Can Clustering Improve Glucose Forecasting with Genetic Programming Models?

Sergio Contador
Universidad Complutense de Madrid
Madrid, Spain
scontador@ucm.es

J. Ignacio Hidalgo*
Universidad Complutense de Madrid
Madrid, Spain
hidalgo@ucm.es

Oscar Garnica
Universidad Complutense de Madrid
Madrid, Spain
ogarnica@ucm.es

J. Manuel Velasco
Universidad Complutense de Madrid
Madrid, Spain
mvelascc@ucm.es

Juan Lanchares
Universidad Complutense de Madrid
Madrid, Spain
julandan@ucm.es

ABSTRACT

This study investigates how to improve the predictions of glucose values obtained with genetic programming models. A set of statistical techniques are used to discover glucose profiles that identify similar situations in patients with type 1 diabetes mellitus, and incorporate this knowledge to the models. Glucose time series are divided into 4-hour non-overlapping slots and clustered using the technique based on decision trees called chi-square automatic interaction detection, to classify glucose profiles into groups using two decision variables: day of the week and time slot of the day. The objective is to customize models for different glucose profiles that appear in the patient's day-to-day. Genetic programming models created with glucose values from the original data-set are compared to those of models created with classified glucose values. Significant differences (p-value < 0.05) and associations are observed between the glucose profiles. In general, using classified glucose values in models created with genetic programming, the accuracy of the predictions improves in comparison with those of models created with the original data-set. We concluded that the classification process can be useful to correct and improve habits or clinical therapies in patients, and obtain more accurate models through automatic learning techniques and artificial intelligence.

CCS CONCEPTS

• **Computing methodologies** → **Modeling methodologies**; *Artificial intelligence*; *Ensemble methods*; • **Applied computing** → **Health informatics**; *Consumer health*;

KEYWORDS

Diabetes mellitus. Continuous glucose monitoring. Clustering. Classification. Chi-square automatic interaction detection. Genetic programming. Symbolic regression. Akaike information criterion. Parkes error grid.

ACM Reference Format:
Sergio Contador, J. Ignacio Hidalgo, Oscar Garnica, J. Manuel Velasco, and Juan Lanchares. 2019. Can Clustering Improve Glucose Forecasting with Genetic Programming Models?. In *Genetic and Evolutionary Computation Conference Companion (GECCO '19 Companion), July 13–17, 2019, Prague, Czech Republic.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3319619.3326809

1 INTRODUCTION

Diabetes Mellitus (DM) is a chronic disease of high prevalence in the world population, which increases the morbidity and mortality of people suffering from it and causes a significant deterioration in the quality of life [13]. According to estimates from the World Health Organization [7], 422 million adults worldwide suffered from DM in 2014, compared to 108 million in 1980.

Diabetes is a group of diseases characterized by a high level of glucose resulting from defects in the body's ability to produce or use insulin, a natural hormone that produces the pancreas, necessary to convert sugar, carbohydrates and other foods into the energy necessary for daily life. Insulin is prescribed to many people with diabetes, either because their body does not produce insulin (Type 1 Diabetes Mellitus T1DM) or because the body does not produce enough insulin or the cells do not use it (Type 2 Diabetes Mellitus T2DM). T2DM is the most common type of diabetes, while only 5 % of people have T1DM. With the help of insulin therapy and other treatments, patients can learn how to improve the control of their disease and have a long, healthy and happy life.

T1DM is a chronic autoimmune disorder whereby the immune system attacks the insulin-secreting cells of the pancreas. The result is that the cells do not assimilate the sugar and, as a result, there is an increase in blood glucose levels which is known as hyperglycemia (blood glucose levels above 180 mg/dl). If this situation lasts for a long period of time, the patient may develop severe long-term complications, including heart disease, blindness, kidney failure or foot ulcers [2, 8]. This disease can only be treated with synthetic insulin injected into the bloodstream.

*Corresponding author

GECCO '19 Companion, July 13–17, 2019, Prague, Czech Republic

ACM ISBN 978-1-4503-6748-6/19/07…$15.00
https://doi.org/10.1145/3319619.3326809

However, this is not an easy task. An excessive dose of insulin can cause hypoglycemia (defined as a blood glucose value less than 70 mg/dl). If the hypoglycemia is severe (glucose less than 50 mg/dl), it can lead to loss of consciousness and a diabetic coma.

For these reasons, diabetics must control blood glucose levels throughout their lives, trying to keep them at normal levels, similar to those of a subject who does not suffer from this disease. This is quite complicated, especially in patients with T1DM. When patients are going to have a meal, they need to estimate the units of insulin that will be injected so that, after the meal, glucose levels remain within a healthy range. The estimate should be made based on many factors, but mainly the patient should know their glucose level at that time, and estimate the amount of food consumed, usually measured in carbohydrate rations. Therefore, the control of blood glucose in patients with T1DM requires the prediction of future glucose values, values that depend on the amount of food intake and insulin and/or glucagon. From a scientific point of view, this process is complex and is not clearly defined by the variables involved. Fortunately, the latest advances in the field of glucose modeling allow the automation of some parts of the process making the task easier for diabetic patients.

This research is motivated by the challenge of accurately predicting future glucose levels so that an automated or manual system can decide when and how much insulin to inject in order to maintain blood sugar levels in a healthy range. At the same time, it is imperative to avoid predictions that may trigger unnecessary treatments or, even worse, treatments that are harmful to the patient.

The main objective of this research is to improve the predictions of glucose obtained with Genetic Programming (GP) models. For this, we split the glucose time series into 4-hour non-overlapping slots and cluster them using decision trees, as a step previous to the generation of the prediction models. The aim is to customize models for different glucose profiles that appear in the patient's day-to-day.

The rest of the paper is organized as follows. Section 2 explains the state of the art. Section 3 describes the techniques used for the modeling of glucose and for calculating the accuracy and quality of the models obtained. Section 4 explains the data and the methodology used. The experimental results are shown in section 5. The conclusions and future work are discussed in section 6.

