SAPIENZA
UNIVERSITÀ DI ROMA

# On the role of retrospective glycemic monitoring as an educational, motivational, and modifying therapy compliance tool in type 2 diabetic patients


Facoltà di Ingegneria dell'Informazione, Informatica e Statistica

Corso di Laurea Magistrale in Engineering in Computer Science

Candidate

Nicola Di Santo

ID number 1853048

Thesis Advisor

Prof. Aris Anagnostopoulos

Co-Advisor

Prof. Ioannis Chatzigiannakis



Academic Year 2020/2021

Thesis defended on January 2021
in front of a Board of Examiners composed by:

- (chairman)

---

**On the role of retrospective glycemic monitoring as an educational, motivational, and modifying therapy compliance tool in type 2 diabetic patients**
Master's thesis. Sapienza – University of Rome

© 2021 Nicola Di Santo. All rights reserved

This thesis has been typeset by LaTeX and the Sapthesis class.

Version: December 28, 2020

Author's email: nicola96ds@gmail.com

*Dedicated to my family,*
*thank you for everything.*

# Abstract

***Background*:** In recent years, a promising new technology has become available: the continuous blood glucose monitoring (CGM). It uses subcutaneous devices that detect and record for several consecutive days the values of the glycemic values of the interstitial fluid, in constant equilibrium with the plasma compartment. The role of retrospective glycemic monitoring as an educational, motivational, and modifying therapy compliance tool in type 2 diabetic patients has not been evaluated yet, although its potential in this direction has been understood

***Methods:*** This project aims to verify the effectiveness of continuous retrospective monitoring of glycemia as an educational intervention to improve patient adherence to ongoing therapy and the degree of glyco-metabolic compensation, in a population of type 2 diabetics with complex management (obese patients on multi-injection insulin therapy and in metabolic decompensation, despite maximal insulin therapy) who are not eligible for insulin pump therapy nor for monitoring real-time continuous blood glucose. We first provide a data visual exploration and analysis, then some data mining techniques are applied to process data and extract meaningful results.

***Results*** We show that patients under CGM feedback have a good response to the training using visual analytics. Then we cluster subjects using automated methods and an age-based criterion to show different responses to the training for distinct groups of patients. Then we have analyzed glycemic-event distribution during part of the days and we find out that hyperglycemic events are significantly frequent between the morning and the late morning, while at night they are much less frequent. We have applied the Sliding Window forecasting methodology to CGM data obtaining similar performances to the same technique applied to T1DM CGM data but with a more variable error distribution. Then a novelty method has been defined that achieves more than 90% into CGM signal forecasting.

***Conclusions*** Using continuous retrospective monitoring of glycemia as an educational intervention on type 2 diabetic patients with complex management effectively improves the adherence to ongoing therapy. Moreover, with the defined forecasting technique, it is possible to give feedback to the patient on the future levels of blood glucose to predict and avoid serious hypo/hyperglycemic events.

## Acknowledgments

*I would like to thank my advisor Prof. Aris Anagnostopoulos and co-advisor Prof. Ioannis Chatzigiannakis and the research team of Policlinico Umberto I lead by Professor Marianna Maranghi without which this work would not have been possible.*

# Contents

# Chapter 1

# Introduction

In recent years, a promising new technology has become available: continuous blood glucose monitoring (CGM). It is achieved by employing the simple installation of subcutaneous devices that detect and record for several consecutive days the values of the glycemic values of the interstitial fluid, in constant equilibrium with the plasma compartment. The continuous monitoring of blood glucose is indicated by guidelines, mainly in the management of type 1 diabetic patients for the largest number of studies performed in this population; in adult diabetic patients (also type 2) it is indicated to highlight any nocturnal or unnoticed hypoglycemia and when there are significant therapeutic changes [20].

To date, studies on the use of CGM in type 2 diabetic are few and have used this technology as: (1) a means of verifying the effectiveness of specific changes in hypoglycemic therapy in terms of reduction of daily glycemic variability [29, 18, 28, 27, 47], (2) in the type 2 diabetic population on multi-injection insulin therapy, the superiority of CGM respect to home glycemic self-monitoring (SMBG) has been proved in the detection of hypoglycemia [50].

However, the role of retrospective glycemic monitoring as an educational, motivational, and modifying therapy compliance tool in type 2 diabetic patients has not been evaluated yet, although its potential in this direction has been understood [4].

