore and uncover the significance of this relationship. Furthermore, it is interesting to highlight that in the solution of cluster number 8, the expression includes terms for carbohydrate intake and HR, while the term for insulin bolus is absent. This solution demonstrates satisfactory performance due to the time series behaviour of the cluster. Nevertheless, this finding calls for specialized clinicians to conduct in-depth studies and assess the validity and applicability of such a solution. However, thanks to the transparency of the model, a healthcare specialist could choose to replace this expression with another one that obtained a good result in the validation phase.

The variables in the models are interpretable in both medical and computational terms. However, the models do not necessarily explain known situations, i.e., the physiological variables that appear in a solution may not coincide with what is expected according to classical compartmental models. By contrast, the models can establish conclusions unknown to the medical staff, i.e., a combination of variables that influence the evolution of glucose, which would not be taken into account in a classical model, may appear. Once the real influence of these variables (or a combination of them) has been contrasted, it can lead to novel medical conclusions.

VI. CONCLUSION AND FUTURE WORK

In conclusion, this paper proposes a novel approach to glucose prediction in diabetes management that emphasises interpretability: Interpretable Sparse Identification by Grammatical Evolution (ISIGE). The proposed technique combines clustering with ISIGE to obtain finite difference equations that predict postprandial glucose levels up to two hours after meals.

In this study, we have employed data from 24 different participants with diabetes mellitus type-I. The data were divided into four-hour segments with two hours before and two hours after a meal, i.e. carbohydrate intake. Clustering was performed based on the blood glucose values for the two-hour window before the meal, dividing the cases into 15 clusters. Forecasts were calculated for the two-hour window after the meal. The results of the newly proposed method are compared to SINDy and artificial neural networks (ANN).

Using Parkes Error Grid to quantify the safety and robustness of the predictions, the essential conclusions of this work are:

- Predictions produced by comparison methods (SINDy and ANNs) are in zones A and B, indicating that the predictions of these models are generally safe. ISIGE achieves 93.6% predictions in zones A and B, which is the highest among all methods.
- SINDy has the highest number of predictions (59%) in zone A, and 92.4% in either zone A or B.
- All models have fewer than 10% of predictions in zone C, with ISIGE standing out with only 6.2%. SINDy and ISIGE have below 1.0% of predictions in zone D, with the lowest being ISIGE, with an average of only 0.2%. None of the studied methods has predictions in zone E.
- Although only few predictions are in zones C and D, they represent a risk for participants and should be monitored.

We have used the mean of root mean square error (MRMSE) to quantify the accuracy of the predictions. In this sense, the conclusions are:

- The average MRMSE over all clusters is 38.96 [mg/dL] for SINDy and 37.79 [mg/dL] ISIGE. The results indicate that the models have different performance levels across different clusters.
- Although ANN is one of the best-performing techniques for glucose prediction and one of the most commonly used methods in this field, in our study, both SINDy and ISIGE obtained better results in most clusters.

Transparency plays a pivotal role in enhancing the interpretability of machine learning models in the healthcare domain.

By focusing on the dimensions of simulatability, decomposability, and algorithmic transparency, we can assess and analyze the transparency of models. Transparent models exhibit simplicity, compactness, and transparent procedures, making them more amenable to human understanding. The insights gained from this study provide valuable guidance for future developments in interpretability, helping to advance the explainability of machine learning models in healthcare and beyond.

The use of personalized models that use longitudinal data from one individual can be considered an alternative to the "population" model and should be explored as future work. In future versions of our proposal, we will devote part of our efforts to optimizing the code, which can significantly reduce the execution time by modifying certain aspects of the evaluation linked to the tool. In this current research, we have explored the application of K-means clustering to group pre-meal glucose time series by their numerical values to aid in predicting future blood glucose levels. This approach has proven effective in identifying distinct patterns associated with glycemic control. However, we acknowledge the importance of considering both the value and shape of time series data for a more comprehensive understanding of glucose dynamics. In future work, we intend to focus on a combined clustering approach. In this line of research, we would first group time series by their numerical values to capture the range within which glucose values fluctuate. Subsequently, we would apply a shape-based clustering technique, such as K-shape [39], to further refine the clusters based on the shape of the time series. This combined method may offer a more refined and nuanced analysis of glycemic patterns, potentially improving the accuracy of blood glucose prediction models [40]. The transparency and explainability features of the proposed technique hold great promise for addressing two critical challenges in the field of healthcare monitoring systems, such as the handling of missing data, or errors in continuous glucose monitors [41]. This line of research deserves to be further developed in the future. We conclude that ISIGE can potentially assist diabetes clinicians in developing personalised treatment plans for their patients, offering a promising alternative to traditional black-box models like ANNs.

Acknowledgment

The authors would like to thank Stephan Winkler for the discussions at the beginning of this research and to Felix Tena for the implementation of the ANNs.

References

[1] S. Mirshekarian, H. Shen, R. Bunescu, and C. Marling, "LSTMs and neural attention models for blood glucose prediction: Comparative experiments on real and synthetic data," in *Proc. IEEE 41st Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2019, pp. 706–712.

