Evolving Classification Rules for Predicting Hypoglycemia Events

Marina De La Cruz
Universidad Complutense de Madrid
Madrid, Spain
marcru06@ucm.es

Carlos Cervigón
Universidad Complutense de Madrid
Madrid, Spain
ccervigon@fdi.ucm.es

Jorge Alvarado
Universidad Complutense de Madrid
Madrid, Spain
jorgal@ucm.es

Marta Botella-Serrano
Hospital Universitario Príncipe de Asturias
Alcalá de Henares, Spain
mata.botella@madridsalud.org

J.Ignacio Hidalgo
Universidad Complutense de Madrid
Instituto de Tecnología del Conocimiento
Madrid, Spain
hidalgo@ucm.es

***Abstract*—People with diabetes have to properly manage their blood glucose levels in order to avoid acute complications. This is a difficult task and an accurate and timely prediction may be of vital importance, especially of extreme values. Perhaps one of the main concerns of people with diabetes is to suffer a hypoglycemia (low value) event and moreover, that the event will be prolonged in time. It is crucial to predict events of hyperglycemia (high value) and hypoglycemia that may cause health damages in the short term and potential permanent damages in the long term. The aim of this paper is to describe our research on predicting hypoglycemia events using Dynamic structured Grammatical Evolution. Our proposal gives white box models induced by a grammar based on if-then-else conditions. We trained and tested our system with real data collected from 5 different diabetic patients, producing 30-minute predictions with encouraging results.**

***Index Terms*—Diabetes, Hypoglycemia prediction, Rule System, Structured Grammatical Evolution**

I. Introduction

The progression of *Diabetes Mellitus* (DM) in recent years has made it one of the most relevant diseases of the 21st century. According to World Health Organization (WHO) estimations, diabetes will be among the leading causes of death worldwide by 2030. Type I diabetes is an autoimmune disease that causes the destruction of the insulin-producing cells (beta cells) of the pancreas. A healthy pancreas is responsible for regulating blood glucose levels by producing insulin. Thanks to insulin, the body's cells can absorb glucose from the bloodstream. In the case of a person with diabetes, the absence of insulin prevents the assimilation of glucose, causing an increase in the levels of glucose in the blood stream. As a result, the person needs to inject insulin to be able to maintain these values in a healthy range, however there are a lot of external factors that affect the variability of the blood glucose levels. As a result, is very common for a people with diabetes (PwD) to suffer hypoglycemia events, not only as a result of over injecting insulin, but also due to other facts such as doing exercise.

Exercising is also very important for PWD and it can specifically help with weight management and insulin sensitivity. A good control of the body weight and a stable sensibility to insulin value reduce the amount of insulin a diabetic patient needs to inject per meal and, consequently the reaction on their blood sugar levels, leading to more stable values. However, as a result of the increasing in insulin sensitivity, and the fact that the muscles need sugar while exercising, it can result in a state of hypoglycemia or low blood sugar for the patient that can be dangerous if it is not addressed properly, and might lead to a loss of consciousness.

The symptoms of hypoglycemia vary depending on the person but usually encompass: shaking, sweating, hunger, fast or irregular heartbeat, numbness in extremities, etc... These symptoms can be easily confused whilst exercising and a person with diabetes might not be aware that she/he is suffering an hypoglycemia event. Another thing to take into account is that the body can get used to being in a state of hypoglycemia, this means that the more hypoglycemic episodes a patient suffers the less they will feel the initial symptoms until their blood sugar drops lower and lower, which is not ideal when trying to recognize the state if the patient is not actively looking at their current blood sugar levels.

The motivation for this research lies in the need for reliable predictive models of blood glucose low levels for PwD. The prediction of a future hypoglycemic episode can lead the patient to take measures to remedy it before it happens and to take better decisions in the future to avoid acute complications. We train classification models by means of Structured Grammatical Evolution (SGE). We present a method to obtain *White-box* models composed by a set of *if-then-else* conditions. The conditions to evaluate include information of a set of physiological variables during the last two hours previous to the prediction time. In particular, we include the values of glucose levels measured by a Continuous Glucose Monitoring System (CGM), and heart rate information, number of steps and calories burned (these last three obtained by wearing a smartwatch). As a result, we obtain models that predict a class which represents the future state of the person (hypoglycemia-non hypoglycemia) in the short term (30 minutes). We inves-

tigate the performance of both static SGE and Dynamic SGE. Experimental results with a set of data of real people from a Hospital in Spain are encouraging, and shows that GE can produce good rule-based models.

The rest of the paper is organized as follows. Section II describes the problem of predicting hypoglycemia in people with diabetes and revises other works in the literature. Section III explains our approach, the workflow and the main techniques used in this research, whereas Section IV describes the experimental setup and the forecasting results that these techniques have obtained. Conclusions and future work are exposed in Section V.

II. Hypoglycemia prediction

Having identified that one of the most important aspects to control in patients with diabetes is hypoglycemia, let us start presenting the problem and analyzing the related work. We already mentioned that hypoglycemia is understood as a state of low blood sugar that can be due to multiple factors, such as physical exercise, an incorrect diet or the injection of excessive doses of insulin. In this work we consider four important variables related to these factors: glucose, heart rate, steps and calories burned. The latter two are used because it has been shown that physical exercise is highly recommended in PwD and can contribute to stabilizing blood glucose values, but it could also be dangerous and lead to a state of hypoglycemia because of the demand for glucose at a muscular level.