2 STATE OF THE ART

A solution that has proven to be suitable for predicting glucose levels is the use of algorithms based on Grammatical Evolution (GE) [10]. However, one of the main obstacles that we found to train GE models is the scarcity of significant amounts of data. As in many other fields of medicine, the collection of data on real diabetic patients is very complex. GE models trained with a small data-set usually suffer from over-fitting and have limited predictive power. To avoid this situation, in [27] it has been proposed to increase the data records of glucose with synthetic data with good results. Other comparisons have also been made between techniques related to GE, such as GP with strict selection. However, the latter has a high execution time. Another solution that works well is presented in [11] where models based on GP are used to make predictions up to 120 minutes. Although there are some classic approaches, there is still much work to be done for predictions within a horizon of more than 60 minutes.

In [3], the authors propose an improved a method for predicting the patient's blood glucose trend based on the minimal model proposed in [25], where the parameters are obtained using a genetic algorithm. Predictions are made for a 24-hour horizon. Leading research groups on artificial pancreas have presented other personalized control approaches [9, 12, 17, 20] that follow the clinical practice. Some proposals provide models for the average case [21], and others for the particularities of each patient. Several papers apply classical modeling techniques, resulting in models or profiles defined by linear equations with a limited set of inputs [15, 22].

A work on GP-based induction of a glucose-dynamics model for telemedicine is presented in [6]. The work aims to create a regression model that allows the determination of blood glucose values from interstitial glucose in patients with T1DM, using it in a telemedicine portal. Blood glucose values in the database are expanded using the Steil-Rebrin model [26]. To make the most accurate estimate, the parameters of this equation are adjusted using an evolutionary algorithm with root mean square error as the fitness function.

3 BACKGROUND

3.1 Chi-Square Automatic Interaction Detection

Decision trees are techniques that explore the data to get hidden information. The objective under the construction of a decision tree is to create a model to predict the value of a dependent/objective variable on the independent/predictor variables considered. The decision tree has three types of nodes, namely the root node, the internal nodes and the terminal nodes, each representing a class. Each path in the construction of the decision tree is associated with a decision rule established by the algorithm itself. Thus, according to the established rules, the data-set is recursively divided into independent subsets of smaller data (divisive algorithm). One of the widely used algorithms is CHAID (CHi-square Automatic Interaction Detection) [14]. This algorithm recursively partitions the data by means of a target variable using multiple divisions between the different input variables. A division must reach a threshold level of significance, using the independence test χ^2 between the nominal values of the target variable and the branches, if not the node is not divided. The Bonferroni setting is used for the number of categorical values of the input variable, thus mitigating the bias towards entries with many values. It also uses significance tests as the Gini index to determine the number of branches. The search ends when the algorithm can no longer join more branches or there are no significant divisions. The last division is chosen as the solution. Note that typically the last division is not the most significant division examined.

3.2 Symbolic Regression by Tree-Based Genetic Programming

GP is a method for automated synthesis of computer programs based on the concept of evolutionary computation [18]. It uses principles of natural evolution and simulates natural selection, breeding and random mutation to evolve a population of computer programs starting from a randomly initialized population. The objective is to find a computer program that solves a given programming problem usually specified using a set of test cases by simulating many generations. One task where genetic programming has proven to be particularly effective is in symbolic regression, where the objective is to identify a prediction model for a real-valued target variable, and the model is represented as a mathematical function. The structure of the model, as well as its parameters, must both be determined by the solution method to fit the given data-set.

The set of terminals uses the time series of glucose, insulin, carbohydrates and constants. Each value is multiplied by a real value weight that is initialized randomly using a Normal N distribution (μ = 1, σ = 1), and is randomly mutated by adding a sampled value of N (μ = 0, σ = 0.05). Constants with real values are initialized randomly using a uniform distribution U (-20.0, 20.0), and the constant mutation adds a sampled value of N (μ = 0, σ = 1.0).

The set of functions used are: $\{+, -, *, /, log(x), exp(x)\}$, where the protected variants of the functions division and logarithm are used [18]. The models in the initial population are generated using *Probabilistic Tree Creator* [19] with a limit to the maximum depth and maximum number of nodes allowed for the trees. The same depth and size restrictions are applied in the crossover and mutation operations. The crossover uses a sub-tree crossover operator. The mutation uses a variety of operators that replace a complete subset of the tree with a tree initialized randomly, or mutate all the nodes of the tree, or mutate only one node of the selected tree in a random way. The mutation operator selected randomly is executed after each crossover operation with a different mutation rate. A maximum number of generations and parent's proportion aptitude selection is used. The objective function in all cases is calculated with R^2 Pearson's coefficient of determination between the actual blood glucose values and the values obtained with the model [16]. The selected prediction models are linearly scaled to minimize the sum of the quadratic errors between the real values and the results of the model. The parameters of the GP were not tuned specifically for this task. Robust standard configurations were applied.

The models in GP are generated using the time series of glucose, insulin, and carbohydrates. In addition to these series, a set of features described below are used.