The aim of this study is therefore to verify the effectiveness of continuous retrospective monitoring of glycemia as an educational intervention to improve patient adherence to ongoing therapy and the degree of glyco-metabolic compensation, in a population of type 2 diabetics with complex management - obese patients on multi-injection insulin therapy and in metabolic decompensation despite maximal insulin therapy- who are not eligible for insulin pump therapy nor for monitoring real-time continuous blood glucose.

# Chapter 2

# Background

In this chapter, it is described the diabetes disease in general and all its most common forms. The self-monitoring approach is also discussed, together with related technologies. Then it is examined the importance of blood glucose forecasting and current state of the art techniques.

## 2.1 Diabetes

Commonly the term diabetes is used to indicate a chronic disease, which can be classified in the group of diseases known as diabetes mellitus, characterized by a high concentration of glucose in the blood, which is in turn caused by an (absolute or relative) deficiency of insulin in the human organism, a hormone that by stimulating the uptake of glucose in muscle and fat cells decreases its concentration in the blood. People with diabetes are also at increased risk of other diseases including heart, peripheral arterial and cerebrovascular disease, obesity, cataracts, erectile dysfunction, and nonalcoholic fatty liver disease. They are also at increased risk of some infectious diseases, such as tuberculosis.[32]

Two different types of diabetes are widely known although, according to the classification of the World Health Organization (WHO), new hybrid types of diabetes have recently been classified[32]. The two macro-classes of diabetes are type 1 diabetes mellitus (T1DM) and type 2 diabetes mellitus(T2DM). The latter is the most widespread and subtle pathology. In fact, this pathology can occur in the absence of symptoms and lead to a delayed diagnosis and complications. It is estimated that between 30 and 80 percent of cases are undiagnosed [51].

In Italy, the diabetes treatment absorbs 6.65% of the overall health expenditure, with a cost per patient that is more than double the national average [10]. Given the considerable burden that diabetes entails for public health, prevention and Improving the care of people with this disease should be a primary goal for most communities and health systems.

### 2.1.1 Type 1 Diabetes Mellitus

Type 1 diabetes mellitus (T1DM) is an autoimmune disease, that destruct pancreatic $\beta$ cells, and that usually involves association with lack of insulin. In about half of

the cases, it starts at the age of 20 and affects 3% of the world population. [32]. The causes depend on several factors and, in general, involve an immunological stimulus joint with some genetic predisposition, that slowly start destroying the beta cells. When more than 80% of the beta cells are destroyed, patients start facing form 1 diabetes mellitus. Symptoms usually develop quickly and include polyuria, polydipsia (secondary to polyuria), paradoxical polyphagia, and weight loss as well as some less frequent symptoms as blurred vision, feeling tired, and poor healing [32].

Some research shows how the difficulties encountered by people suffering from a chronic disease such as type 1 diabetes mellitus affect the patient's psychological and emotional level, sometimes impacting the management of the disease itself. When the sick person is overwhelmed by the responsibilities, periodic checks, and restrictions (mostly food) that the disease itself entails, the likelihood that psychological disorders such as depression and anxiety will also develop simultaneously with the chronic disease. [42, 32].

### 2.1.2 Type 2 Diabetes Mellitus

Type 2 diabetes mellitus (T2DM) accounts for between 90% and 95% of diabetes, with the highest proportions in low- and middle-income countries. It is a disease that mainly affects adults although adolescents can also be involved, and is mainly caused by an unhealthy lifestyle with a prevalence of sedentary factors and consumption of super-processed foods, largely due to lifestyle changes of the last century [30, 32].

Also in T2DM a $\beta$-cell dysfunction is observed. Most of the T2DM population is obese or has a very high BMI, with a prevalence of fat in the abdominal area. Patients with this condition usually develop an increase in insulin and develop resistance to it[45, 7, 32]. For most people with T2DM, insulin treatment is not required for survival but may be required to lower blood glucose and avoid complications. T2DM often remains undiagnosed for many years because the hyperglycemia is not severe enough to provoke noticeable symptoms of diabetes [15, 32]. Many factors increase the risk of developing T2DM including age, obesity, and unhealthy lifestyles. The causes, especially the genetic ones, have not yet been clearly determined [32].

