

# Diabetes and Prediabetes Classification Using Glycemic Variability Indices From Continuous Glucose Monitoring Data

Journal of Diabetes Science and Technology  
2018, Vol. 12(1) 105–113  
© 2017 Diabetes Technology Society  
Reprints and permissions:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/1932296817710478  
journals.sagepub.com/home/dst  


Giada Acciaroli, MS<sup>1</sup>, Giovanni Sparacino, PhD<sup>1</sup>, Liisa Hakaste, MD, PhD<sup>2,3</sup>,  
Andrea Facchinetti, PhD<sup>1</sup>, Giorgio Maria Di Nunzio, PhD<sup>1</sup>,  
Alessandro Palombit, MS<sup>1</sup>, Tiinamaija Tuomi, MD, PhD<sup>2,3,4</sup>,  
Rafael Gabriel, MD, PhD<sup>5</sup>, Jaime Aranda, MD, PhD<sup>6</sup>,  
Saturio Vega, MD, PhD<sup>7</sup>, and Claudio Cobelli, PhD<sup>1</sup>

## Abstract

**Background:** Tens of glycemic variability (GV) indices are available in the literature to characterize the dynamic properties of glucose concentration profiles from continuous glucose monitoring (CGM) sensors. However, how to exploit the plethora of GV indices for classifying subjects is still controversial. For instance, the basic problem of using GV indices to automatically determine if the subject is healthy rather than affected by impaired glucose tolerance (IGT) or type 2 diabetes (T2D), is still unaddressed. Here, we analyzed the feasibility of using CGM-based GV indices to distinguish healthy from IGT&T2D and IGT from T2D subjects by means of a machine-learning approach.

**Methods:** The data set consists of 102 subjects belonging to three different classes: 34 healthy, 39 IGT, and 29 T2D subjects. Each subject was monitored for a few days by a CGM sensor that produced a glucose profile from which we extracted 25 GV indices. We used a two-step binary logistic regression model to classify subjects. The first step distinguishes healthy subjects from IGT&T2D, the second step classifies subjects into either IGT or T2D.

**Results:** Healthy subjects are distinguished from subjects with diabetes (IGT&T2D) with 91.4% accuracy. Subjects are further subdivided into IGT or T2D classes with 79.5% accuracy. Globally, the classification into the three classes shows 86.6% accuracy.

**Conclusions:** Even with a basic classification strategy, CGM-based GV indices show good accuracy in classifying healthy and subjects with diabetes. The classification into IGT or T2D seems, not surprisingly, more critical, but results encourage further investigation of the present research.

## Keywords

classification, continuous glucose monitoring, glycemic variability, impaired glucose tolerance, type 2 diabetes

The concept of glycemic variability (GV) is broadly used in diabetes context to characterise the fluctuations of blood glucose (BG) profiles, which are often involved in the pathogenesis of diabetes-related complications.<sup>1-8</sup> Tens of different metrics were proposed in the literature to quantify the GV, including indices derived from the distribution of glucose readings or the amplitude and duration of the glycemic excursions, indices based on risk and quality of glycemic control, just to name a few (see Ohara et al,<sup>8</sup> Rodbard,<sup>9</sup> Le Floch and Kessler,<sup>10</sup> Kovatchev and Cobelli<sup>11</sup>).

The GV concept has become even more intriguing since the advent of continuous glucose monitoring (CGM) sensors. These sensors measure glucose concentration every 1-5 minutes for several consecutive days, allowing the characterization

<sup>1</sup>Department of Information Engineering, University of Padova, Padova, Italy

<sup>2</sup>Endocrinology, Abdominal Centre, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

<sup>3</sup>Folkhälsan Research Center, and Research Program for Diabetes and Obesity, University of Helsinki, Helsinki, Finland

<sup>4</sup>Finnish Institute for Molecular Medicine, University of Helsinki, Helsinki, Finland