Data are added in periods of time calculated using the following equation:

$$mean(X, t, range) = \frac{\sum_{t \in range}(X_t)}{n}, \quad range \in [t_1, ..., t_n] \quad (1)$$

The objective is to reduce the number of values used to generate the models. For each time t the set of features $F(t)$ that describe the historical values F_{his} of the time series (values of glucose G, insulin I, and carbohydrates C) up to t, as well as the future values F_{fut} of insulin and carbohydrates are defined as:

$$\begin{aligned} F(t) = F_{his}(G,t) \cup F_{his}(I,t) \cup F_{his}(C,t) \cup \\ F_{fut}(I,t) \cup F_{fut}(C,t) \cup \{G, I, C\} \end{aligned} \quad (2)$$

$$\begin{aligned} F_{his}(X,t) = \{&mean(X,t,[-15,0]), mean(X,t,[-30,-15]), \\ &mean(X,t,[-45,-30]), mean(X,t,[-60,-45]), \\ &mean(X,t,[-90,-60]), mean(X,t,[-120,-90]), \\ &mean(X,t,[-150,-120]), mean(X,t,[-180,-150]), \\ &mean(X,t,[-210,-180]), mean(X,t,[-240,-210])\} \end{aligned} \quad (3)$$

$$\begin{aligned} F_{fut}(X,t) = \{&mean(X,t,[0,15]), mean(X,t,[15,30]), \\ &mean(X,t,[30,45]), mean(X,t,[45,60]), \\ &mean(X,t,[60,75]), mean(X,t,[75,90]), \\ &mean(X,t,[90,105]), mean(X,t,[105,120]), \\ &mean(X,t,[120,135]), mean(X,t,[135,150]), \\ &mean(X,t,[150,165]), mean(X,t,[165,180]), \\ &mean(X,t,[180,195]), mean(X,t,[195,210]), \\ &mean(X,t,[210,225]), mean(X,t,[225,240])\} \end{aligned} \quad (4)$$

3.3 Akaike Information Criterion

AIC (Akaike Information Criterion) [1] is a method used to measure the relative quality of models based on the entropy of information. It provides a relative estimate of the information lost when a model is used to represent the process that generates the data. The criterion is defined by the following equation:

$$\text{AIC} = n \cdot \log(\text{SE} + 1) + 2(m + 1) \quad (5)$$

where n is the sample size, m is the number of model parameters and SE is the residual quadratic error, which is defined by the following equation:

$$\text{SE} = \sum_{i=1}^{n} (X_i - Y_i)^2 \quad (6)$$

where X_i is the actual value of the observable at sample i, and Y_i is the estimated value in the prediction.

The model that obtains the lowest AIC value is the model that best represents the process that generates the data. When the sample size is small, the following correction is used [5]:

$$\text{AIC}_c = \text{AIC} + \frac{2(m+1)(m+2)}{n-m-2} \quad (7)$$

3.4 Parkes Error Grid

Parkes error grid [23] was published in 2000 as an alternative to Clarke error grid [4]. These methods were developed to calculate the clinical accuracy of continuous blood glucose monitoring systems for the entire range of glucose values, using the differences between the reference values and the values measured by the measurement system. Analogously, it can be used to calculate the differences between the values estimated in a prediction and the current or reference values. The values are represented in

a graph with cartesian coordinates where the X axis represents the reference values and the Y axis the values of the prediction, where Y = X is the ideal prediction. The special feature of this representation is that the graph is divided into five zones depending on the degree of accuracy of the glucose estimates. The difference between the two methods lies in the definition of the zones. In Parkes error grid the zones are redefined based on the zones of Clarke error grid and in the limits established by 100 medical experts in diabetes in a survey carried out in the conference *American Diabetes Association Meetings* in June 1994. The new zones are defined as follows:

- Zone A: glucose values without effect on clinical action. The estimates are accurate.
- Zone B: glucose values with alteration of the clinical action, or with little or no effect in the clinical treatment.
- Zone C: glucose values with alteration of the clinical action and with probability of effects in the clinical treatment.
- Zone D: glucose values with alteration of the clinical action and with probability of having a significant medical risk.
- Zone E: glucose values with alteration of the clinical action and with probability of having dangerous consequences.

These definitions are more flexible than in Clarke error grid, and allow the incorporation of expert knowledge to assign the zones based on experience. In addition, this new method has been introduced as an evaluation tool accepted in ISO15197:2013 [24] of requirements for blood glucose monitoring systems.

The objective is to maximize the predictions included in zones A and B, and minimize the predictions included in zones C, D and E.

4 METHODOLOGY

Fourteen patients with T1DM were selected on the data collection process. First, data collected from a group of four patients with glucose measurements every five minutes using Continuous Glucose Monitoring (CGM) was analyzed and recorded with the intakes and carbohydrate estimates, as well as the insulin doses injected by the patients. Furthermore, the study was extended to ten new subjects with a greater number of records collected with the same device, in order to increase the representation of the results. At the moment of the submission, results with only one of the patients is presented. The data-set contains 37854 observations distributed in 148 days. We expect to present a complete study during the Workshop.

We compare two different methods. The first one is explained in Figure 1 and is the motivation of this work. We will refer to this approach as *CHAID-GP-AIC* in the rest of the paper. It consists on three steps. The first step is to classify the glucose values in the data-set with CHAID [1]. In the application of the algorithm, the original data-set is divided into two parts using stratified random sampling. For the construction of the decision trees, a maximum tree depth of 3, a minimum number of cases in the parent node of 100 and 50 in the child node are selected. Glucose values are used as the dependent variable, and the days of the week and the different time slots as the independent variables.

[1]We use IBM SPSS predictive analysis software.

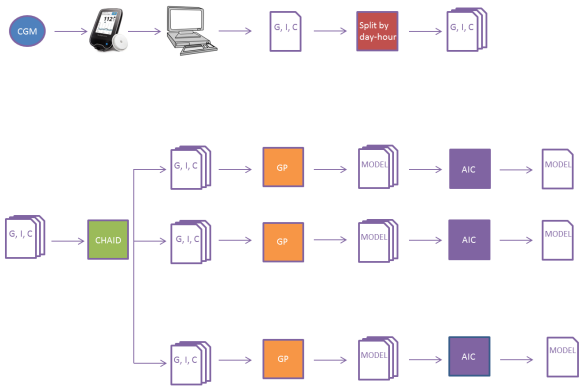

Figure 1: Flow diagram describing the creation process of models based on *CHAID-GP-AIC*.

There are seven categories for the variable day of the week and six categories for the variable time slot. The categories of the variables are represented by letters and numbers as shown in Table 1. Once the classification is obtained, the groups of glucose values are selected as training data-sets.