## 2.2 Diabetes Self Management

In recent years it was developed the idea that, in order to reduce costs and to reduce the stress of patients associated with a multitude of visits in a short time, it is necessary that subjects with diabetes be educated to self-monitor the progress of the disease and take action consequently[31, 11]. The two most popular monitoring techniques are self-monitoring of blood glucose (SMBG) and continuous glucose monitoring (CGM). However, it has been observed that such an approach could be overwhelming for patients[11].

### 2.2.1 Self Monitoring Blood Glucose

Self-monitoring blood glucose (SMBG) is one of the various self-monitoring procedures for diabetic patients and belongs to the larger self-management group for

diabetic patients. It consists of a reader, called "glucose meter", which allows the patient to independently monitor the level of glucose in the blood. In order to make self-measurement effective, patients must be properly trained by medical personnel, in order to make the measurements as accurate and error-free as possible. The guidelines for this type of monitoring were released in the early '90s by the American Diabetes Association (ADA) and are now widely used in the treatment of patients with diabetes. The measurements allow the patient to adjust their therapy, eating habits, exercise etc. based on the surveys carried out. However, it has been found that patients may be overwhelmed by this practice, and is sometimes identified as one of the stressors that lead the patient to develop neurological conditions such as depression due to the overhead of managing the disease [3, 1, 2]. To fully benefit from self-monitoring, it is essential that patients are trained by a health professional.

### 2.2.2 Continuous glucose monitoring

Continuous glucose monitoring (CGM) is a technique for continuous monitoring (24/7) of blood glucose levels, as the name suggests. The technique has been developing since the early 2000s and, to date, has incredible improvements in terms of measurement compared to the first versions. With the CGM it is foreseen the insertion of a small subcutaneous sensor with a lifetime of about 15 days, through which it is possible to obtain continuous monitoring. In the most recent versions of the technology, it is possible to connect the sensor to your smartphone through short-range and secure communication protocols such as NFC or Bluetooth for real-time analysis. Some devices are connected to an insulin pump and inject it if needed. In some patients in whom real-time monitoring is not applicable due to some complicating factors, it is possible to use the data obtained through CGM to carry out a retrospective analysis of the patient's habits (diet, physical activity, etc.) in order to educate him to a healthier lifestyle that allows him to live together and better manage the disease.

There are several devices on the market but there is no standard that allows this type of monitoring to be standardized under a single framework. It is also challenging for non-medical personnel to interface with the data produced by this device and to be able to give it the right interpretation. These factors, together with the cost factors and the low life of the sensors, have made the technology not widespread even if lately the trend seems to be reversing[37].

## 2.3 Blood Glucose Forecasting

With the advent of continuous and real-time monitoring of blood glucose levels, a key aspect of research has become the possibility of predicting future hypo/hyperglycemic events in order to be able to prevent them with specific countermeasures. Several techniques has been evaluated in the literature, all with the objective of predicting exact sensors outcome within a given predictive horizon. The majority of techniques are short-range predictors with a horizon of less than or equals to 60 min. Those are based on Deep Artificial Neural Networks (DANN) [25], supervised learning techniques [33] or classical machine learning approach with algorithms like random forest and features selected with the help of expert professionals [23]. Some attempts

to broaden the predictive horizon also exist [12] as well as innovative forecasters based on a predictive sliding window and fixed horizon[38].

# Chapter 3

# Methodologies

In this chapter all the methodologies used to obtain, visualize, process, and analyze data are presented.

## 3.1 CGM Experiment Setup

The study involves a total of 30 subjects. Patients are recruited within three months at the Diabetology Outpatient Clinics of the Department of Internal Medicine and Medical Specialties of Policlinico Umberto I in Rome: 15 subjects assigned to the CGM group and 15 to the SMBG group.

During the study will be carried out educational meetings scheduled according to the indications of the current guidelines in diabetes treatment. The meetings will be organized in small groups and will be aimed at the clarification of aspects related to the disease and clinical management of the disease.

During examination 1 (V1), which will be carried out 2 weeks after the screening visit (V0), patients can be assigned to the CGM (continuous glucose monitoring) group or self-monitoring blood glucose (SMBG) group. Training will be carried out dedicated to the use of the glucose meter by the nursing staff involved in the study; all subjects enrolled in the study will be given the same type of glucometer (Bayer's CONTOUR® LINK meter[1], Medtronic[2]) and will be provided with the necessary equipment for glycemic detection. The initial setting procedures of the glucometer will be reviewed with the trial participants, together with its main functions, to ensure the performance of the Glycemic measurement is in standardized conditions.