It is therefore of vital importance to detect or predict these states of hypoglycemia. Previous hypoglycemia data may be the best predictors of new hypoglycemia events and, if we can correctly predict a future episode of hypoglycemia, persons can act to remedy it before it happens.

Hypoglycemia is identified when the blood glucose level falls below 70 mg/dL [1], but as a reference of possible hypoglycemia states, the American Diabetes Association (ADA) work-group has proposed a classification with five different grades, from Severe Hypoglycemia to Pseudo-hypoglycemia [2], [3].

In general, the problem addressed is the prediction of the blood glucose level value over a time horizon using historical information of physiological variables. Some approaches use values of the previous 90 minutes, 120 minutes, in the case of postprandial blood glucose, and other approaches work with periods up to 24 hours. There are also approaches that work with models adapted to individuals and general models constructed with information of several persons.

For our particular study we will use historical information of blood glucose, heart rate, steps and calories burned data (for each of these variables we will use samples taken every 5 minutes). Hypoglycemia and glucose prediction is a widely discussed topic. Many studies have proposed methods to characterize glycemic variability, which as mentioned above is considered an important risk factor in patients with diabetes. In [4] they have developed a new metric, the glycemic variability percentage (GVP), to assess glycemic variability. In [5] they propose a new metric called coefficient of gradient variability (CVG), which allows characterizing glycemic variability and the risk of hypoglycemia.

A description of the different techniques that are commonly applied to make predictions can be found in [6]. Several papers have used Genetic Programming and variants ([7]–[13]), Grammatical evolution (GE) has achieved good results in time series prediction problems [14], [15]. Probabilistic fitting in evolutionary optimized models has also been used successfully [16]. In [17], authors applied a Particle Swarm Optimization algorithm (PSO) to calibrate a model of glucose dynamics to predict short-term glucose level.

Machine learning (ML) is an area with great impact and application in the field of medicine [18], [19]. It has been shown that ML techniques applied to prediction has produced very encouraging results. With specific reference to the prediction of hypoglycemia states, a review of the techniques that use ML is given in [20], where the different techniques applied to hypoglycemia prediction are summarized in detail, considering also the type of data used, predictions horizons, the age of the population, the type of treatment and the gender of the patient.

For instance artificial Neural Networks [21] [22] [23] [24] [25], Random Forest [26] [27] [28], Kernel based SVM [29] [30], Logistic Regression [31] [32], etc.) .

Machine learning techniques often use black box representations to make sense of the models, without knowing the level of detail. They are very complicated to understand at a detailed level, so only the inputs and outputs of the system are analyzed and studied without taking into account its inner workings. While machine learning techniques use black box representations to make sense of the models, without knowing the level of detail, our approach uses white box models in which the deductive approach is mixed with all the theoretical part that governs the system internally. We use ”explainable” models in which the internal model can be calibrated and adjusted to do data validation, prediction, etc.

Other work has shown that Structured Grammatical Evolution (SGE) perform better than traditional Grammatical Evolution (GE) [14]. In [33], they use Structured Grammatical Evolution for glucose prediction, but the approach is different from ours: different variables are used (past glucose values, insulin injections, and the amount of carbohydrate ingested by a patient), different grammars and different metrics are also used to optimize the solutions of the evolutionary algorithm. In [34], a comparison is made between evolutionary grammars (GE), Contex-Free-Grammar Genetic Programming and Structured (CFG-GP) and Grammatical Evolution (SGE), and they showed that SGE has a better performance than GE on several problems.

III. Methodology

Although there can be different degrees of hypoglycemia, in this paper we address the problem of identifying hypoglycemic events over a set period of time as if it were a single class between two possible classifications: Hypoglycemia and Non-Hypoglycemia. We created a classification system that can identify a hypoglycemic state when glucose values are equal

or less than 70mg/dL mg. For this classification, we use an evolutionary algorithm with Static structured Grammatical Evolution and Dynamic structured Grammatical Evolution.

A. Structured Grammatical Evolution

Grammatical Evolution (GE) [35] is a variation of Genetic Programming (GP) that uses a grammar to generate the phenotype from the genotype of the individuals, the genotype is made up of a list of numbers each one of which will generate a terminal o non-terminal symbol of the grammar forming the phenotype until there is no more possible expansions and it has been completely created. Although GE has been successfully applied to different problems, it presents some drawbacks derived from the process of decoding the solutions. GE has two main issues: (i) since each allele depends on the previous one, a small change on it can completely change the final resulting phenotype, and, (ii) since we use a modulus operation to determine the next rule, we could potentially change an allele but the generated phenotype might be equivalent. On the other hand, the ability to change the problem by using the grammar rather than editing the code is a great feature of GE.

In order to partially solve this concerns, Lourenço et al. proposed Structured Grammatical Evolution [34]. In comparison to Grammatical Evolution where individuals are made up of a list of numbers of a set length and each number is used to generate the next symbol of the phenotype, in Structured Grammatical Evolution, individuals are made up of a list of lists.

Each internal list represents a non-terminal of the grammar and the numbers contained are the possible expansions of this rule, therefore each internal lists length can be at most the maximum number of expansions for a certain rule and each one of the numbers contained in the list will represent what production we take for a rule, the numbers can go from 0 to C_{n-1}, being C_{n-1} the number of derivation options for a non-terminal.