Table 1: Codes for the variables day of the week and time slot.

day of the week	identifier
Monday	M
Tuesday	T
Wednesday	W
Thursday	R
Friday	F
Saturday	S
Sunday	U

time slot	identifier
00:00h-04:00h	0
04:00h-08:00h	1
08:00h-12:00h	2
12:00h-16:00h	3
16:00h-20:00h	4
20:00h-24:00h	5

The second step is to create GP models for the different data-sets obtained in the previous step. Each model is generated using the open source software HeuristicLab version 3.3.15 [28]. The parameters selected for each model are the same. A population size of 1500 individuals, a maximum tree depth of 11, a maximum number of generations and a maximum number of nodes of 100, and a mutation rate of 0.15 are selected. We obtain models for future glucose values at intervals of 30 minutes for a maximum of 4 hours (time horizons at 30, 60, 90, 120, 150, 180, 210 and 240 minutes). Cross-validation of 10 iterations is used to create the models (*10-fold Cross-Validation*), where each iteration is repeated 10 times.

The final step is to select the best model of the 10 repetitions by the AIC criterion (the best model is the model with the lowest AIC value). We select one model for each time horizon and fold.

The second method is explained in Figure 2 and correspond with previous approaches. We will refer to this approach as *GP-AIC* in the rest of the paper and needs only the last two of the previous steps.

Both methods are compared in terms of Parkes error grid to evaluate the accuracy of the predictions for each time horizon.

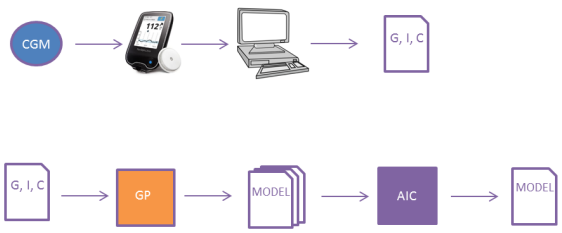

Figure 2: Flow diagram describing the creation process of models based on *GP-AIC*.

5 EXPERIMENTAL RESULTS

The mean, standard deviation and percentages of time where the patient has glucose levels below 70 [mg/dl], above 250 [mg/dl], and in the range [70, 180] [mg/dl] known as time in range are shown in Table 2. The patient has a high standard deviation and the percentage in range above 250 [mg/dl] is higher than the percentage in range below 70 [mg/dl]. This is normal in diabetic patients due to the disease and diabetes control.

Table 2: Features of glucose values where mean is the average glucose, std is the standard deviation, and T_G is the percentage of time that the glucose values are in each range (< 70 [*mg/dl*], > 250 [*mg/dl*] and [70, 180] [*mg/dl*] or time in range).

mean [mg/dl]	std [mg/dl]	$T\_G_{<70}$ [%]	$T\_G_{>250}$ [%]	$T\_G_{[70,180]}$ [%]
157.67	62.25	4.48	7.31	37.57

Table 3 shows the results obtained with the CHAID algorithm. The final depth of the tree is 2, the number of nodes is 33, and the number of terminal nodes is 23. The first predictor (variable) used in the construction of the tree is the day of the week and the second is the time slot. Significant differences are observed in the glucose profiles classified for each of the categories of the independent variables. In the first level groups are made by the day of the week. There are four different glycemic patterns; one for Tuesday, another for Thursday, another for Friday and one for Monday-Wednesday-Saturday-Sunday, where the lowest value of average glucose is for Friday (163.53±58.06 [mg/dl]) and the highest is for Tuesday (179.01±55.61 [mg/dl]). In the second level groups are made by time slots. The colors separate the size of the groups at this level. It should be noted that glucose values on Friday usually have two different behaviors, one for the time period [20:00h-04:00h] and another for the rest of the day.

Associations are also observed in the glucose profiles between the categories of the independent variables. Associations for the variable day of the week are formed by two, three and up to four categories. The same happens for the associations found in the variable time slot. The most common combinations for the variable day of the week are of one category. The least common ones are with four categories. The most common combination in the variable time slot is of one category. The least common are with three categories and four, although there are many more associations in time slots than in days of the week.

Next, we analyze the accuracy of the GP models using Parkes error grid under the assumption that the best model obtains a higher percentage of values in zone A+B, and a lower percentage of values in zones C, D, and E. This assumption is a consequence of the meaning of the zones (see section 3.4). Table 5 presents the results. GP models created with *GP-AIC* are represented as *GP*, and the average percentage of GP models created with *CHAID-GP-AIC* are represented as *Avg*. The rest of the acronyms represent GP models created with *CHAID-GP-AIC*.

In general, the accuracy of predictions is better for shorter time horizons and gradually gets worse as the time horizon increases from 30 to 240 minutes. Note that in most of the cases, the percentage of values in zone E is null (only 13 out of 128 cases are not null). This is preferable since the predictions being likely to have dangerous consequences in the treatment of patients are avoided. Remarkably, model $T01$ with time horizon of 30 minutes has an 100 % accuracy; all the predictions are in zone A+B.

In predictions with time horizons at 30, 60 and 90 minutes, $T01$ has the best results. For the rest of the time horizons (120, 150, 180, 210 and 240 minutes), $R0$ has the best results. Note that models created with glucose values classified in categories with fewer elements obtained the best results.

A comparison between *GP-AIC* and *CHAID-GP-AIC* is made in Figure 3 and Figure 4. The ratio of predictions in zone A+B for the different clusters of data are shown in Figure 3. It is observed that in 11 out of 15 cases, models created with *CHAID-GP-AIC* are more accurate (8 out of 15) or at least of equal accuracy (3 out of 15) than models created with *GP-AIC*. Moreover, in 4 cases ($MWSU01$, $T01$, $R0$ and $F05$) *CHAID-GP-AIC* predictions are the best for all time horizons. Figure 4 shows the ratio between the values obtained in zone A+B for the time horizons of models created with *CHAID-GP-AIC* and *GP-AIC*. For all time horizons, except 120 and 150 minutes, *CHAID-GP-AIC* models are more accurate, although a deeper study is necessary.