Both groups (SMBG/CGM) will have a subcutaneous sensor applied to the abdominal region for the continuous detection of interstitial glycemia and will be subsequently connected to a data recorder (iProTM2, MMT-7741, Medtronic)[3]. The patient will be instructed on how to monitor capillary glycemia.

At each removal visit of the continuous blood glucose monitoring system, for both treatments, a short comment/training session with the patient is scheduled to relate the blood glucose trend to the dietary and physical activity diary. In the SMBG group, the results of the SMBG capillary blood glucose monitoring will be

[1] https://www.diabetes.ascensia.it/

[2] https://www.medtronic.com/

[3] https://hcp.medtronic-diabetes.com.au/sites/default/files/ipro2_step_by_step_guide_aug18_final.pdf

| Column | Description |
|---|---|
| Date | Date of measured value (yyyy-mm-dd) |
| Time | Time of the measure (hh:m:ss) |
| Timestamp | Date time field (yyyy-mm-dd hh:m:ss) |
| ISIG Value | Raw electrical output of the glucose sensor (in nA) |
| Sensor Glucose (mg/dL) | The glucose sensor value in mg/dl |
| Raw-Values | Data of the current table row in non tabular form (csv) |
| Patient | Identifier for the subject those measures belong to |
| Visit | Identifier for the visit those measures refer to |
| Cgm | A flag specifying the group of the subject |

**Table 3.1.** The final raw sensor data-set fields description.

commented on, while in the CGM group both the SMBG glycemic self-monitoring results and the results of the generated reports from CGM will be commented on.

After the procedure of monitoring and education between visits is completed, the collected sensors' data are post-processed to extrapolate some useful insights and evidence difference and similarity between the two groups as well as the importance of the educational process for patients.

### 3.1.1 Dataset

For this study, a data-set with results obtained from both sensor monitoring and clinical screening has been built and will drive some of the analysis in this document together with the raw sensors values dataset. The former is composed of 88 columns and an accurate description can be found in Appendix A. It is made of 28 rows, one for each patient. The low number of subjects involved in this study results is a small data-set ($28 \times 88$). Even if we can enrich it with other features derived from sensors measures statistics, we reach 133 features that still result in a small-sized data-set. From here the choice of driving our research using raw sensors data. Initially, those data are divided into a file per visit for each patient, resulting in $28 \times 3$ files. Each file contains a timestamp column, the ISIG Value[4] column, and the Sensor Glucose (mg/dl) column. To facilitate processing those data are put all in the same table and columns to differentiate between patients, visits, and groups are added. The final raw sensor data-set results in 130707 rows, each representing values from sensor sampled with a frequency of 5 min. A schematic description in Table 3.1

### 3.1.2 Population

The selected patients for this study are type 2 diabetic subjects with at least 5 years of diabetes. Subject are aged between 40 and 75 years, in multi-injection insulin therapy (at least 3 insulin administrations per day, of which at least 1 is basal insulin) ± metformin (2000 mg/day), with an insulin requirement greater than or equal to at least 0.3-0.5 IU/kg, not eligible for insulin pump therapy and continuous

[4]It represents the raw electrical output of the glucose sensor (in nA)

real-time glycemic monitoring, in glycemic decompensation for at least 6 months (HbA1c $\geq$ 7.5%), with a body mass index between 25 and 40 $kg/m^2$.

In Table 3.2, population is divided into four groups. Division is dictated by age criterion: first age groups has 6 patients and age interval: 48-60, second has 8 patients and age interval: 61-64, third has 5 patients and age interval: 65-66, last one has 9 patients and age interval: 67-78. Age ranges are defined by 0.2, 0.5, 0.7 quantiles over age distribution in the population. Those age groups will be considered in all further analyses.