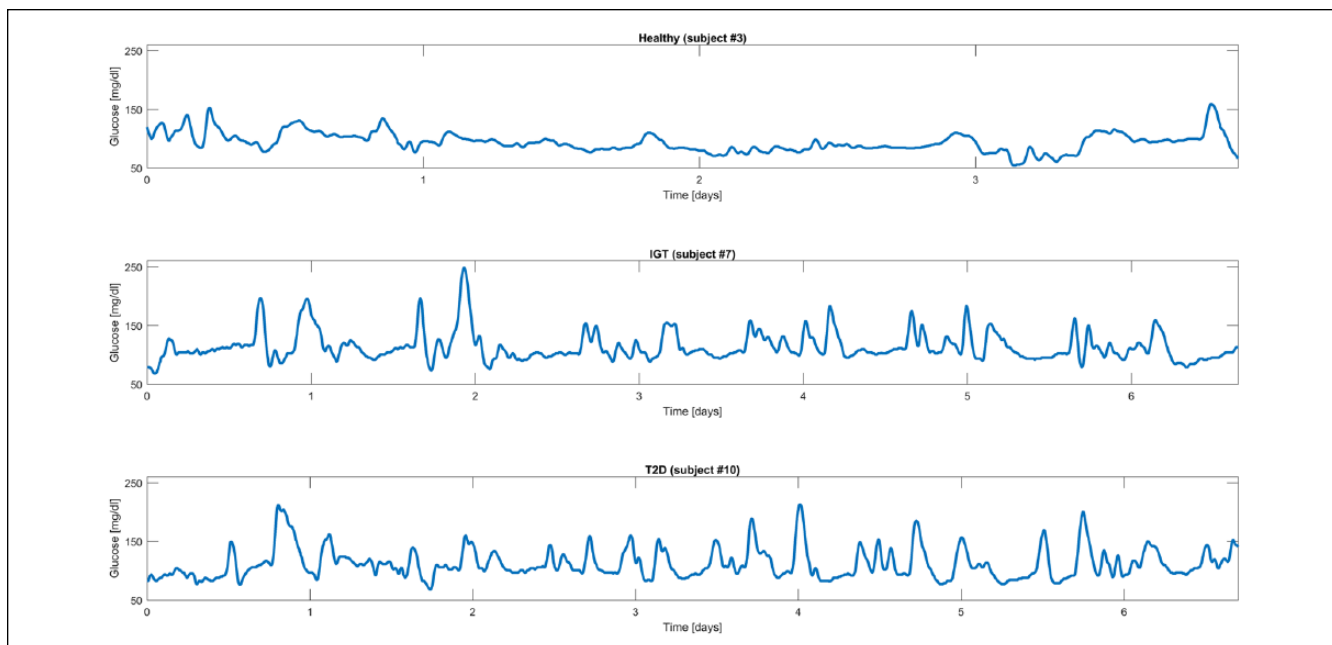
<sup>5</sup>Escuela Nacional de Sanidad, Instituto de Salud Carlos III, Madrid, Spain

<sup>6</sup>Servicio de Endocrinología Hospital General de Cuenca, Cuenca, Spain

<sup>7</sup>Centro de Salud de Arevalo, Avila, Spain

## Corresponding Author:

Claudio Cobelli, PhD, Department of Information Engineering, University of Padova, Via Gradenigo 6/B, Padova, PD 35131, Italy.  
Email: cobelli@dei.unipd.it



**Figure 1.** Representative CGM traces from representative healthy (first panel), IGT (second panel), and T2D (third panel) subjects.

of the dynamic properties of BG profiles by capturing components otherwise invisible with the traditional self-monitoring of blood glucose (SMBG) management.<sup>11</sup> GV indices computable from CGM traces<sup>12-16</sup> have been used in several studies to assess the impact of GV on the risk of developing some diabetes complications,<sup>17-19</sup> to quantify the quality of the glycemic control,<sup>20</sup> and to stratify CGM traces in relation to the need of therapeutic actions.<sup>21</sup>

Recently, several studies employed CGM technology not only in the population with diabetes,<sup>22,23</sup> but also in subjects affected by states of prediabetes and in obese individuals,<sup>24,25</sup> where a progressively increasing GV as moving from normal subjects to subjects with prediabetes has been observed.<sup>26</sup> Such progressive changes in glucose dynamics in different categories of subjects suggest that GV metrics extracted from CGM signals could be used to detect impaired glycemia in certain groups of subjects<sup>27</sup> much earlier than the standard techniques used for the diagnosis and classification of diabetes, based on oral glucose tolerance test (OGTT) and HbA1c values.<sup>23</sup>

The problem of how to best exploit the plethora of GV indices to distinguish among different categories of subjects is still debated since there are unresolved issues like, for example, GV indices that carry on redundant information.<sup>28,29</sup> In this context, even a basic question like using GV indices to automatically recognize if a subject is healthy rather than affected by a state of prediabetes, such as impaired glucose tolerance (IGT), or by type 2 diabetes (T2D), is still unaddressed in the literature to the best of our knowledge.

In the present work, we use a machine learning approach to distinguish healthy from IGT&T2D and IGT from T2D subjects using a set of 25 well established CGM-based indices.

## Methods

### Database

The data set consists of 102 subjects belonging to three different classes, preliminary determined for each individual by a standard OGTT: 34 healthy subjects, 39 IGT subjects, and 29 T2D subjects (IGT and T2D subjects participated in the Botnia Study in Finland, approved by the Ethics committee of the Helsinki University Hospital with informed consent from all study participants). Each subject was monitored by either the Guardian Real Time or the iPro CGM systems (Medtronic MiniMed, Inc, Northridge, CA) under normal life conditions for an average period of 5 days. Figure 1 shows an exemplificative CGM trace for each of the three categories (healthy subject in the first panel, IGT subject in the second panel, and T2D subject in the third panel).