The execution times of the models are shown in Table 4. Note that the execution times are similar for all models except *GP*, where it is an order of magnitude higher than the others. Thus, the models created with *CHAID-GP-AIC* are faster than models created with *GP-AIC* (*GP*). This is logical because models with fewer observations are faster than models with more observations. All the algorithms are executed using an Intel (R) Core (TM) i7 5820k with 64 GB of RAM (DDR4) in Windows 7 64 bits at 3.30 GHz.

6 CONCLUSIONS AND FUTURE WORK

Significant differences (p-value < 0.05) and associations are observed between the glucose profiles classified using the independent variables day of the week and time slot. The accuracy of predictions with models created with GP is better for shorter time horizons and gradually gets worse as the time horizon increases from 30 to 240 minutes. In the majority of cases, the percentages of values in zone E are null, avoiding dangerous consequences

Table 3: Results obtained with the CHAID algorithm for the groupings obtained with different days and time slots. mean is the average glucose, std is the standard deviation, and T_G is the percentage of time that the glucose values are in each range (< 70 [*mg/dl*], > 250 [*mg/dl*] and [70, 180] [*mg/dl*] or time in range). The colors in the time slot represent the number of elements in the group: *blue* represents groupings of 1 element, *green* grouping of 2, *yellow* 3 and *red* 4.

mean [mg/dl]	std [mg/dl]	$T\_G_{<70}$ [%]	$T\_G_{>250}$ [%]	$T\_G_{[70,180]}$ [%]	day of the week	time slot				
170.62	61.36	2.22	10.45	57.56	MWSU	01		2	3	4	5
179.01	71.15	0.65	13.65	55.61	T	01		24		3	5
167.22	61.34	1.47	9.81	61.28	R	0	123			4	5
163.53	58.06	2.40	8.43	62.55	F	05		1234			

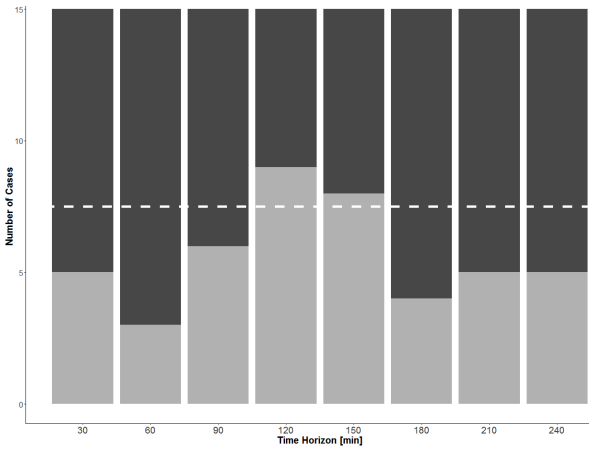

Figure 3: Ratio of models with better predictions (higher values in zone A+B) for the different clusters of data. Dark gray segments indicate that the best prediction is made by *CHAID-GP-AIC* while light gray means a better model from *GP-AIC*.

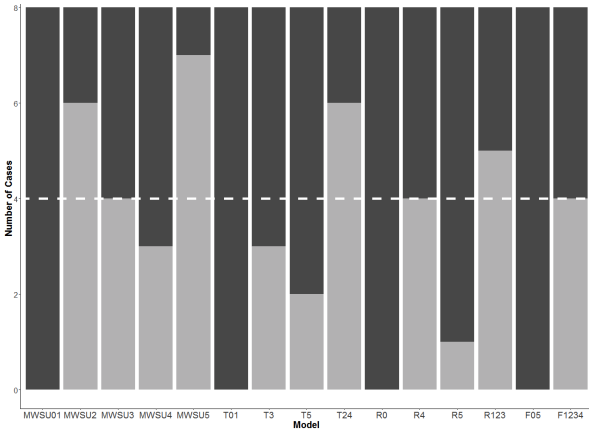

Figure 4: Ratio of models with better predictions (higher values in zone A+B) for the different time horizons. Dark gray segments indicate that the best prediction is made by *CHAID-GP-AIC* while light gray means a better model from *GP-AIC*.

Table 4: Average execution times by model.

model	GP	$MWSU$	T	R	F
time [sec]	391.86	67.91±16.77	36.05±3.26	37.04±6.80	50.48±14.07

in the treatment of patients. It is significant that there is a 100 % accuracy in one of the models where all predictions are in zone A+B. In general, using classified glucose values in models created with traditional GP, the accuracy of the prediction of the glucose values improves in comparison with models created with the original data-set.

Significant differences found in the classification process can be useful to correct and improve habits or therapies in patients, and obtain more accurate models through automatic learning techniques and artificial intelligence. The results obtained will facilitate the mathematical modeling of glucose, and can be used for the creation of an individualized classifier for each patient, that classifies the glucose profiles according to the variables day of the week and time slot. Using this classifier, the glucose values of the patient can be forecast knowing the day of the week and the time slot, obtaining more accurate models. During the workshop, we will present results for more patients.

ACKNOWLEDGMENTS

This work has been supported by: Fundación Eugenio Rodríguez Pascual 2019 grant (Desarrollo de un sistema para mejora del control de la diabetes mellitus basado en inteligencia artificial); The Spanish Ministerio de Economía y Competitividad grant TIN2014-54806-R; Madrid Regional Goverment grant B2017/BMD3773 (GenObIA-CM); Madrid Regional Goverment grant Y2018/NMT-4668 (Micro-Stress- MAP-CM); FEDER;

REFERENCES

[1] Hirotogu Akaike. 1998. *Information Theory and an Extension of the Maximum Likelihood Principle.* Springer New York, New York, NY, 199–213. https://doi.org/10.1007/978-1-4612-1694-0_15

[2] American Diabetes Association. 2016. Intensive Diabetes Treatment and Cardiovascular Outcomes in Type 1 Diabetes: The DCCT/EDIC Study 30-Year Follow-up. *Diabetes Care* (2016).