| | Age group 1 | Age Group 2 | Age Group 3 | Age Group 4 |
|---|---|---|---|---|
| Male % | 83.33 | 62.5 | 40 | 44.44 |
| Female % | 16.67 | 37.5 | 60 | 55.56 |
| CGM % | 50 | 62.5 | 40 | 55.56 |
| SMBG % | 50 | 37.5 | 60 | 44.44 |
| Smoking % | 33.33 | 25 | 0 | 11.11 |
| Ex Smoking % | 50 | 62.5 | 60 | 44.44 |
| No Smoking % | 16.67 | 12.5 | 40 | 44.44 |
| Height (mt.) | 1.73 (1.69 - 1.78 ) | 1.66 (1.58 - 1.69 ) | 1.59 (1.55 - 1.72 ) | 1.62 (1.58 - 1.68 ) |
| Insulin Therapy Duration (months) | 27.0 (24.0 - 61.5 ) | 38.0 (22.0 - 69.0 ) | 108.0 (36.0 - 132.0 ) | 36.0 (20.0 - 108.0 ) |
| Basal Insulin Unit | 28.0 (23.5 - 32.5 ) | 19.0 (15.5 - 29.0 ) | 28.0 (28.0 - 30.0 ) | 30.0 (22.0 - 44.0 ) |
| Number Of Administrations Per Day | 4.0 (4.0 - 4.0 ) | 4.0 (4.0 - 4.0 ) | 4.0 (4.0 - 4.0 ) | 4.0 (4.0 - 4.0 ) |
| Weight V0 (Kg) | 88.5 (81.5 - 94.15 ) | 92.25 (84.75 - 97.5 ) | 93.9 (82.0 - 98.0 ) | 83.0 (69.4 - 93.0 ) |
| Weight V2 (Kg) | 86.0 (78.35 - 92.88 ) | 89.65 (85.25 - 95.0 ) | 93.0 (82.7 - 95.5 ) | 82.95 (69.5 - 87.98 ) |
| BMI V0 ($kg/m^2$) | 29.46 (28.31 - 31.04 ) | 34.64 (29.07 - 39.31 ) | 36.95 (32.09 - 38.76 ) | 31.25 (29.41 - 32.63 ) |
| BMI V2 ($kg/m^2$) | 28.71 (7.0 - 29.7 ) | 34.28 (28.88 - 37.97 ) | 37.52 (32.88 - 37.78 ) | 31.28 (29.61 - 32.38 ) |

**Table 3.2.** Table describing statistics of age groups defined in the population section. Data are presented as percentage if % symbol is next to the statistic or as IQR (interquartile ranges) otherwise

| | Visit 1<br>Screening | Visit 2<br>First followup | Visit 3<br>Second followup |
|---|---|---|---|
| Subjects | | | |
| Number[5] | 28 | 25 | 25 |
| Diabetes duration (years) | 16.00 (11.75-23.00) | 16.00 (11.75-23.00) | 16.00 (11.75-23.00) |
| Age | 64.50 (61.00-68.00) | 64.50 (61.00-68.00) | 64.50 (61.00-68.00) |
| Sex[6] | 1.00 (0.00-1.00) | 1.00 (0.00-1.00) | 1.00 (0.00-1.00) |
| BMI | 31.79 (29.12-36.48) | -[7] | 31.35 (28.72-36.49) |
| Blood Glucose Data | | | |
| Mean Glucose (mg/dL) | 174.43 (153.57-194.13) | 189.27 (162.12-216.86) | 183.81 (162.85,201.89) |
| SD (mg/dL) | 59.33 (34.44-57.75) | 61.88 (40.65-64.06) | 54.92(36.4,57.89) |
| Coefficient of variation (%) | 34.01 (22.42-29.75) | 32.69 (25.08-29.54) | 29.88 (22.35-28.67) |
| CGM metrics | | | |
| Mean Glucose (mg/dL) | 170.34 (154.9-186.35) | 179.86 (142.61-192.79) | 180.34(150.11,206.11) |
| SD (mg/dL) | 52.69 (38.98-46.68) | 58.37 (34.71-52.61) | 61.96 (40.46-52.99) |
| Coefficient of variation (%) | 30.93 (25.16-25.05) | 32.45 (24.34-27.29) | 34.36 (26.95-25.71) |
| GMI (%) | 7.38 (7.02-7.77) | 7.61 (6.72-7.92) | 7.62(6.9,8.24) |
| TIR (70–180 mg/dL) (%) | 59.78 (42.54-72.46) | 53.12 (37.77-70.08) | 53.03 (32.43-74.24) |
| TAR(> 180 mg/dL) (%) | 39.53 (25.95-55.1) | 43.52 (26.02-62.03) | 46.26 (24.3-67.57) |
| Level 1 [181–250 mg/dL] (%) | 84.25 (76.45-96.22) | 73.6 (59.65-91.2) | 79.45 (68.22-100.0) |
| Level 2 [>250 mg/dL] (%) | 16.22 (4.43-24.52) | 22.78 (8.63-39.91) | 21.02 (0.0-32.14) |
| TBR(<70 mg/dL) (%) | 0.69 (0.0-1.47) | 3.35 (0.0-0.82) | 0.71 (0.0-0.71) |
| Level 1 [54–69 mg/dL] (%) | 30.96 (0.0-62.5) | 26.21 (0.0-40.0) | 20.89 (0.0-48.39) |
| Level 2 [<54 mg/dL] (%) | 9.04 (0.0-0.0) | 13.79 (0.0-21.74) | 11.11 (0.0-4.35) |
| Insulin | | | |
| Insulin therapy duration (month) | 36.00 (23.50 - 108.00) | 36.00 (23.50 - 108.00) | 36.00 (23.50 - 108.00) |
| Basal insulin unit | 28.00 (20.00 - 34.50) | 28.00 (20.00 - 34.50) | 28.00 (20.00 - 34.50) |
| Rapid insulin unit | 32.00 (21.50 - 42.00) | 32.00 (21.50 - 42.00) | 32.00 (21.50 - 42.00) |
| HbA1 | 8.65 (7.90 - 9.50) | | |
| Glycemic Variability Indexes | | | |
| J Index | 41.16 (31.51 - 55.73) | 45.69 (33.13 - 63.55) | 44.00 (33.67 - 63.25) |
| MAGE | 170.92 (152.03 - 202.90) | 185.35 (156.80 - 217.68) | 181.60 (158.34 - 210.04) |
| GVI (%) | 14.60 (9.44 - 20.58) | 13.90 (10.10 - 19.71) | 14.08 (9.26 - 20.52) |
| LGI | 0.00 (0.00 - 0.12) | 0.00 (0.00 - 0.11) | 0.00 (0.00 - 0.12) |
| HGI | 6.74 (3.76 - 11.65) | 8.30 (4.12 - 14.29) | 7.47 (4.17 - 14.47) |