1) Static Structured Grammatical Evolution: In this case each internal list is generated completely to its maximum length, the maximum number of expansions for each non-terminal element of the grammar, even if those expansions are not being used to decode the individual. This forces us to eliminate recursion from the grammar file, since recursion means that there could exist an infinite number of expansions for the recursive rule. We have performed this by setting a maximum recursive depth, defined by the user, and transforming the recursive rule into a set of rules that mimic the recursion up to the set depth.

```
<expr> ::= <expr> <op> <expr>
```

As an example, for depth 4 this rule will be transformed into:

```
<expr> ::= <expr1> <op> <expr1>  |
           <term> <op> <term>
<expr1> ::= <expr2> <op> <expr2> |
            <term> <op> <term>
<expr2> ::= <expr3> <op> <expr3> |
            <term> <op> <term>
<expr3> ::= <term> <op> <term>
```

2) Dynamic Structured Grammatical Evolution: For the dynamic version the lists are generated dynamically, that is, when the individuals are generated we generate the lists up to what is needed to complete the individual (all the generated individuals must be valid from creation). After the application of crossover and mutation operators, the individual might stop being valid, in which case when we generate the phenotype. We will add the needed derivation numbers to the lists to complete the individual, transforming it in valid. In case of recursion we could potentially create an individual that never completes the process of decoding: To avoid this from happening and avoiding obtaining individuals that are extremely big, we have added a maximum tree depth to limit the lenght of individuals. If the depth of the tree becomes higuer than the number set by the use,r the algorithm only generates terminal expansions and not recursive ones.

B. Grammar

In Structured GE algorithms, same as in GE, the phenotype of the individuals is generated using a grammar that determines what number of the genotype refers to what expression that forms the final model. The grammar we have chosen to generate the expressions is show in fig. 1. The expressions will be composed of a set of if-else statement that can either return class 0, hypoglycemia, or class 1, not hypoglycemia, the condition of the if statement can be composed of up to 3 individual conditions fused together by using 'and' or 'or' statements, these are made up of the input variables and generated numbers.

- Input Variables
  The input variables are those that appear in fig 1. as 'getVariable(x,k)' with $x \in (0, 100)$ and have been measured every 5 minutes.
  - Glucose values of the two hours before the time of prediction t ($x \in (0, 25)$).
  - Heart rate values of the patient of the two hours before time of prediction t ($x \in (25, 50)$).
  - Amount of steps performed by the patient of the two hours before time of prediction t ($x \in (50, 75)$).
  - Amount of calories burned by the patient in the two hours before time of prediction t ($x \in (75, 100)$).
- Output Variable
  - Expected class at 30 minutes after the time of prediction.

As can be observed in fig 1. we have not taken into account every single possible input to generate the models, we have only taken values that are less correlated from one another from each one of the input types, however, the available inputs that could be used by the grammar are the ones described above.

```
<type> ::= if( <binexpr> ){ result=0; }else{ result=1; }
<binexpr> ::= <relop>( <expr> , <mix> ) | ( <relop>( <expr> , <mix> ) <binop> <binexpr2> )
<binexpr2> ::= <relop>( <expr> , <mix> ) | ( <relop>( <expr> , <mix> ) <binop> <binexpr3> )
<binexpr3> ::= <relop>( <expr> , <mix> )
<binop> ::= && | ||
<expr> ::= <term> <op> <term> | ( <term> <op> <term> ) | <expr> <op> <expr> | ( <expr> <op> <expr> ) |
        ( <number> <op> <expr> )  | ( <number> <op> <term> ) | ( <expr> <op> <number> )  | ( <term> <op> <number> ) |
        <number> <op> <expr> | <number> <op> <term> | <expr> <op> <number> | <term> <op> <number>
<op> ::= + | - | * | /
<term> ::= getVariable(1,k) | getVariable(6,k) | getVariable(11,k) | getVariable(16,k) | getVariable(25,k) | |
        getVariable(26,k) | getVariable(38,k) | getVariable(44,k) | getVariable(50,k) | getVariable(51,k) |
        getVariable(55,k) | getVariable(59,k) | getVariable(63,k) | getVariable(67,k) | getVariable(71,k) |
        getVariable(75,k) | getVariable(76,k) | getVariable(80,k) | getVariable(84,k) | getVariable(88,k) |
        getVariable(92,k) | getVariable(96,k)
<digit> ::= 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0
<mix> ::= <number> | <expr>
<number> ::= <digit>.<digit><digit><digit> | <digit><digit>.<digit><digit><digit> |
                <digit><digit><digit>.<digit><digit><digit>
<relop> ::= lessThan | biggerThan | lessEqualThan | biggerEqualThan | Equal
```

Fig. 1. Grammar file.

C. Fitness Function

As with any evolutionary algorithm, we need a fitness function to determine how good an individual solution is: As we are dealing with a a classification problem we have chosen to use weighted accuracy.

$$\text{WA} = 0.5 * Accuracy + 0.5 * F_{measure} \tag{1}$$

$$F_{measure} = \frac{2 * Recall * Precision}{Recall + Precision} \tag{2}$$

where $Precision$ corresponds to the fraction of elements that we are classifying correctly in the Hypoglycemia class from the total amount we have classified in this class and $Recall$ to the fraction of elements that we are classifying correctly from the total amount of elements in the Hypoglycemia class.
The function will be executed for each data point, since we are minimizing the evolutionary algorithm and will use $1 - WA$.