### Indices Used for Classification

The 102 CGM traces were first processed to extract the set of well-established 25 indices already considered in two recent studies by Fabris et al.<sup>28,29</sup> Each GV index has been mean-centered and scaled before entering the classification procedure. The pool of indices comprises metrics

based on statistical properties, that is, standard deviation (SD), coefficient of variation (CV), range, interquartile range (IQR), and J-index,<sup>30</sup> indices based on interday variability of statistical properties, that is, mean of daily SD (SDw) and SD of daily means (SDdm),<sup>14</sup> indices based on the permanence in the euglycemic target range, that is, percentages of values below the target range (<70 mg/dL), above the target range (>180 mg/dL) and within the target range ([70, 180] mg/dL), indices related to significant glycemic excursions, that is, mean amplitude of glycemic excursions (MAGE) index,<sup>31,32</sup> and measures derived from nonlinear transformations of glucose values, that is, low and high blood glucose indices (LBGI, HBGI),<sup>33,34</sup> blood glucose risk index (BGRI),<sup>35</sup> average daily risk range (ADRR),<sup>36</sup> hypoglycemic index, hyperglycemic index and index of glycemic control (IGC),<sup>15</sup> glycemic risk assessment diabetes equation (GRADE) score with its three different glycemic states (%GRADEhypo, %GRADEeu, %GRADEhyper),<sup>37</sup> and, finally, the  $M_{100}$  index.<sup>38</sup> In addition, the pool of indices comprises mean and median glucose. While, from a certain point of view, these two indices are not exactly related with the concept of “variability,” they are normally included in the tools used for glucose time-series analysis.<sup>28,29</sup> Therefore, for the sake of reasoning, in the reminder of the present work all the above-mentioned indices will be referred to as GV indices. The correlation among the indices in the considered sample of subjects is briefly commented at the end of the appendix.

### Classification Strategy Overview

The proposed classification system is based on a logistic regression model.<sup>39</sup> In particular, to assign each subject to a class, a two-step binary logistic regression model is implemented. The first step aims to distinguish healthy from subjects with diabetes (obtained merging IGT and T2D classes). At the second step, all subjects assigned to the IGT&T2D group at the previous step are classified into IGT or T2D.

The two classifiers are trained on a subset of the 25 GV indices (specific for each of the two classification steps) that best distinguish between classes, that is, the subset of features that guarantee higher classification performance, as determined by a feature-selection algorithm. In particular, the subset of indices that best correlate with classification performance automatically selected by the feature selection procedure is composed of 8 over 25 indices for the first classification step, and 5 over 25 indices for the second classification step (see the appendix for details).

In our experiments, the classifier is trained using a stratified 5-fold cross-validation approach in which the data set is subdivided into five equally sized folds where

the proportion of healthy, IGT and T2D subjects in each fold reflects the distribution in the entire data set. Iteratively, one of the five folds composes the test set, used to evaluate classification performance, while the remaining four folds are used to train the classifier. In this way, we guarantee independence between training and test of the classifier and fairness of the results. Training and test phases are briefly described in the following sections.

### Training Phase

During the training phase, the set of 25 GV indices and the class label of each subject are used to build the classifier. The logistic regression model we used in the experiments can be expressed by the following formula:

$$\begin{cases} P(y = C_1|\mathbf{x}) = \frac{1}{1 + e^{-\mathbf{w}^T \mathbf{x}}} \\ P(y = C_2|\mathbf{x}) = 1 - P(y = C_1|\mathbf{x}) \end{cases} \quad (1)$$

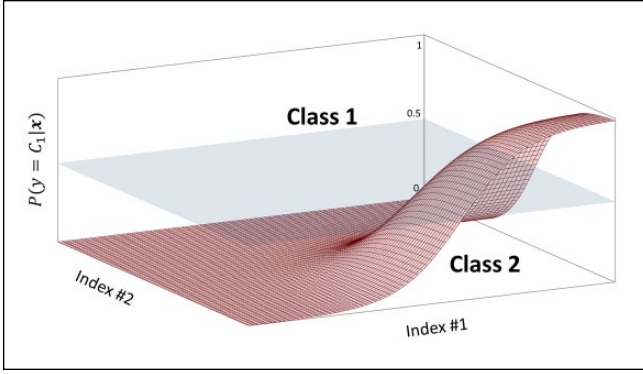
where  $\mathbf{x}$  is the vector containing the 25 GV indices extracted from the CGM trace for each subject,  $P(y = C_k|\mathbf{x})$ ,  $k = 1, 2$  is the posterior probability of the  $k$ -th class given the set of GV indices  $\mathbf{x}$ , and  $\mathbf{w}$  is the vector of unknown parameters to be determined. During the training phase, the unknown parameters vector  $\mathbf{w}$  is estimated by maximizing the log-likelihood  $l(\mathbf{w})$ :

$$\begin{cases} l(\mathbf{w}) = \sum_{i=1}^{N_t} \log[p(y = C_k | \mathbf{x} = \mathbf{x}_i, \mathbf{w})] \\ \hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmax}} \{l(\mathbf{w})\} \end{cases} \quad (2)$$