**Table 3.3.** Data on glycemic control in the three time intervals. Data are presented as median and interquartile ranges. Beyond classical statistical measure like mean and standard deviation (SD), we present BMI (Body Mass Index), GMI (median glucose management indicator), TIR (the time spent in range indicated in the table), TAR (time above the range), TBR (time below the range). Also, glycemic variability indexes are presented, they will be explained in a later section. They are: J Index (denoting glycemic control qualities), MAGE (mean amplitude of glycemic excursion), GVI (glucose variability index expressed as percentage), LGI, and HGI (Low and High Glucose Index) denoting risk of hypo/hyperglycemic events.

## 3.2 Visualization and Analysis

Our analysis will focus on sensors measured values, namely blood glucose measured in milligrams over deciliter (mg/dL). We will focus on three different periods, called V0,

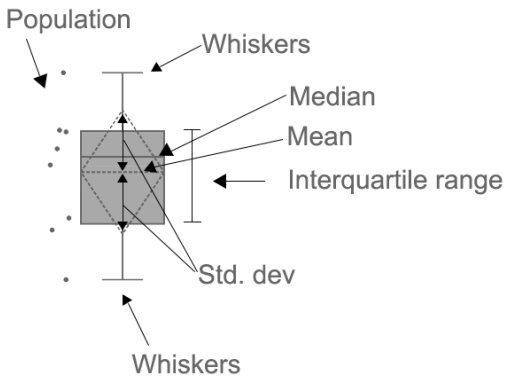



**Figure 3.1.** Example of Boxplot used in this report. Dashed diamond stands for mean ± std deviation, box represent median and interquartile ranges, Whiskers represent min-max, dots on the left side are points corresponding to the actual population.

V1, V2, as described in the previous section and we will try to prove the effectiveness of the educational intervention over the subjects after V1. As a first step, we are going to examine those data visually and we will try to find patterns or insights that confirm our goal.

To carry out the task two different approaches have been taken to show the results: The first approach we analyze the overall distribution of measured values as well as the percentage of time the patient is in the range between 70-180 mg/dL (TIR); above range (TAR) with values greater than 180 mg/dL, above range with low severity (TAR Level 1) with values between 181-250 mg/dL and above range with high severity (TAR Level 2) with values greater than 250 mg/dL. In the second approach, the population is divided into groups using a clustering algorithm (k-means++) to point out different responses to the educational intervention in terms of measured blood glucose distribution, focusing on groups, using the same time indicators of the first approach.