[3] Carlos Cervigon, J. Ignacio Hidalgo, Marta Botella, and Rafael-J Villanueva. 2017. A genetic algorithm approach to customizing a glucose model based on usual therapeutic parameters. *Progress in Artificial Intelligence* 6, 3 (2017), 255–261.

[4] W. Clarke, D. Cox, L. Gonder-Frederick, W. Carter, and S. Pohl. 1987. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. *Diabetes Care* 10, 5 (1987), 622–628.

[5] M. Hurvich Clifford and Chih Ling Tsai. 1989. Regression and time series model selection in small samples. *Biometrika* 76, 2 (1989), 297–307. https://doi.org/10.1093/biomet/76.2.297

Table 5: Predictions (in percentage) in the evaluation phase obtained for the different zones with Parkes error grid for all time horizons. Color *blue* in zone A+B is used to represents the percentage of models created with *GP-AIC*, *green* values indicates that *CHAID-GP-AIC* is better while *red* represents worst models. Bold text highlights the best value.

	t+30						t+60					
model	A+B	A	B	C	D	E	A+B	A	B	C	D	E
GP	95.14	65.09	30.05	4.22	0.64	0.00	94.70	59.28	35.42	4.79	0.51	0.00
*MWSU*01	96.83	76.05	20.78	2.64	0.53	0.00	98.96	82.53	16.43	0.97	0.07	0.00
*MWSU*2	98.96	83.45	15.51	1.05	0.00	0.00	95.23	65.40	29.83	4.66	0.11	0.00
*MWSU*3	99.06	77.18	21.88	0.94	0.00	0.00	95.83	59.50	36.33	4.09	0.08	0.00
*MWSU*4	92.56	54.73	37.83	6.79	0.61	0.03	92.24	52.91	39.33	7.11	0.64	0.00
*MWSU*5	93.26	56.05	37.21	6.52	0.22	0.00	90.03	48.62	41.41	8.98	0.99	0.00
*T*01	**100.00**	93.42	6.58	0.00	0.00	0.00	**99.73**	84.97	14.76	0.27	0.00	0.00
*T*3	99.65	81.98	17.67	0.35	0.00	0.00	96.16	59.42	36.74	3.84	0.00	0.00
*T*5	90.22	50.22	40.00	8.20	1.57	0.00	91.01	52.81	38.20	7.75	1.24	0.00
*T*24	93.17	61.81	31.36	6.78	0.06	0.00	96.89	65.82	31.07	3.11	0.00	0.00
*R*0	99.65	88.12	11.53	0.35	0.00	0.00	99.53	70.82	28.71	0.47	0.00	0.00
*R*4	95.06	72.26	22.80	4.19	0.75	0.00	97.42	70.00	27.42	2.47	0.11	0.00
*R*5	96.04	58.13	37.91	3.19	0.77	0.00	97.47	68.46	29.01	2.53	0.00	0.00
*R*123	95.47	66.98	28.49	4.09	0.43	0.00	98.14	77.54	20.60	1.55	0.30	0.00
*F*05	99.38	86.07	13.31	0.62	0.00	0.00	97.30	73.26	24.04	2.70	0.00	0.00
*F*1234	97.69	79.44	18.25	2.09	0.23	0.00	95.59	61.72	33.87	3.67	0.73	0.00
Avg	96.47	72.39	24.07	3.19	0.34	0.00	96.10	66.25	29.85	3.61	0.28	0.00
	t+90						t+120					
GP	94.44	60.23	34.21	5.05	0.50	0.01	94.62	55.47	39.15	5.11	0.28	0.00
*MWSU*01	97.10	68.64	28.46	2.70	0.20	0.00	96.43	59.50	36.93	3.14	0.43	0.00
*MWSU*2	92.54	52.68	39.86	6.84	0.62	0.00	88.62	39.83	48.79	10.71	0.68	0.00
*MWSU*3	89.97	44.89	45.08	9.53	0.50	0.00	87.05	38.51	48.54	11.44	1.52	0.00
*MWSU*4	91.93	54.68	37.25	6.84	1.20	0.03	95.40	64.22	31.18	4.33	0.27	0.00
*MWSU*5	95.06	58.90	36.16	4.59	0.36	0.00	94.17	50.55	43.62	5.69	0.11	0.03
*T*01	**98.98**	77.11	21.87	0.96	0.05	0.00	97.71	71.93	25.78	2.25	0.05	0.00
*T*3	95.24	47.33	47.91	4.77	0.00	0.00	93.49	43.49	50.00	6.51	0.00	0.00
*T*5	98.20	74.27	23.93	1.57	0.22	0.00	96.96	56.85	40.11	2.92	0.11	0.00
*T*24	94.35	60.11	34.24	5.54	0.11	0.00	92.38	54.24	38.14	7.40	0.23	0.00
*R*0	98.94	70.00	28.94	1.06	0.00	0.00	**97.76**	67.88	29.88	2.24	0.00	0.00
*R*4	94.09	54.09	40.00	5.91	0.00	0.00	93.01	48.28	44.73	6.88	0.11	0.00
*R*5	95.17	52.86	42.31	4.84	0.00	0.00	93.29	47.36	45.93	6.15	0.55	0.00
*R*123	96.16	57.67	38.49	3.53	0.30	0.00	93.50	48.84	44.66	6.16	0.34	0.00
*F*05	97.64	64.16	33.48	2.36	0.00	0.00	95.28	57.13	38.15	4.04	0.67	0.00