[2] C. Cobelli and A. Mari, "Validation of mathematical models of complex endocrine-metabolic systems. A case study on a model of glucose regulation," *Med. Biol. Eng. Comput.*, vol. 21, no. 4, pp. 390–399, Jul. 1983.

[3] V. Felizardo, N. M. Garcia, N. Pombo, and I. Megdiche, "Data-based algorithms and models using diabetics real data for blood glucose and hypoglycaemia prediction–a systematic literature review," *Artif. Intell. Med.*, vol. 118, 2021, Art. no. 102120. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0933365721001135

[4] S. Contador, J. M. Velasco, O. Garnica, and J. I. Hidalgo, "Glucose forecasting using genetic programming and latent glucose variability features," *Appl. Soft Comput.*, vol. 110, 2021, Art. no. 107609. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494621005305

[5] F. Tena, O. Garnica, J. Lanchares, and J. I. Hidalgo, "Ensemble models of cutting-edge deep neural networks for blood glucose prediction in patients with diabetes," *Sensors*, vol. 21, no. 21, 2021, Art. no. 7090. [Online]. Available: https://www.mdpi.com/1424-8220/21/21/7090

[6] J. I. Hidalgo et al., "Identification of models for glucose blood values in diabetics by grammatical evolution," in *Handbook of Grammatical Evol.*, Cham, Switzerland: Springer, 2018, pp. 367–393, doi: 10.1007/978-3-319-78717-6_15.

[7] S. Contador, J. M. Colmenar, O. Garnica, J. M. Velasco, and J. I. Hidalgo, "Blood glucose prediction using multi-objective grammatical evolution: Analysis of the "agnostic" and "what-if" scenarios," *Genet. Program. Evolvable Mach.*, vol. 23, pp. 161–192, 2022.

[8] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2022.

[9] T. McConaghy, *FFX: Fast, Scalable, Deterministic Symbolic Regression Technology*. New York, NY, USA: Springer, 2011, pp. 235–260.

[10] W. La Cava, T. Helmuth, L. Spector, and J. H. Moore, "A probabilistic and multi-objective analysis of lexicase selection and ε-lexicase selection," *Evol. Comput.*, vol. 27, no. 3, pp. 377–402, 2019.

[11] C. Dalla Man, R. A. Rizza, and C. Cobelli, "Meal simulation model of the glucose-insulin system," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 10, pp. 1740–1749, Oct. 2007.

[12] V. Felizardo, N. M. Garcia, N. Pombo, and I. Megdiche, "Data-based algorithms and models using diabetics real data for blood glucose and hypoglycaemia prediction–a systematic literature review," *Artif. Intell. Med.*, vol. 118, May 2021, Art. no. 102120.

[13] T. Zhu, K. Li, P. Herrero, J. Chen, and P. Georgiou, "A deep learning algorithm for personalized blood glucose prediction," in *Proc. KHD, IJCAI*, 2018, pp. 64–78.

[14] K. Li, J. Daniels, C. Liu, P. Herrero, and P. Georgiou, "Convolutional recurrent neural networks for glucose prediction," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 2, pp. 603–613, Feb. 2020.

[15] B. Lobo, L. Farhy, M. Shafiei, and B. Kovatchev, "A data-driven approach to classifying daily continuous glucose mo nitoring (CGM) time series," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 2, pp. 654–665, Feb. 2022.

[16] T. Hamdi et al., "Artificial neural network for blood glucose level prediction," in *Proc. IEEE Int. Conf. Smart, Monitored Controlled Cities*, 2017, pp. 91–95.

[17] R. Hovorka et al., "Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes," *Physiol. Meas.*, vol. 25, no. 4, pp. 905–920, 2004. [Online]. Available: http://stacks.iop.org/0967-3334/25/i=4/a=010

[18] C. Liu et al., "Long-term glucose forecasting using a physiological model and deconvolution of the continuous glucose monitoring signal," *Sensors*, vol. 19, no. 19, Oct. 2019, Art. no. 4338.

[19] K. Turksoy, E. S. Bayrak, L. Quinn, E. Littlejohn, D. Rollins, and A. Cinar, "Hypoglycemia early alarm systems based on multivariable models," *Ind. Eng. Chem. Res.*, vol. 52, no. 35, pp. 12329–12336, Sep. 2013, doi: 10.1021/ie3034015.

[20] X. Sun, M. Rashid, M. R. Askari, and A. Cinar, "Latent variables model based MPC for people with type 1 diabetes," *IFAC-PapersOnLine*, vol. 54, no. 3, pp. 294–299, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405896321010302

[21] X. Sun, M. Rashid, N. Hobbs, R. Brandt, M. R. Askari, and A. Cinar, "Incorporating prior information in adaptive model predictive control for multivariable artificial pancreas systems," *J. Diabetes Sci. Technol.*, vol. 16, no. 1, pp. 19–28, Dec. 2021.