### 3.2.1 Introducing Boxplots

In this thesis, box-plots are used heavily but with some overload of information on the chart with respect to the usual box-plot. In detail in each box-plot, the rectangle-shaped box will represent the interquartile range, i.e. values going from the 25th percentile to the 75th percentile. Median is represented as a solid line inside this box. Lines extending from the boxes (whiskers) indicate variability outside the upper and lower quartiles. In addition to this classical representation, we add also the mean and standard deviation of the represented population in the form of a diamond-shaped dashed line. On the left of some box-plot, also the points representing the population are drawn to have a clearer idea of the underlying distribution and, in this specific analysis, to better understand the subject behaviors. See Figure 3.1 for an example.

## 3.3 GVI

In this section, we describe a series of analytical measures to characterize glycemic variability. The methodology is driven by the understanding that fluctuations in the blood glucose (BG) curve are a continuous process over time, $BG(t)$. Each point of this process is characterized by its value (blood glucose level) and its rate/direction of change in blood glucose. The CGM presents the $BG(t)$ process as a discrete-time series $BG(t_n), n = 1, 2, ...$ that approximates $BG(t)$ in steps determined by the resolution of the particular device (for example, a new value displayed every 5 min). It is important to remember that traditional statistics do not work well with CGM data because consecutive CGM readings are highly interdependent. Our framework for characterizing glycemic variability is based both on traditional measures, such as standard deviation (std) and coefficient of variation (cv), and on some specific glycemic indexes such as J-Index [40], mean of glycemic excursions (MAGE) [41], glucose variability index (GVI) [36], low and high glucose indices [22].

### 3.3.1 Standard Generic Statistical Indexes

The arithmetic mean is the most commonly used type of average and the one to which, with the term "average", is generally referred to in common speech. It is used to summarize with a single number a set of data on a measurable phenomenon. It is calculated by adding all the available values and dividing the result by the total number of data. The arithmetic mean formula for n elements is

$$\mu = \frac{1}{n}\sum_{i=1}^{n}(x_i) \tag{3.1}$$

The standard deviation is a statistical dispersion index, which is an estimate of the variability of a data population or a random variable. It is one of the ways to express the dispersion of data around a position index, which can be, for example, the arithmetic mean ($\mu$) or an estimate of it. It therefore has the same unit of measurement as the observed values. In statistics, precision can be expressed as the standard deviation. It is computed as:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2} \tag{3.2}$$

The coefficient of variation (CV) is defined as the ratio between the standard deviation and the mean, showing the extent of variability in relation to the population mean. The coefficient of variation should only be calculated for data measured on a ratio scale, i.e. scales that have a significant zero and thus allow relative comparison of two measurements. Its formula is:

$$cv = \frac{\sigma}{\mu} \tag{3.3}$$

### 3.3.2 Blood Glucose Variability Specific Indexes

With the advent of CGM, mathematical indices able to estimate and quantify the variability of glucose in the blood are increasingly useful. Among these, we find

J-index, whose mathematical formulation is specially optimized to highlight the average presence of glucose and its variability[49]. Its formulation has been confirmed and refined through numerous clinical tests. The J-index is considered ideal with values less than 20, good less than 30, inadequate if greater than 40 [14]. It can be calculated as follows

$$J = 0.001 \times (\mu + \sigma)^2 \tag{3.4}$$

Mean Amplitude Of Glycemic Excursions (MAGE) is one of the most famous and accredited indexes of glycemic variability. the MAGE is calculated by averaging the value of all significant glycemic excursions, ie those that are large enough with respect to the metric proposed at the same time as the index evaluation[41]. Defining, therefore $BG$ be the set of BG measures derived from CGM. We can define a tolerance range to glycemic variability and we consider relevant all those values outside this range. Let $\alpha$ be a constant, then the upper limit will be

$$up = \mu + \alpha \times \sigma \tag{3.5}$$

while the lower limit will be

$$dwn = \mu - \alpha \times \sigma \tag{3.6}$$

Then we we define with as $\hat{BG}$ all those values that are not in this tolerance range: $\hat{BG} = \{BG_i : BG_i \in BG \; and \; BG_i > up \; or \; BG_i < dwn\}$, then mage can be simply computed as

$$MAGE = \frac{1}{n} \sum_{i=1}^{n} (\hat{BG_i}) \tag{3.7}$$

with $\hat{BG_i}$ indicating all the available data point within CGM data.