*F*1234	92.77	46.84	45.93	6.05	1.19	0.00	88.67	39.86	48.81	9.01	1.81	0.51
Avg	95.21	58.95	36.26	4.47	0.32	0.00	93.58	52.56	41.02	5.92	0.46	0.04
	t+150						t+180					
GP	92.54	47.75	44.79	6.56	0.89	0.02	91.02	43.83	47.19	8.11	0.87	0.00
*MWSU*01	95.07	53.39	41.68	4.14	0.80	0.00	92.64	48.65	43.99	6.50	0.86	0.00
*MWSU*2	81.18	39.15	42.03	17.46	1.27	0.08	87.21	39.44	47.77	11.95	0.85	0.00
*MWSU*3	87.65	38.70	48.95	10.64	1.71	0.00	91.55	43.70	47.85	7.49	0.97	0.00
*MWSU*4	93.93	54.95	38.98	5.61	0.45	0.00	93.74	51.55	42.19	6.02	0.24	0.00
*MWSU*5	91.76	45.77	45.99	7.96	0.25	0.03	88.73	39.53	49.20	10.75	0.52	0.00
*T*01	96.42	66.63	29.79	3.42	0.16	0.00	93.48	64.01	29.47	6.26	0.27	0.00
*T*3	92.33	42.33	50.00	7.67	0.00	0.00	91.16	41.74	49.42	8.84	0.00	0.00
*T*5	93.03	49.10	43.93	6.97	0.00	0.00	92.81	48.09	44.72	6.74	0.45	0.00
*T*24	89.49	50.73	38.76	10.45	0.06	0.00	90.06	45.03	45.03	9.89	0.06	0.00
*R*0	**98.12**	65.65	32.47	1.88	0.00	0.00	**97.89**	62.24	35.65	2.12	0.00	0.00
*R*4	92.15	49.68	42.47	7.85	0.00	0.00	92.04	49.46	42.58	7.31	0.65	0.00
*R*5	94.29	44.40	49.89	4.84	0.88	0.00	91.76	42.31	49.45	4.62	3.63	0.00
*R*123	82.24	35.52	46.72	14.74	2.76	0.26	89.35	37.54	51.81	9.48	1.08	0.09
*F*05	95.56	50.00	45.56	3.60	0.84	0.00	96.46	45.39	51.07	2.92	0.62	0.00
*F*1234	91.44	39.24	52.20	6.95	1.61	0.00	91.58	39.04	52.54	7.57	0.85	0.00
Avg	91.64	48.35	43.29	7.61	0.72	0.02	92.03	46.51	45.52	7.23	0.74	0.01
	t+210						t+240					
GP	90.21	41.16	49.05	8.69	1.10	0.00	89.87	39.85	50.02	9.01	1.12	0.00
*MWSU*01	92.54	45.09	47.45	6.16	1.30	0.00	90.35	41.65	48.70	7.96	1.69	0.00
*MWSU*2	88.30	41.44	46.86	10.68	1.02	0.00	88.08	40.42	47.66	10.48	1.44	0.00
*MWSU*3	89.78	44.03	45.75	9.36	0.86	0.00	90.72	45.33	45.39	8.45	0.83	0.00
*MWSU*4	91.85	46.58	45.27	7.49	0.64	0.03	90.29	43.85	46.44	8.80	0.91	0.00
*MWSU*5	89.40	39.34	50.06	10.08	0.52	0.00	89.42	43.59	45.83	10.00	0.58	0.00
*T*01	91.88	59.79	32.09	7.70	0.43	0.00	90.85	53.26	37.59	8.66	0.48	0.00
*T*3	92.56	43.14	49.42	6.74	0.70	0.00	88.37	43.37	45.00	11.63	0.00	0.00
*T*5	94.83	45.17	49.66	4.27	0.90	0.00	92.02	45.39	46.63	7.30	0.67	0.00
*T*24	90.96	40.90	50.06	8.93	0.11	0.00	89.66	43.28	46.38	10.23	0.11	0.00
*R*0	**98.47**	57.88	40.59	1.53	0.00	0.00	**98.71**	56.12	42.59	1.29	0.00	0.00
*R*4	93.12	51.51	41.61	6.67	0.22	0.00	95.70	52.26	43.44	4.30	0.00	0.00
*R*5	94.18	41.21	52.97	3.74	2.09	0.00	95.39	43.30	52.09	4.62	0.00	0.00
*R*123	75.99	27.89	48.10	20.99	2.97	0.04	89.70	39.66	50.04	9.40	0.60	0.30
*F*05	95.51	40.73	54.78	3.48	1.01	0.00	95.51	40.34	55.17	3.54	0.96	0.00
*F*1234	89.63	39.01	50.62	8.28	2.09	0.00	90.03	40.48	49.55	8.84	1.13	0.00
Avg	91.27	44.25	47.02	7.74	0.99	0.00	91.65	44.82	46.83	7.70	0.63	0.02

[6] Ivanoe De Falco, Antonio Della Cioppa, Tomas Koutny, Michal Krcma, Umberto Scafuri, and Ernesto Tarantino. 2018. Genetic Programming-based induction of a glucose-dynamics model for telemedicine. *Journal of Network and Computer Applications* 119 (2018), 1–13.

[7] OrganizaciÃĆÃşn Mundial de la Salud. 2016. Informe Mundial Sobre la Diabetes. *World Health Organization 2016* (2016).

[8] Ya-Xiong Tao (Eds.). 2014. *Glucose Homeostatis and the Pathogenesis of Diabetes Mellitus* (1 ed.). Academic Press.

[9] F.H. El-Khatib, S.J. Russell, D.M. Nathan, R.G. Sutherlin, and E.R. Damiano. 2010. A bihormonal closed-loop artificial pancreas for type 1 diabetes. *Sci Transl Med.* 2 (27) (2010).