IV. Experimental Results

A. Experimental set-up

To generate the models we have first divided our data into two subsets: training and test (70% and 30% respectively), after, we balanced the training model so that the amount of data for each class is approximately 50%, this must be performed so that the Fitness Function does not give more importance to non-hypoglycemic values as they are the most common. Balancing the training data helps to make the results less dependent on the partitioning of the data. The tests are performed without data balancing. We use this process as an alternative to other methods such as cross validation, which will be explored in the future.
As a result, we obtain a data file for each patient on which to execute the algorithm.

For the evolutionary algorithm we have set its parameters as such:

- Static replacement policy, the worst 2 individuals of the population
- 200 population size
- 600 generations
- Tournament selection operator with pressure 2
- Uniform crossover operator
  - 75% probability
  - 25% interchange probability per gene (internal lists)
- Basic mutation of all alleles
  - 15% probability

For each type of Structured Grammatical Evolution we define their respective depths

- Static: 4 recursive depth
- Dynamic: 7 maximum tree depth

These depth values have been elected to be approximately equivalent, so that the resulting if-else statements we obtain as results have a similar length.

We will execute 30 runs of the algorithm and obtain 30 models, the tests will be performed on the one with the lowest fitness (best model during training).

B. Data

The data that we have used to train and test our algorithm come from a set of persons after signing a informed consent. The persons are patients of the Hospital Universitario Príncipe de Asturias, in Alcalá de Henares, in the Region of Madrid (Spain). The main features of the patients appear in table I. A total of 11 people contribute, 7 female and 4 male, whose ages range goes from 20 to 56 years old and that follow 2 different treatments. 3 patients treat their blood sugar levels through injecting multiple doses of insulin, and the other 8 wore a continuous subcutaneal infuser of insulin. Their HBA1c value ranges from 6.4 to 8.5, this value reflects their average blood-glucose level for the last 3 months approximately, and can be a reflection of how good their control has been on average, the

ID	Gender	Age	IMC	HBA1c	Treatment	Years DMT1
HUPA001	F	56.3	22.76	8.2	ISCI	15.46
HUPA002	M	48.6	23.82	7.1	ISCI	36.47
HUPA003	F	43.4	18.72	7.3	ISCI	12.45
HUPA004	M	41.2	27.16	7.8	ISCI	8.5
HUPA005	F	20.9	22.57	6.9	ISCI	39.5
HUPA007	M	37.6	30.64	6.6	ISCI	10.1
HUPA011	F	35.0	23.92	7.8	ISCI	27.3
HUPA014	F	50.0	25.39	8.5	MDI	12.9
HUPA015	F	43.1	22.33	6.4	MDI	11.2
HUPA016	F	29.9	26.33	6.5	ISCI	20.1
HUPA027	M	26.4	22.21	7.0	MDI	23.7

TABLE I

CHARACTERISTICS OF THE PARTICIPANTS: ID, GENDER (M=MALE; F=FEMALE), AGE, BMI, HBA1C, TREATMENT (MDI: MULTIPLES DOSES OF INSULIN; ISCI. INFUSION SUBCUTANEAL CONTINUOUS OF INSULIN)YEARS OF EVOLUTION OF DMT1.

recommended value is for it to be around 7 or less. However, this value should not be taken as a reflection of a good control as it is only an average and does not show the nuances of the day to day values, which is what we are investigating in this study.

As stated before, we separate the data into to different files, one for training the models and the other for testing the results. A clear difference in the volume of hypoglycemic data versus non-hypoglycemic data has been observed in most of the datasets. This is an inevitable reality, since most of the patients should try to have less that 4% of their data values falling in hypoglycemia, as a recommendation for a healthy management of the glucose values. As a result, for some of the patients we have a very small amount of values on which train an individual model and for this reason we have tried (and succeeded) to generate a general model that includes the data from all the patients and observe how this model differs from the individual one when testing on the individual patient data.

Later, to expand on this idea, we have also generated a general model excluding a couple of patients (HUPA001, HUPA002 and HUPA003) and tested this one with all the data from the patients that were not included in the training phase.

C. Results

All of the results shown have been performed over a 30-minute prediction horizon, an example of a possible model we might obtain is:

```
if((biggerThan((101.257-getVariable(1,k)),
        20.272/getVariable(16,k)*76.250)
    || biggerThan(getVariable(67,k)/
       getVariable(38,k) , 21.816))){
     result=0;} else{result=1;}
```

Where getVariable(1,k) corresponds to Glucose(t), getVariable(16,k) to Glucose (t-50), getVariable(67,k) to Steps in (t-45) and getVariable(38,k) to HeartRate in (t-65).