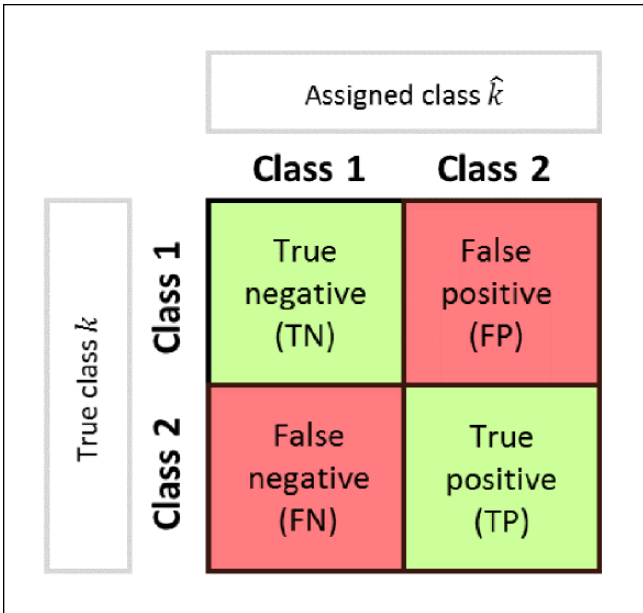
where  $N_t$  is the number of subjects composing the training group and  $\hat{\mathbf{w}}$  is the estimated parameters vector. Practically, the vector  $\hat{\mathbf{w}}$  defines the sigmoid function that best subdivides the probabilistic space of the two classes. Letting  $n$  be the number of indices considered for the classification ( $n < 25$  after the feature-selection step, see the appendix), the  $n$ -dimensional vector of indices  $\mathbf{x}$  represents, together with its class label, the position of each subject in the probabilistic space, which is subdivided into two areas corresponding to the two classes by the  $n$ -dimensional sigmoid function defined by  $\hat{\mathbf{w}}$ . A simplified example of sigmoid function with  $n = 2$  is shown in Figure 2.

### Test Phase

The test phase is used to evaluate classification performance on an unseen subset of subjects, that is, the test group. In particular, the set of GV indices  $\mathbf{x}$  is used to compute, for



**Figure 2.** Simplified representation of sigmoid function with  $n = 2$  dimensions. The sigmoid surface divides the probabilistic space into two areas, corresponding to the two classes.



**Figure 3.** Scheme of confusion matrix for the evaluation of classification performance.

each subject, the two posterior probabilities of Equation 1, where the parameters vector  $\boldsymbol{w}$  is given by  $\hat{\boldsymbol{w}}$ , estimated during the training phase. The class label  $k = \{1, 2\}$  is then determined by choosing the class with higher a posteriori probability:

$$\hat{k} = \underset{k}{\operatorname{argmax}} P(y = C_k | \boldsymbol{x}, \hat{\boldsymbol{w}}) \quad (3)$$

Practically, in relation to the graphical representation of Figure 2, solving Equation 3 corresponds to: look at the vector of coordinates  $\boldsymbol{x}$  in the probabilistic space; assign the subject to class 1 if the correspondent value of the sigmoid function is above 0.5, to class 2 if is under.

### Assessment Criteria

Classification performance is quantified by comparing the class predicted by the classifier,  $\hat{k}$ , with the true class  $k$  for each subject. We use a confusion matrix to summarize the performance of the classifier, as shown in Figure 3. A confusion matrix reports the four possible outputs of the comparison between the true and the predicted class, that is, true negative (TN), false negative (FN), true positive (TP), and false positive (FP). The elements of the matrix allow the definition of the following performance metrics:

- Accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

that represents the fraction of subjects correctly classified among the total number of subjects examined.

- Precision:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

that is, the fraction of TP among the total number of positives. It gives intuitively the ability of the classifier not to label as positive a sample that is negative.

- Recall:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

that is, the fraction of TP among the sum of TP and FN (the total number of elements of that class), which intuitively measures the ability of the classifier to detect all the positive samples.

- F1 score:

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (7)$$

that represents the harmonic mean of precision and recall. Its value is between 0 and 1, where 0 and 1 indicate, respectively, poor and good classification performance.

### Results

Attained 5-fold cross-validation results are shown in terms of accuracy in Table 1, where we reported mean and SD among the five cross-validation folds. In particular, the first classification step distinguishing healthy from subjects with diabetes (comprising both IGT and T2D classes) has mean accuracy of 91.4% with 9.8% SD, whereas the second classification step, which further distinguishes

**Table 1.** Classification Accuracy (First and Second Classification Steps and Global Result).

Classification	First step Healthy/IGT&T2D	Second step IGT/T2D	Global Healthy/IGT/T2D
Accuracy* % (Without feature selection)	91.4 ± 9.8 (81.6 ± 15.5)	79.5 ± 15.9 (60.1 ± 21.3)	86.6 ± 11.7 (73.1 ± 17.6)

Results when no feature selection is performed are also reported (in parentheses) for comparison.

\*Average over the five cross-validation folds.

