IGT and T2D Subjects Automatically Classified Using a Selection of CGM-based Glycemic Variability Indices

E. Longato[1], G. Acciaroli[1], A. Facchinetti[1], L. Hakaste[2], T. Tuomi[2], A. Maran[3], and G. Sparacino[1]

[1]Department of Information Engineering, University of Padova, Italy [2]Folkhälsan Research Centre, University of Helsinki, Finland [3]Department of Medicine, University of Padova, Italy

1. INTRODUCTION AND AIM

Glycemic Variability (GV) indices are useful for characterizing the dynamic properties of CGM-acquired glucose concentration profiles. However, there is still no consensus on how to use the plethora of indices proposed in the literature due, partly, to the high degree of correlation between them.

Recently, we demonstrated that CGM-based GV indices can be successfully used to distinguish healthy from impaired glucose tolerance (IGT) and type 2 diabetes (T2D) conditions [1]. A further distinction between IGT and T2D proved quite critical.

In the present work, we investigate whether a suitable subset of GV indices can be used, in a machine learning framework, to accurately distinguish between subjects affected by IGT and T2D.

2. DATABASE

The dataset comprises 62 subjects extracted from the Botnia Perspective Study and the Botnia PPP Study within the EU FP7 Mosaic project [2]. Each subject was monitored in both 2014 (1st visit) and 2015 (2nd visit).

	# IGT subjects	# T2D subjects
1st Visit (2014)	36	26
2nd Visit (2015)	37	25

Table 1. Cohort composition at the two visits. Subjects are divided according to their metabolic state at visit time.

In both visits, subjects were monitored for a 6-day period by the iPro CGM system (Medtronic MiniMed, Inc., Northridge, CA) at a frequency of one sample every 5 minutes.

In addition to CGM traces, age, sex, body mass index (BMI) and waist circumference (WC) were recorded for each patient at each visit.

The metabolic state of each subject was assessed via the result of an oral glucose tolerance test (OGTT), a gold standard technique for the diagnosis of IGT and T2D.

3. THE GV INDICES SUBSET

Expert-knowledge-driven feature selection was performed by a diabetologist on the basis of each index's immediate clinical significance, ease of interpretation, and redundancy.

Age, sex, BMI, and WC were also included in the final model, as they give a broad, but valuable, characterization of a subject's general metabolic state, highlighting risk factors such as central obesity.

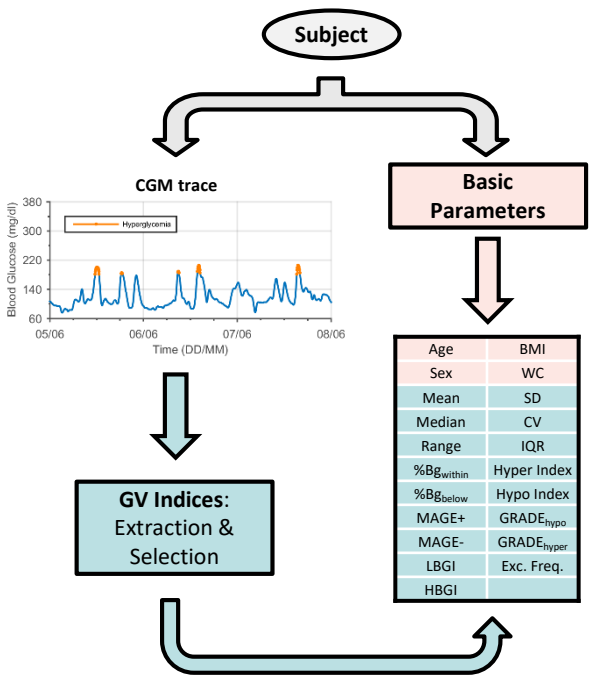

Fig. 1. Feature set building process: from raw data to the final model.

4. CLASSIFICATION STRATEGY

We implemented a linear Support Vector Machine (SVM) model to automatically classify each subject as affected by either IGT or T2D.

Hyperparameter tuning was embedded in a nested cross-validation step.

The dataset was divided in a training and a test sets: the training set comprised the data from the 1st visit, the test set comprised the data from the 2nd visit.

5. RESULTS

Classification accuracy is computed as the percentage of correctly classified subjects in the test set.

The full confusion matrix calculated on the test set is shown in the table below.

		PREDICTED DIAGNOSIS: IGT	PREDICTED DIAGNOSIS: T2D
TRUE DIAGNOSIS	IGT	29	8
TRUE DIAGNOSIS	T2D	3	22

Table 2. Test set confusion matrix. Each cell reports the number of subjects included in the test set with a given combination of predicted and true diagnosis.

Overall, **51 out of 62 subjects were correctly classified**: 8 subjects affected by IGT were instead assigned to the T2D class, whereas only 3 patients affected by full-fledged T2D were mistakenly believed to be affected by IGT.

A qualitative visual representation of our results, obtained by means of Principal Component Analysis, is also shown below, highlighting the satisfactory performance of the classifier (Fig. 2).

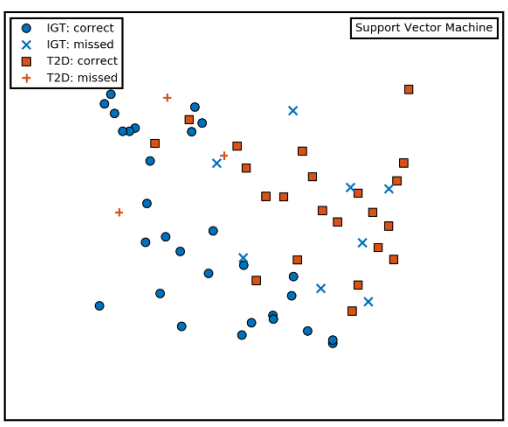

Fig. 2. PCA representation of test-set classification performance.

We achieved **82.3% accuracy** in classifying the metabolic state of a cohort of IGT and T2D subjects, using only a subset of indices trusted by an expert diabetologist and some basic parameters.

The subset of 17 GV indices and 4 basic parameters selected by the clinician improved classification accuracy by ~30% compared to the same technique applied to our starting pool of 37 GV indices and 4 basic parameters.

6. CONCLUSION

CGM-based GV indices and basic clinical parameters can be used to quite accurately distinguish the subtle differences between IGT and T2D glucose recordings, by relying only on metrics trusted by an expert clinician.

Future work will investigate the identification of the minimal subset of GV indices needed for accurate classification and the trade-off between performance and ease of interpretation.

REFERENCES:

[1] G. Acciaroli, et al., "Diabetes and Prediabetes Classification Using Glycemic Variability Indices From Continuous Glucose Monitoring Data", J Diabetes Sci Technol 2017.

[2] Available online at: http://www.mosaicproject.eu/.

CONTACT: Enrico Longato: enrico.longato.1@phd.unipd.it
Giada Acciaroli: giada.acciaroli@phd.unipd.it

ATTD Advanced Technologies & Treatments for Diabetes

DEPARTMENT OF INFORMATION ENGINEERING UNIVERSITY OF PADOVA