In table II we show the Mean, Median, Standard deviation of the fitness and best Individual for the 30 runs for each patient, for each algorithm used. On average, we can observe that Dynamic Structured Grammatical Evolution gives slightly better fitness results than the static one, this can be due to the fact that the Static SGE contains all the possible values on their chromosomes, the length is not variable, but a lot of these alleles will not be used on the final phenotype, even though they are going through mutation and crossover. On the other hand, on the Dynamic SGE we have variable length lists and, even though we might obtain longer lists through crossover with another individual, that issue is much less pronounced than with Static SGE. This fact provokes a more effective evolutionary process and results in a better Mean for the fitness over the 30 runs. In Fig. 2 and Fig. 3 we can observe a graphical representation of this table using boxplots, from the graphs we can infer that there is no significant statistic variation between the performance of both algorithms. Likewise, the individual with the best fitness (in training), is not always found on the Dynamic SGE, and there is not a significant difference between the performance of the best individual found by both algorithms, as we will see in the next table. The last two rows of the table show the values obtained for the 2 general models we have tested on, the first one (All) has been trained with the data of all the patients and the other (All w/o 1,2,3) has been trained with the data of all except the first 3 patients.

In table III we show the Recall (percentage of well classified samples for each class over the total for each class) of the best model obtained for each patient (on training) and the recall of the general model, over the test data of each patient, the recall percentage is given for both classes, Hypoglycemia and Not Hypoglycemia.

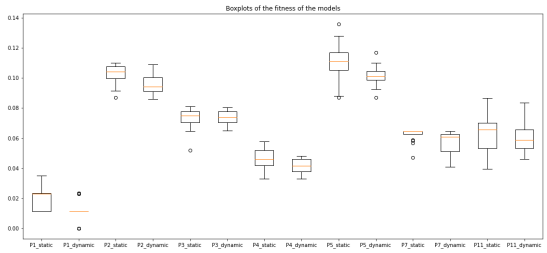

Fig. 2. Boxplot for the best fitness of the 30 generations of the patients HUPA001 to HUPA011.

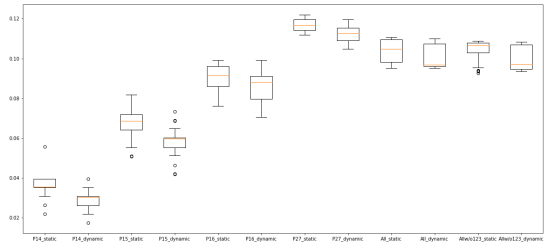

Fig. 3. Boxplot for the best fitness of the 30 generations of the patients HUPA014 to HUPA027 and the 2 general models.

TABLE II
STATISTICS: MEAN, MEDIAN, STANDARD DEVIATION AND BEST FITNESS FOR ALL THE MODELS WE HAVE TRAINED

Patient	Algorithm Type	Mean	Median	Standard deviation	Best Fitness
HUPA001	Static	0.0185	0.0230	0.0070	0.0115
HUPA001	Dynamic	0.0123	0.1041	0.0066	0
HUPA002	Static	0.1023	0.0942	0.0063	0.0870
HUPA002	Dynamic	0.0952	0.0748	0.0063	0.0861
HUPA003	Static	0.0734	0.0739	0.0061	0.0519
HUPA003	Dynamic	0.0738	0.0462	0.0042	0.0651
HUPA004	Static	0.0466	0.0417	0.0068	0.0329
HUPA004	Dynamic	0.0416	0.1112	0.0045	0.0329
HUPA005	Static	0.1101	0.1010	0.0107	0.0869
HUPA005	Dynamic	0.1009	0.0646	0.0056	0.0869
HUPA007	Static	0.0628	0.0608	0.0036	0.0471
HUPA007	Dynamic	0.0569	0.0658	0.0080	0.0409
HUPA011	Static	0.0634	0.0588	0.0116	0.0395
HUPA011	Dynamic	0.0595	0.0354	0.0072	0.0462
HUPA014	Static	0.0363	0.0304	0.0056	0.0218
HUPA014	Dynamic	0.0283	0.0304	0.0050	0.0175
HUPA015	Static	0.0674	0.0686	0.0081	0.0508
HUPA015	Dynamic	0.0581	0.0597	0.0069	0.0417
HUPA016	Static	0.0900	0.0914	0.0068	0.0761
HUPA016	Dynamic	0.0857	0.0881	0.0075	0.0705
HUPA027	Static	0.1169	0.1167	0.0030	0.1117
HUPA027	Dynamic	0.1127	0.1128	0.0038	0.1047
All	Static	0.1037	0.1047	0.0056	0.0951
All	Dynamic	0.1006	0.0969	0.0056	0.0951
All w/o 1,2,3	Static	0.1042	0.1066	0.0052	0.0926
All w/o 1,2,3	Dynamic	0.0996	0.0969	0.0056	0.0936