**Table 2.** Precision, Recall, and F1 Score (Healthy, IGT, and T2D Classes).

Class	Healthy	IGT	T2D
Precision* % (Without feature selection)	87.8 ± 12.5 (74.2 ± 25.8)	87.3 ± 12.9 (71.7 ± 18.2)	81.7 ± 12.3 (66.8 ± 21)
Recall* % (Without feature selection)	85.7 ± 20.2 (72.3 ± 24.8)	86.4 ± 10.1 (69.1 ± 28.8)	82 ± 18.5 (68.7 ± 18.8)
F1 score* (Without feature selection)	0.86 ± 0.16 (0.72 ± 0.23)	0.87 ± 0.11 (0.69 ± 0.24)	0.82 ± 0.15 (0.66 ± 0.18)

Results when no feature selection is performed are also reported (in parentheses) for comparison.

\*Average over the five cross-validation folds.

between IGT and T2D, has mean accuracy of 79.5% with 15.9% SD. The global classification into the three classes shows mean accuracy among the five folds of 86.6% with 11.7% SD.

A deeper analysis of classification performance is done by quantifying the number of subjects correctly and wrongly classified at each step. At the first classification step, 4 healthy subjects are wrongly classified as subjects with diabetes (IGT&T2D) and 7 subjects with diabetes are wrongly classified as healthy, while all other 91 subjects are correctly classified. At the second classification step, that is applied in cascade to the first one, 5 IGT subjects are wrongly classified as T2D, 7 T2D subjects are wrongly labeled as IGT, while all the other 49 subjects are correctly classified.

In Table 2, we reported precision, recall, and F1 score for each of the three classes: healthy (second column), IGT (third column), and T2D (fourth column). Precision and recall are computed in percentage, showing values above 80% for all the three classes, whereas the F1 score, which can vary between 0 and 1, is greater than 0.8 for all the three classes.

In both Table 1 and Table 2, we also reported the classification performance obtained when no feature selection is performed, that is, when the classifiers are trained on the complete set of 25 GV indices. This analysis is performed and documented to emphasize the relevance of the feature selection process. Indeed, all performance metrics result lower when compared to that obtained after the feature selection process.

## Discussion

A first aspect to be noted from Table 1 is that the subdivision of the subjects between healthy and subjects with diabetes (comprising both IGT and T2D classes) is much more accurate than the subsequent subdivision of IGT&T2D into IGT or T2D. Results of the first classification step ensure that we are able to correctly distinguish between normal and pathologic condition in nine cases over ten. This result appears even more interesting if one notes that the subjects of the healthy group present a mean BMI of 30.01 kg/m<sup>2</sup> at limit with obese status. On the other hand, the further classification between IGT and T2D subjects appears less effective, with mean accuracy of 79.5%. Analyzing the structure of the classification problem, which is a two-step procedure, we note that the second classification step suffers from error propagation. Indeed, subjects that are wrongly classified by the first classification step, that is, healthy subjects that are labeled as subjects with diabetes will obviously result in errors also in the second classification step. However, all the other metrics shown in Table 2 (precision, recall, and F1 score) are substantially stable among the three classes with values always above 80%.

The second interesting aspect is the outcome of the feature selection, which, for the first classification step, is composed of the following 8 indices: mean, CV, range, percentage of values below target, percentage of values within target, HBGI, ADRR, and BGRI; and for the second classification step is composed of the following 5 indices: median, SD, CV, MAGE, LBGI. In fact, on one

hand, the analysis of these two subsets of indices confirms the high redundancy carried by the original set of 25 GV indices, as already observed in Fabris et al,<sup>28,29</sup> whereas, on the other hand, evidences the specificity of each index in describing particular characteristics of the glucose dynamics. Indeed, the two subsets have no overlap, with the exception of CV, a standard statistical feature that is powerful to assess glucose variability<sup>40</sup> and appears useful for both classification steps when combined with more specific GV metrics. This aspect suggests that more class-specific indices could be included in the set of features that serves as input to the machine-learning algorithm, with the aim of improving class-to-class subdivision.

## Conclusions

Tens of GV indices were proposed in the literature, even more since the advent of CGM sensors, and used for several purposes, such as to quantify the quality of glycemic control or to assess the risk of developing diabetes-related complications, but their usability and reliability for classification problems is still scarcely investigated.

In the present work, given a data set of 102 subjects, we analyzed the performance of an automatic classifier that distinguishes between healthy, IGT, and T2D subjects using GV metrics extracted from CGM traces. Results confirm that CGM-based GV indicators can effectively distinguish CGM traces of healthy and subjects with diabetes (IGT&T2D), with an overall mean classification accuracy of 91.4%. The subdivision of subjects into IGT or T2D appears, not surprisingly, more critical, but results (79.5% accuracy) encourage further investigation of the present research. Furthermore, the global subdivision into the three classes shows 86.6% accuracy.

Future developments will consider the application of the classification algorithm to different and possibly larger data sets that could permit the construction of more robust classifiers, as well as the implementation of different, more sophisticated machine-learning techniques. Possibly, inputs of the classification algorithms could also be extended by adding other different GV indices and, if available, some basic clinical parameters (eg, age, height, weight) that could further improve the GV indices-based classification performance.

## Appendix

When dealing with classification problems, it is important to check whether all available information is actually needed to efficiently solve the problem. In general, it is good

practice to reduce the set of features that describe the object by removing the “noisy” features that are not crucial (or even that influence negatively) for distinguishing between the considered classes.