[22] N. Hobbs, I. Hajizadeh, M. Rashid, K. Turksoy, M. Breton, and A. Cinar, "Improving glucose prediction accuracy in physically active adolescents with type 1 diabetes," *J. Diabetes Sci. Technol.*, vol. 13, no. 4, pp. 718–727, Jan. 2019.

[23] C. Marling and R. Bunescu, "The OhioT1DM dataset for blood glucose level prediction: Update 2020," *CEUR Workshop Proc.*, vol. 2675, pp. 71–74, 2020.

[24] A. Sharkovsky, Y. L. T. Maistrenko, and E. Y. Romanenko, *Difference Equations and Their Applications*, vol. 250. Berlin, Germany: Springer, 2012.

[25] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proc. Nat. Acad. Sci.*, vol. 113, no. 15, pp. 3932–3937, 2016.

[26] M. Berger, H. Cüppers, H. Hegner, V. Jörgens, and P. Berchtold, "Absorption kinetics and biologic effects of subcutaneously injected insulin preparations," *Diabetes Care*, vol. 5, no. 2, pp. 77–91, 1982.

[27] E. R. Garrett, "The Bateman function revisited: A critical reevaluation of the quantitative expressions to characterize concentrations in the one compartment body model as a function of time with first-order invasion and first-order elimination," *J. Pharmacokinetics Biopharmaceutics*, vol. 22, no. 2, pp. 103–128, Apr. 1994, doi: 10.1007/BF02353538.

[28] A. K. J. Gradel et al., "Factors affecting the absorption of subcutaneously administered insulin: Effect on variability," *J. Diabetes Res.*, vol. 2018, 2018, Art. no. 1205121.

[29] C. Meijner and S. Persson, "Blood glucose prediction for type 1 diabetes using machine learning long short-term memory based models for blood glucose prediction," M.S. thesis, Chalmers Univ. Technol. Univ. Gothenburg. 2017. Accessed: Mar. 3, 2024. [Online]. Available: https://publications.lib.chalmers.se/records/fulltext/251317/251317.pdf

[30] S. Mirshekarian, R. Bunescu, C. Marling, and F. Schwartz, "Using LSTMs to learn physiological models of blood glucose behavior," in *Proc. IEEE 39th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2017, pp. 2887–2891.

[31] Q. Sun, M. V. Jankovic, L. Bally, and S. G. Mougiakakou, "Predicting blood glucose with an LSTM and BI-LSTM based deep neural network," in *Proc. IEEE 14th Symp. Neural Netw. Appl.*, 2018, pp. 1–5.

[32] J. L. Parkes, S. L. Slatin, S. Pardo, and B. H. Ginsberg, "A new consensus error grid to evaluate the clinical significance of inaccuracies in the measurement of blood glucose," *Diabetes Care*, vol. 23, no. 8, pp. 1143–1148, 2000.

[33] X. Yu et al., "Online glucose prediction using computationally efficient sparse kernel filtering algorithms in type-1 diabetes," *IEEE Trans Control Syst. Technol.*, vol. 28, no. 1, pp. 3–15, Jun. 2020.

[34] B. P. Kovatchev, L. A. Gonder-Frederick, D. J. Cox, and W. L. Clarke, "Evaluating the accuracy of continuous glucose-monitoring sensors: Continuous glucose-error grid analysis illustrated by TheraSense freestyle navigator data," *Diabetes Care*, vol. 27, no. 8, pp. 1922–1928, Aug. 2004.

[35] Z. C. Lipton, "The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery," *Queue*, vol. 16, no. 3, pp. 31–57, Jun. 2018, doi: 10.1145/3236386.3241340.

[36] V. Belle and I. Papantonis, "Principles and practice of explainable machine learning," *Front. Big Data*, vol. 4, 2021, Art. no. 688969. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fdata.2021.688969

[37] A. Meurer et al., "SymPy: Symbolic computing in python," *PeerJ Comput. Sci.*, vol. 3, 2017, Art. no. e103.

[38] L. J. Rothberg, T. Lees, R. Clifton-Bligh, and S. Lal, "Association between heart rate variability measures and blood glucose levels: Implications for noninvasive glucose monitoring for diabetes," *Diabetes Technol. Therapeutics*, vol. 18, no. 6, pp. 366–376, 2016.

[39] J. Paparrizos and L. Gravano, "K-Shape: Efficient and accurate clustering of time series," *SIGMOD Rec.*, vol. 45, no. 1, pp. 69–76, Jun. 2016, doi: 10.1145/2949741.2949758.

[40] X. Sun et al., "Prior informed regularization of recursively updated latent-variables-based models with missing observations," *Control Eng. Pract.*, vol. 116, 2021, Art. no. 104933. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0967066121002100

[41] J. Feng, K. Turksoy, S. Samadi, I. Hajizadeh, and A. Cinar, "Hybrid online sensor error detection and functional redundancy for artificial pancreas control systems," *IFAC-PapersOnLine*, vol. 49, no. 7, pp. 753–758, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405896316304840