TABLE III
COMPARISON OF THE RECALL OF THE INDIVIDUAL AND GENERAL MODEL FOR EACH PATIENT

Patient	Algorithm	Individual Models		General Models	
		Recall Hypo	Recall No Hypo	Recall Hypo	Recall No Hypo
HUPA001	Static	1. 000	0.9369	0.8824	0.964
HUPA001	Dynamic	0.9412	0.944	0.9412	0.956
HUPA002	Static	0.9067	0.8864	0.9422	0.8172
HUPA002	Dynamic	0.9067	0.8892	0.9644	0.8296
HUPA003	Static	0.9	0.9155	0.925	0.8848
HUPA003	Dynamic	0.8875	0.9184	0.95	0.8656
HUPA004	Static	0.9677	0.9356	0.957	0.9356
HUPA004	Dynamic	0.9677	0.9461	0.957	0.9368
HUPA005	Static	0.9118	0.8317	0.7941	0.9307
HUPA005	Dynamic	0.9706	0.8425	0.9412	0.9181
HUPA007	Static	0.9293	0.9074	0.9697	0.8959
HUPA007	Dynamic	0.9798	0.9131	0.9596	0.9054
HUPA011	Static	0.9032	0.9379	0.9032	0.9379
HUPA011	Dynamic	0.9032	0.9344	0.9032	0.9406
HUPA014	Static	0.9787	0.9195	0.9574	0.9634
HUPA014	Dynamic	0.9574	0.925	0.9574	0.9543
HUPA015	Static	0.9767	0.8654	0.907	0.929
HUPA015	Dynamic	0.9535	0.8986	0.907	0.9392
HUPA016	Static	0.9711	0.8462	0.9422	0.872
HUPA016	Dynamic	0.9595	0.8669	0.9595	0.8184
HUPA027	Static	0.8898	0.882	0.8898	0.8932
HUPA027	Dynamic	0.8857	0.8963	0.9116	0.8815

TABLE IV
COMPARISON OF THE RECALL FOR HUPA001, HUPA002 AND HUPA003 WITH THE GENERAL MODEL WITHOUT THESE PATIENTS AND THEIR INDIVIDUAL MODELS

Patient	Algorithm	Individual Model		General Model w/o P1 P2 P3	
		Recall Hypo	Recall No Hypo	Recall Hypo	Recall No Hypo
HUPA001	Static	1.	0.9369	0.9483	0.9582
HUPA001	Dynamic	0.9412	0.944	0.9483	0.9585
HUPA002	Static	0.9067	0.8864	0.9454	0.814
HUPA002	Dynamic	0.9067	0.8892	0.9441	0.8119
HUPA003	Static	0.9	0.9155	0.9358	0.882
HUPA003	Dynamic	0.8875	0.9184	0.9358	0.8803

In Fig. 4 we can observe a graphical representation that shows the boxplots for the 4 models, in it we can see that there is no significant statistical variation between the individual and general models. However, the main advantage of using a general model in this case is that we can train it with a lot more data, making it more robust, as some of the patients don't have a lot of hypoglycemic values to train with.

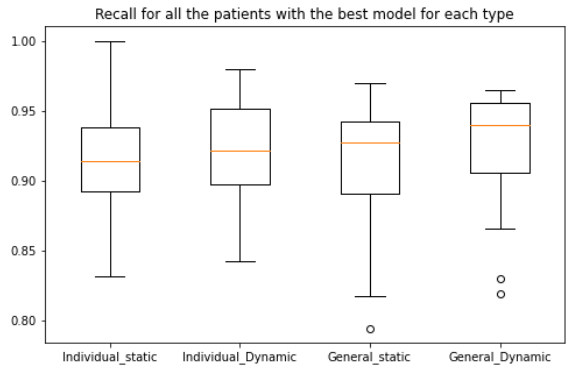

Fig. 4. Boxplot for the recall of the general and individual models

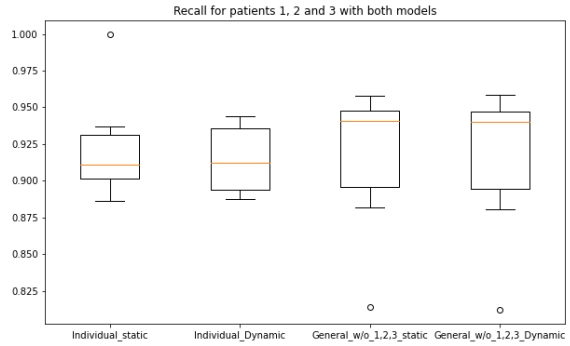

Fig. 5. Boxplot for the recall of the general model without HUPA001, HUPA002 and HUPA003 and their respective individual models

In table IV we expand on the idea that the data from some patients can be used to generate models that perform well on other patients, whose data has not been included in the training. On the table, we can see that the recall percentages for the model on the three patients whose data we have not included are not far from their respective Individual Models. In Fig. 5 we observe once more the boxplots with the recall of both models. Finally, an example of a convergence graph over the 30 execution of the dynamic SGE is shown in Fig. 6.

This type of model creation might not work on all patients, as glucose sensibility and variability can be very different between people, but it is interesting to study as data to train this type of models can be scarce due to the patient's control of their glucose values.

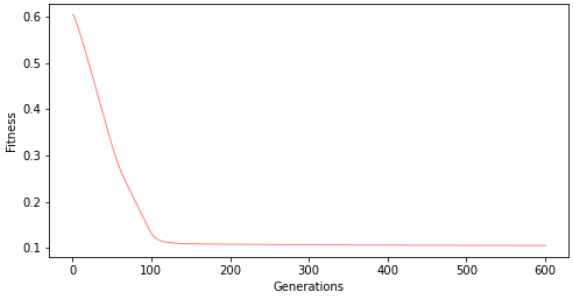

Fig. 6. Convergence Graph for all the population over the 30 execution of the dynamic SGE.

V. Conclusions

In this paper, we investigated the performance of two types of Structured Grammatical Evolution for the prediction of the classification of future glucose values on a short prediction horizon (30 minutes), with input data obtained automatically (through a smartwatch and a CGM). We have also tested the creation of a general model that can be applied to all patients and how this model compares with the individual ones. The main benefit is that the obtained models are very understandable as they are comprised of an if-else statement that performs the classification and uses the input data from the patient to determine the resulting class.