Here, we analyze if the whole set of features, that is, the complete set of 25 GV indices, is actually needed to distinguish between classes, or if it is instead possible to reduce it by selecting a specific subset of indices. This problem can be solved through many feature selection methods. In this article, we used a forward sequential feature selection (SFS) approach.<sup>41</sup> The forward SFS method finds the best subset  $S^*$  of features, for example, the best subset of GV indices in our specific case, following an iterative procedure:

1. Initialize the set of features  $S$  to the empty set  $S = \emptyset$ ;
2. Until all features have been added to  $S$ :
  - 2.1. For every feature  $f_i$  not in  $S$ :
    - 2.1.1. Build a classifier using  $S \cup \{f_i\}$  as input
    - 2.1.2. Compute the performance measure (F1 score)
  - 2.2. Find  $f^* = f_i$  such that adding it to  $S$  maximizes the performance measure
  - 2.3. Update  $S$  to  $S = S \cup f^*$  and store the value of the performance measure
3.  $S^*$  will be composed of the first  $Z$  elements of  $S$  such that the value of the relative performance measure (stored at step 2.3) is maximized.

This procedure is run at each of the five cross-validation iterations by subdividing the training set into actual training part and validation part (following an 80% to 20% partition rule). The training part is used to build the different classifiers (point 2.1.1 in the algorithm above) and the validation part is used to compute the performance measure (here given by the F1 score on the validation group). The features selected by the SFS nested in the cross-validation scheme are quite stable among the folds, that is, many features recur in the five subsets while few others appear to depend on the training set. The global output of the SFS reported in the Discussion section is instead obtained on the entire data set.

For sake of completeness, we report in Table A1 the correlation matrix of the 25 GV indices estimated from the considered data set. It clearly shows that indices are somehow mutually related. Interestingly, compared to the same matrix reported by Fabris et al<sup>29</sup> for a homogeneous population of T2D subjects, here lower correlation coefficients are found, because of the greater heterogeneity of the subjects (healthy, IGT, and T2D).

**Table A1.** Correlation Matrix of the 25 GV Indices in the Entire Population.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
1. Mean	1																									
2. SD	.56	1																								
3. %CV	.27	.93	1																							
4. SDw	.62	.96	.88	1																						
5. SDdm	.24	.81	.79	.65	1																					
6. Median	.98	.45	.15	.52	.12	1																				
7. IQR	.53	.93	.82	.86	.87	.42	1																			
8. Range	.61	.91	.86	.93	.59	.52	.75	1																		
9. %val below tgt	-.40	.24	.44	.17	.40	-.45	.23	.14	1																	
10. %val in tgt	.02	-.57	-.65	-.51	-.61	.10	-.55	-.44	-.89	1																
11. %val above tgt	.69	.80	.61	.80	.61	.61	.80	.71	.10	-.54	1															
12. MAGE	.60	.96	.87	.98	.66	.49	.85	.92	.16	-.48	.76	1														
13. J-index	.89	.86	.62	.87	.59	.83	.83	.81	-.08	-.33	.89	.85	1													
14. M-value	.45	.85	.77	.81	.78	.37	.84	.72	.48	-.81	.88	.78	.78	1												
15. LBGI	-.68	.01	.27	-.07	.26	-.71	.01	-.09	.90	-.72	-.08	-.36	.29	.29	1											
16. HBGI	.77	.84	.63	.84	.65	.70	.84	.75	.08	-.51	.97	.81	.95	.89	-.14	1										
17. ADRR	.22	.85	.91	.82	.72	.12	.75	.80	.58	-.80	.67	.80	.58	.87	.44	.67	1									
18. BGRI	-.13	.52	.62	.45	.62	-.20	.52	.37	.84	-.95	.52	.42	.26	.80	.80	.48	.80	1								
19. Hypo index	-.31	.22	.43	.18	.29	-.34	.14	.22	.86	-.77	.10	.14	-.06	.47	.80	.07	.60	.75	1							
20. Hyper index	.46	.85	.72	.79	.83	.36	.89	.67	.26	-.61	.85	.78	.78	.91	.10	.87	.75	.62	.18	1						
21. IGC	-.17	.43	.58	.37	.48	-.22	.36	.37	.86	-.87	.31	.34	.15	.67	.76	.29	.75	.85	.97	.43	1					
22. GRADE	.72	.77	.63	.78	.54	.67	.71	.74	.30	-.64	.85	.73	.86	.88	.02	.90	.71	.56	.36	.69	.51	1				
23. %GRADE hypo	-.57	-.01	.23	-.06	.16	-.58	-.03	-.05	.82	-.65	-.10	-.10	-.33	.23	.88	-.16	.42	.68	.81	.00	.74	.06	.06	1		
24. %GRADE eu	.20	-.42	-.58	-.39	-.44	.26	-.36	-.36	-.81	.85	-.37	-.34	-.13	-.62	-.77	-.30	-.74	-.86	-.79	-.39	-.83	-.46	-.86	-.86	1	
25. %GRADE hyper	.67	.85	.71	.88	.57	.57	.77	.80	.07	-.47	.91	.87	.86	.79	-.13	.89	.68	.43	.05	.76	.25	.78	-.16	-.36	-.36	1