[10] J. Ignacio Hidalgo, Esther Maqueda, Jose L. Risco-Martan, Alfredo Cuesta-Infante, J. Manuel Colmenar, and Javier Nobel. 2014. glUCModel: A monitoring and modeling system for chronic diseases applied to diabetes. *Journal of Biomedical Informatics* 48 (2014), 183 – 192.

[11] Jose Ignacio Hidalgo, Jose Manuel Velasco, Juan Lanchares, Sergio Contador, and Oscar Garnica. 2018. An analysis of solutions based on Genetic Programming to solve problems of symbolic regression of data from continuous glucose monitoring. *XVIII Conferencia AEPIA* (2018), 764–771. arXiv:https://www.tandfonline.com/doi/pdf/10.1080/01621459.1972.10481279 https://sci2s.ugr.es/caepia18/proceedings/proceedings.php

[12] Roman Hovorka, Janet M Allen, Daniela Elleri, Ludovic J Chassin, Julie Harris, Dongyuan Xing, Craig Kollman, Tomas Hovorka, Anne Mette F Larsen, Marianna Nodale, Alessandra De Palma, Malgorzata E Wilinska, Carlo L Acerini, and David B Dunger. 2010. Manual closed-loop insulin delivery in children and adolescents with type 1 diabetes: a phase 2 randomised crossover trial. *The Lancet* 375 (2010), 743 – 751.

[13] George Kaplan, Debbie E Goldberg, Susan Everson-Rose, Richard D Cohen, Riitta Salonen, Jaakko Tuomilehto, and Jukka Salonen. 1996. Perceived Health Status and Morbidity and Mortality: Evidence from the Kuopio Ischaemic Heart Disease Risk Factor Study. 25 (05 1996).

[14] G. V. Kass. 1980. An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 29, 2 (1980), 119–127. http://www.jstor.org/stable/2986296

[15] Christina-Maria Kastorini, George Papadakis, Haralampos J. Milionis, Kallirroi Kalantzi, Paolo-Emilio Puddu, Vassilios Nikolaou, Konstantinos N. Vemmos, John A. Goudevenos, and Demosthenes B. Panagiotakos. 2013. Comparative analysis of a-priori and a-posteriori dietary patterns using state-of-the-art classification algorithms: A case/case-control study. *Artificial Intelligence in Medicine* 59, 3 (2013), 175 – 183.

[16] Maarten Keijzer. 2003. Improving Symbolic Regression with Interval Arithmetic and Linear Scaling. (2003), 70–82.

[17] B. Kovatchev, C. Cobelli, E. Renard, S. Anderson, M. Breton, S. Patek, W. Clarke, D. Bruttomesso, A. Maran, S. Costa, A. Avogaro, C. Dalla Man, A. Facchinetti, L. Magni, G. De Nicolao, J. Place, and A. Farret. 2010. Multinational study of subcutaneous model-predictive closed loop control in type 1 diabetes mellitus: Summary of the results. *Diabetes Sci Technol* 4 (2010), 1374–1381.

[18] John R. Koza. 1992. *Genetic Programming: On the Programming of Computers by Means of Natural Selection.* MIT Press, Cambridge, MA, USA.

[19] Sean Luke. 2000. Two fast tree-creation algorithms for genetic programming. *IEEE Trans. Evolutionary Computation* 4 (2000), 274–283.

[20] Lalo Magni, Marco Forgione, Chiara Toffanin, Chiara Dalla Man, Boris Kovatchev, Giuseppe De Nicolao, and Claudio Cobelli. 2009. Run-to-Run Tuning of Model Predictive Control for Type 1 Diabetes Subjects: In Silico Trial. *Journal of Diabetes Science and Technology* 3, 5 (2009), 1091–1098.

[21] Lucy Mays. 2015. Diabetes Mellitus Standards of Care. *Nursing Clinics of North America* 50, 4 (2015), 703 – 711. Pathophysiology and Care Protocols for Nursing Management.

[22] Mirko Messori, Chiara Toffanin, Simone Del Favero, Giuseppe De Nicolao, Claudio Cobelli, and Lalo Magni. 2016. Model individualization for artificial pancreas. *Computer Methods and Programs in Biomedicine* (2016), –.

[23] J. L. Parkes, S. L. Slatin, S. Pardo, and B. H. Ginsberg. 2000. A new consensus error grid to evaluate the clinical significance of inaccuracies in the measurement of blood glucose. *Diabetes Care* 23, 8 (2000), 1143–1148. https://doi.org/10.2337/diacare.23.8.1143 arXiv:http://care.diabetesjournals.org/content/23/8/1143.full.pdf

[24] Andreas Pfützne, David C. Klonoff, Scott Pardo, and Joan L. Parkes. 2013. Technical Aspects of the Parkes Error Grid. *Journal of Diabetes Science and Technology* 7, 5 (2013), 1275–1281. https://doi.org/10.1177/193229681300700517 arXiv:https://doi.org/10.1177/193229681300700517 PMID: 24124954.

[25] Thierry Prud'Homme, Alain Bock, Gregory Francois, and Denis Gillet. 2011. Preclinically assessed optimal control of postprandial glucose excursions for type 1 patients with diabetes. In *Automation Science and Engineering (CASE), 2011 IEEE Conference on*. IEEE, 702–707.

[26] G.M. Steil, K. Rebrin, F. Hariri, S. Jinagonda, S. Tadros, C. Darwin, and M.F. Saad. 2005. Interstitial fluid glucose dynamics during insulin-induced hypoglycaemia. *Diabetologia* 48, 9 (2005), 1833–1840.

[27] Jose Manuel Velasco, Oscar Garnica, Juan Lanchares, Marta Botella, and J. Ignacio Hidalgo. 2018. Combining data augmentation, EDAs and grammatical evolution for blood glucose forecasting. *Memetic Computing* (22 Jun 2018). https://doi.org/10.1007/s12293-018-0265-6

[28] S. Wagner and M. Affenzeller. 2005. HeuristicLab: A Generic and Extensible Optimization Environment. (2005), 538–541.