In the future, we can test the performance of the algorithm on a longer prediction horizon and perform a classification with more than 2 classes. The study could also be extended to predict states other than hypoglycemia based on other data such as carbohydrates, insulin, stress data, etc. In addition, the number of experiments can be expanded by adding data from new patients.

Acknowledgment

This work has been supported by Madrid Regional Goverment and FEDER funds under grant Y2018/NMT-4668 (Micro-Stress- MAP-CM). We acknowledge support from Spanish Ministry of Economy and Competitiveness under project RTI2018-095180-B-I00, Madrid Regional Goverment - FEDER grant B2017/BMD3773 (GenObIA-CM). Devices for adquiring data from patients were adquired under the support of Fundación Eugenio Rodriguez Pascual 2019 grant - *Desarrollo de sistemas adaptativos y bioinspirados para el control glucémico con infusores subcutáneos continuos de insulina y monitores continuos de glucosa (Development of adaptive and bioinspired systems for glycaemic control with continuous subcutaneous insulin infusors and continuous glucose monitors)*.

References

[1] Jean-François Yale, Breay Paty, and Peter A. Senior. Hypoglycemia. *Canadian Journal of Diabetes*, 42:S104–S108, 2018. Diabetes Canada 2018 Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada.

[2] Elizabeth R. Seaquist, John Anderson, Belinda Childs, Philip Cryer, Samuel Dagogo-Jack, Lisa Fish, Simon R. Heller, Henry Rodriguez, James Rosenzweig, and Robert Vigersky. Hypoglycemia and Diabetes: A Report of a Workgroup of the American Diabetes Association and The Endocrine Society. *Diabetes Care*, 36(5):1384–1395, 04 2013.

[3] Javier Morales and Doron Schneider. Hypoglycemia. *The American Journal of Medicine*, 127(10):S17–S24, 2014.

[4] Thomas Peyser, Andrew Balo, Bruce Buckingham, Irl Hirsch, and Arturo Garcia. Glycemic variability percentage: A novel method for assessing glycemic variability from continuous glucose monitor data. *Diabetes Technology and Therapeutics*, 20, 12 2017.

[5] Jingzhen Li, Jingyi Lu, Igbe Tobore, Yuhang Liu, Abhishek Kandwal, Lei Wang, Xiaojing Ma, Wei Lu, Yuqian Bao, Jian Zhou, et al. Gradient variability coefficient: a novel method for assessing glycemic variability and risk of hypoglycemia. *Endocrine*, pages 1–7, 2022.

[6] Omar Diouri, Monika Cigler, Martina Vettoretti, Julia K. Mader, Pratik Choudhary, Eric Renard, and HYPO-RESOLVE Consortium. Hypoglycaemia detection and prediction techniques: A systematic review on the latest developments. *Diabetes/Metabolism Research and Reviews*, 37(7):e3449, 2021.

[7] Jose Ignacio Hidalgo, Jose Manuel Velasco, Juan Lanchares, Sergio Contador, and Oscar Garnica. *An analysis of solutions based on Genetic Programming to solve problems of symbolic regression of data from continuous glucose monitoring*. AEPIA, Granada, Spain, 2018.

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

[9] Jose Velasco, Oscar Garnica, Juan Lanchares, Marta Botella, and Ignacio Hidalgo. Combining data augmentation, edas and grammatical evolution for blood glucose forecasting. *Memetic Computing*, 10, 06 2018.

[10] Mirko Messori, Chiara Toffanin, Simone Del Favero, Giuseppe De Nicolao, Claudio Cobelli, and Lalo Magni. Model individualization for artificial pancreas. *Computer Methods and Programs in Biomedicine*, 171:133–140, 2016.

[11] 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. 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):175–183, 2013.

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

[13] J. Ignacio Hidalgo, J. Manuel Colmenar, Gabriel Kronberger, Stephan M. Winkler, Oscar Garnica, and Juan Lanchares. Data based prediction of blood glucose concentrations using evolutionary methods. *Journal of Medical Systems*, 41, 2017.

[14] J. Ignacio Hidalgo, J. Manuel Colmenar, José L. Risco-Martin, Alfredo Cuesta-Infante, Esther Maqueda, Marta Botella, and José Antonio Rubio. Modeling glycemia in humans by means of grammatical evolution. *Applied Soft Computing*, (20):40–53, 2014.

[15] J. Ignacio Hidalgo, J. Manuel Colmenar, J. Manuel Velasco, Gabriel Kronberger, Stephan M. Winkler, Oscar Garnica, and Juan Lanchares. *Identification of Models for Glucose Blood Values in Diabetics by Grammatical Evolution*, pages 367–393. Springer International Publishing, Cham, 2018.

[16] Carlos Cervigón, J. Manuel Velasco, Clara Burgos-Simón, Rafael J. Villanueva, and J. Ignacio Hidalgo. Probabilistic fitting of glucose models with real-coded genetic algorithms. pages 736–743, 2021.

[17] Clara Burgos Simón, Carlos Cervigón, José Ignacio Hidalgo, and Rafael J. Villanueva. A computational technique to predict the level of glucose of a diabetic patient with uncertainty in the short term. *Comput. Math. Methods*, 2(2), 2020.

[18] Rahul C. Deo. Machine learning in medicine. *Circulation*, 132(20):1920–1930, 2015.