## Abbreviations

ADRR, average daily risk range; BG, blood glucose; BGRI, blood glucose risk index; CGM, continuous glucose monitoring; CV, coefficient of variation; GRADE, glycemic risk assessment diabetes equation; GV, glucose variability; HBGI, high blood glucose index; IGC, index of glycemic control; IGT, impaired glucose tolerance; IQR, interquartile range; LBGI, low blood glucose index; MAGE, mean amplitude of glycemic excursions; OGTT, oral glucose tolerance test; SD, standard deviation; SDw, mean of daily SD; SDdm, SD of daily means; SFS, sequential feature selection; SMBG, self-monitoring of blood glucose; T2D, type 2 diabetes.

## Authors' Note

The present study has been presented in preliminary form at the 16th Annual Diabetes Technology Meeting, November 10-12, 2016, Bethesda, MD, with the work titled "Good Accuracy of CGM-Based Glucose Variability Indices for IGT and T2D Classification," by Acciaroli et al, winning the Gold Prize of the Diabetes Technology Society Student Research Award.

## Acknowledgments

The Botnia Study Group is acknowledged for recruiting and studying the participants from the Botnia Study (<http://www.botnia-study.org/en/home/>).

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was performed as part of the FP7-MOSAIC project funded by the European Union under the 7th framework program (grant agreement FP7-600914). The Botnia Study has also been financially supported by grants from the Sigrid Juselius Foundation, Folkhälsan Research Foundation, Ollqvist Foundation, Signe and Ane Gyllenberg Foundation, Swedish Cultural Foundation in Finland, Finnish Diabetes Research Foundation, Foundation for Life and Health in Finland, Helsinki University Central Hospital Research Foundation, Närpes Health Care Foundation, and Ahokas Foundation.

## References

- Nalysnyk L, Hernandez-Medina M, Krishnarajah G. Glycaemic variability and complications in patients with diabetes mellitus: evidence from a systematic review of the literature. *Diabetes Obes Metab*. 2010;12(4):288-298.
- Hirsch IB. Glycemic variability and diabetes complications: does it matter? Of course it does!. *Diabetes Care*. 2015;38(8):1610-1614.
- Rizzo MR, Barbieri M, Marfella R, Paolisso G. Reduction of oxidative stress and inflammation by blunting daily acute glucose fluctuations in patients with type 2 diabetes. *Diabetes Care*. 2012;35(10):2076-2082.
- Timmons JG, Cunningham SG, Sainsbury CA, Jones GC. Inpatient glycemic variability and long-term mortality in hospitalized patients with type 2 diabetes. *J Diabetes Complications*. 2016;31(2):479-482.
- Kilpatrick ES, Rigby AS, Atkin SL. For debate. Glucose variability and diabetes complication risk: we need to know the answer. *Diabet Med*. 2010;27(8):868-871.
- Su G, Mi S, Tao H, et al. Association of glycemic variability and the presence and severity of coronary artery disease in patients with type 2 diabetes. *Cardiovasc Diabetol*. 2011;10(1):1.
- Rezki A, Chiheb S, Merioud B, Fysekidis M, Cosson E, Valensi P. Role of glycemic variability and lipids in the changes in cutaneous microcirculatory blood flow in patients with impaired glucose tolerance or type 2 diabetes. *Diabetes Metab*. 2016;42(4):295.
- Ohara M, Fukui T, Ouchi M, et al. Relationship between daily and day-to-day glycemic variability and increased oxidative stress in type 2 diabetes. *Diabetes Res Clin Pract*. 2016;122:62-70.
- Rodbard D. The challenges of measuring glycemic variability. *J Diabetes Sci Technol*. 2012;6(3):712-715.
- Le Floch JP, Kessler L. Glucose variability comparison of different indices during continuous glucose monitoring in diabetic patients. *J Diabetes Sci Technol*. 2016;10(4):885-891.
- Kovatchev B, Cobelli C. Glucose variability: timing, risk analysis, and relationship to hypoglycemia in diabetes. *Diabetes Care*. 2016;39(4):502-510.
- Lane JE, Joseph PS, Howard Z. Continuous glucose monitors: current status and future developments. *Curr Opin Endocrinol Diabetes Obes*. 2013;20(2):106-111.
- Rodbard D. Continuous glucose monitoring: a review of successes, challenges, and opportunities. *Diabetes Technol Ther*. 2016;18(S2):S2-S3.
- Rodbard D. New and improved methods to characterize glycemic variability using continuous glucose monitoring. *Diabetes Technol Ther*. 2009;11:551-565.
- Rodbard D. Interpretation of continuous glucose monitoring data: glycemic variability and quality of glycemic control. *Diabetes Technol Ther*. 2009;11(suppl 1):55-67.
- Clarke W, Kovatchev B. Statistical tools to analyze continuous glucose monitor data. *Diabetes Technol Ther*. 2009;11(suppl 1):S45-S54.
- Gorst C, Kwok CS, Aslam S, et al. Long-term glycemic variability and risk of adverse outcomes: a systematic review and meta-analysis. *Diabetes Care*. 2015;38(12):2354-2369.
- Xu W, Zhu Y, Yang X, et al. Glycemic variability is an important risk factor for cardiovascular autonomic neuropathy in newly diagnosed type 2 diabetic patients. *Int J Cardiol*. 2016;215:263-268.
- Hoffman RP, Dye AS, Huang H, Bauer JA. Glycemic variability predicts inflammation in adolescents with type 1 diabetes. *J Pediatr Endocrinol Metab*. 2016;29(10):1129-1133.
- Marling CR, Shubrook JH, Vernier SJ, Wiley MT, Schwartz FL. Characterizing blood glucose variability using new metrics with continuous glucose monitoring data. *J Diabetes Sci Technol*. 2011;5(4):871-878.
- Augstein P, Heinke P, Vogt L, et al. Q-Score: development of a new metric for continuous glucose monitoring that enables stratification of antihyperglycaemic therapies. *BMC Endocr Disord*. 2015;15(22).
- Vigersky R, Shrivastav M. Role of continuous glucose monitoring for type 2 in diabetes management and research. *J Diabetes Complications*. 2017;31(1):280-287.