[19] Jan A. Roth, Manuel Battegay, Fabrice Juchler, Julia E. Vogt, and Andreas F. Widmer. Introduction to machine learning in digital healthcare epidemiology. *Infection Control; Hospital Epidemiology*, 39(12):1457–1462, 2018.

[20] Omer Mohammed Mujahid, Iván Contreras, and Josep Vehí. Machine learning techniques for hypoglycemia prediction: Trends and challenges. *Sensors (Basel, Switzerland)*, 21, 2021.

[21] Arthur Bertachi, Clara Viñals, Lyvia Biagi, Ivan Contreras, Josep Vehí, Ignacio Conget, and Marga Giménez. Prediction of nocturnal hypoglycemia in adults with type 1 diabetes under multiple daily injections using continuous glucose monitoring and physical activity monitor. *Sensors*, 20(6), 2020.

[22] Mohammad Reza Vahedi, Koenrad B. MacBride, Woo Wunsik, Yosep Kim, Chris Fong, Andrew J. Padilla, Mohammad Pourhomayoun, Alex Zhong, Sameer Kulkarni, Siddharth Arunachalam, and Boyi Jiang. Predicting glucose levels in patients with type1 diabetes based on physiological and activity data. In *Proceedings of the 8th ACM MobiHoc 2018 Workshop on Pervasive Wireless Healthcare Workshop*, MobileHealth'18, New York, NY, USA, 2018. Association for Computing Machinery.

[23] Clara Mosquera-Lopez, Robert Dodier, Nichole Tyler, Navid Resalat, and Peter Jacobs. Leveraging a big dataset to develop a recurrent neural network to predict adverse glycemic events in type 1 diabetes. *IEEE Journal of Biomedical and Health Informatics*, pages 1–1, 2019.

[24] Yonghao Jin, Fei Li, Varsha G Vimalananda, and Hong Yu. Automatic detection of hypoglycemic events from the electronic health record notes of diabetes patients: Empirical study. *JMIR Med Inform*, 7(4):e14340, Nov 2019.

[25] Hrushikesh N. Mhaskar, Sergei V. Pereverzyev, and Maria D. van der Walt. A deep learning approach to diabetic blood glucose prediction. *Frontiers in Applied Mathematics and Statistics*, 3, 2017.

[26] Wonju Seo, You-Bin Lee, Seunghyun Lee, Sang-Man Jin, and Sung-Min Park. A machine-learning approach to predict postprandial hypoglycemia. *BMC Medical Informatics and Decision Making*, 19(1):210, 2019.

[27] Josep Vehí, Iván Contreras, Silvia Oviedo, Lyvia Biagi, and Arthur Bertachi. Prediction and prevention of hypoglycaemic events in type-1 diabetic patients using machine learning. *Health Informatics Journal*, 26(1):703–718, 2020. PMID: 31195880.

[28] Ravi Reddy, Navid Resalat, Leah M. Wilson, Jessica R. Castle, Joseph El Youssef, and Peter G. Jacobs. Prediction of hypoglycemia during aerobic exercise in adults with type 1 diabetes. *Journal of Diabetes Science and Technology*, 13(5):919–927, 2019. PMID: 30650997.

[29] Predicting and Preventing Nocturnal Hypoglycemia in Type 1 Diabetes Using Big Data Analytics and Decision Theoretic Analysis. *Diabetes technology & therapeutics*, 22(11):801–811, nov 2020.

[30] Amparo Güemes, Giacomo Cappon, Bernard Hernandez, Monika Reddy, Nick Oliver, Pantelis Georgiou, and Pau Herrero. Predicting quality of overnight glycaemic control in type 1 diabetes using binary classifiers. *IEEE Journal of Biomedical and Health Informatics*, 24(5):1439–1446, 2020.

[31] Jinying Chen, John Lalor, Weisong Liu, Emily Druhl, Edgard Granillo, Varsha G Vimalananda, and Hong Yu. Detecting hypoglycemia incidents reported in patients' secure messages: Using cost-sensitive learning and oversampling to reduce data imbalance. *J Med Internet Res*, 21(3):e11990, Mar 2019.

[32] Darpit Dave, Daniel J. DeSalvo, Balakrishna Haridas, Siripoom McKay, Akhil Shenoy, Chester J. Koh, Mark Lawley, and Madhav Erraguntla. Feature-based machine learning model for real-time hypoglycemia prediction. *Journal of Diabetes Science and Technology*, 15(4):842–855, 2021. PMID: 32476492.

[33] Nuno Lourenço, J. Manuel Colmenar, J. Ignacio Hidalgo, and Óscar Garnica. Structured grammatical evolution for glucose prediction in diabetic patients. page 1250–1257, 2019.

[34] Nuno Lourenço, Joaquim Ferrer, Francisco Pereira, and Ernesto Costa. A comparative study of different grammar-based genetic programming approaches. pages 311–325, 03 2017.

[35] Conor Ryan, JJ Collins, and Michael Neill. Grammatical evolution: Evolving programs for an arbitrary language. In Wolfgang Banzhaf, Riccardo Poli, Marc Schoenauer, and Terence Fogarty, editors, *Genetic Programming*, volume 1391 of *Lecture Notes in Computer Science*, pages 83–96. Springer Berlin / Heidelberg, 1998.