23. Chon S, Lee YJ, Fraterrigo G, et al. Evaluation of glycemic variability in well-controlled type 2 diabetes mellitus. *Diabetes Technol Ther.* 2013;15(6):455-460.
24. Warnock A, Salkind S, Senicki S, et al. The effect of real-time continuous glucose monitoring (RT-CGM) in prediabetes. *Diabetes.* 2016;65(suppl 1):A221-A360.
25. Zou CC, Liang L, Hong F, Zhao ZY. Glucose metabolism disorder in obese children assessed by continuous glucose monitoring system. *World J Pediatr.* 2008;4(1):26.
26. Salkind S, Huizenga R, Fonda S, Walker S, Vigersky R. Glycemic variability in nondiabetic morbidly obese persons: results of an observational study and review of the literature. *J Diabetes Sci Technol.* 2014;8(5):1042-1047.
27. Madhu SV, Muduli SK, Avasthi R. Abnormal glycemic profiles by CGMS in obese first-degree relatives of type 2 diabetes mellitus patients. *Diabetes Technol Ther.* 2013;15(6):461-465.
28. Fabris C, Facchinetti A, Sparacino G, et al. Glucose variability indices in type 1 diabetes: parsimonious set of indices revealed by sparse principal component analysis. *Diabetes Technol Ther.* 2014;16(10):644-652.
29. Fabris C, Facchinetti A, Fico G, Sambo F, Arredondo MT, Cobelli C. Parsimonious description of glucose variability in type 2 diabetes by sparse principal component analysis. *J Diabetes Sci Technol.* 2016;10(1):119-124.
30. Wójcicki JM. "J"-index. A new proposition of the assessment of current glucose control in diabetic patients. *Horm Metab Res.* 1995;27:41-42.
31. Service FJ, Molnar GD, Rosevear JW, Ackerman E, Gatewood LC, Taylor WF. Mean amplitude of glycemic excursions, a measure of diabetic instability. *Diabetes.* 1970;19:644-655.
32. Baghurst PA. Calculating the mean amplitude of glycemic excursions from continuous glucose monitoring data: an automated algorithm. *Diabetes Technol Ther.* 2011;13:296-302.
33. Kovatchev B, Cox D, Gonder-Frederick L, Clarke W. Symmetrization of the blood glucose measurement scale and its applications. *Diabetes Care.* 1997;20:1655-1658.
34. Kovatchev B, Straume M, Cox D, Farhy L. Risk analysis of blood glucose data: a quantitative approach to optimizing the control of insulin dependent diabetes. *J Theor Med.* 2000;3:1-10.
35. Clarke W, Kovatchev B. Statistical tools to analyse continuous glucose monitor data. *Diabetes Technol Ther.* 2009;11:S45-S54.
36. Kovatchev B, Otto E, Cox D, Gonder-Frederick L, Clarke W. Evaluation of a new measure of blood glucose variability in diabetes. *Diabetes Care.* 2006;29:2433-2438.
37. Hill NR, Hindmarsh PC, Stevens RJ, Stratton MJ, Levy JC, Matthews DR. A method for assessing quality of control from glucose profiles. *Diabet Med.* 2007;24:753-758.
38. Schlichtkrull J, Munck O, Jersild M. The M-value, an index of blood-sugar control in diabetics. *Acta Med Scand.* 1965;177:95-102.
39. Friedman J, Hastie T, Tibshirani R. *The Elements of Statistical Learning.* Berlin, Germany: Springer; 2001.
40. DeVries JH. Glucose variability: where it is important and how to measure it. *Diabetes.* 2013;62(5):1405-1408.
41. Aha DW, Bankert RL. A comparative evaluation of sequential feature selection algorithms. In: Fisher D, Lenz HJ, eds. *Learning from Data.* New York, NY: Springer; 1996:199-206.