Through the Glucose Variability Index (GVI) index it is possible to relate the length of the line drawn by the measurements of the CGM sensor using the trigonometric analysis. In particular, the measurement is based on the observation that the length of the line drawn by a variable signal will be greater as the variability increases. The size is normalized by comparing it to the length of a flat line in the same time interval under examination, which represents the shortest possible length (ie. Given the least possible variability, due to a constant trend of the signal) and therefore the optimal of stationarity. This normalization makes the index invariant over time and therefore can be applied to different time intervals without losing accuracy. We call $L$ the length of curve made by sensor measurements, $L_0$ length of a flat line between the considered time instants. We express the measure as a percentage:

$$GVI = (\frac{L}{L_0} - 1) \times 100 \tag{3.8}$$

The last two indexes used in this work to examine glycemic variability are the Low Glucose index (LGI) and the High Glucose Index (HGI). Measurements from the CGM signal are asymmetrical. This implies that the range of glucose values in the blood for hypoglycemia is much less wide than that for hyperglycemia. Consequently, the indices explained so far could erroneously indicate hypoglycemias as low risk or in any case at lower risk than hyperglycemias due to this asymmetry. The LGI and HGI indices exploit the transformation of the signal into a symmetrical space

to quantify the risk of variability, trying to eliminate any type of bias determined by the asymmetry of the original space of the CGM data [22, 21]. Let $BG$ be the blood glucose values from sensors then, a non linear transformation called symmetrization is defined as

$$f(BG) = 1.509 \times [(ln(BG))^{1.084} - 5.381] \tag{3.9}$$

Then we compute the risk function $r(BG)$ :

$$r(BG) = 10 \times f(BG)^2 \tag{3.10}$$

and we separate its left part and its right part

$$rl(BG) = r(BG),\ if \quad f(BG) < 0\ and\ 0\ otherwise \tag{3.11}$$

$$rh(BG) = r(BG),\ if \quad f(BG) > 0\ and\ 0\ otherwise \tag{3.12}$$

Let $BG_1, BG_2, ...BG_n$ a series of CGM readings, composing the whole signal, then Low Glucose Index is computed as:

$$LGI = \frac{1}{n}\sum_{i=1}^{n}(rl(BG_i)) \tag{3.13}$$

while the High Glucose Index is computed as:

$$HGI = \frac{1}{n}\sum_{i=1}^{n}(rh(BG_i)) \tag{3.14}$$

## 3.4 Clustering

In statistics, clustering or group analysis is a set of multivariate data analysis techniques aimed at selecting and grouping homogeneous elements in a data set. Clustering techniques are based on measures relating to the similarity between elements. In many approaches this similarity, (or dissimilarity, they are one the inverse of the other) is conceived in terms of distance in a multidimensional space. The goodness of the analyzes obtained by the clustering algorithms depends a lot on the choice of the metric, and therefore on how the distance is calculated. Clustering algorithms group elements on the basis of their mutual distance, and therefore whether or not they belong to a set depends on how far the element under consideration is from the set itself. Two different strategies for clustering exists[24]:

- Agglomerative clustering, which consists of starting by assigning each point to a different cluster and then merge clusters together based on the notion of closeness.
- Clustering based on point assignment, where points are assigned to the cluster they fit better.

Different definition of closeness exists depending on the space used to formalize the problem, in this work it is assumed that the space in euclidean and the notion of closeness is defined according to Euclidean distance in this space.

### 3.4.1 K-Means Algorithm

The most popular algorithm of the point-assignment family is k-means[26]. K-means algorithms are applied in an euclidean space and assume the number of clusters k is given as input to the algorithm. The procedure is very simple, the algorithm takes as input all the points, select k initial centers based on a strategy, and then iteratively assigns remaining points to the closest center, adjusting at each iteration the centers of the cluster based on the new points added to the clusters. The strategy used in this work to select the k initial center to the k-means procedure is called k-means++[5]. This initialization strategy selects the first center at random and then all the subsequent centers are selected with probability proportional to its squared distance from already existing cluster centers. This procedure is based on the fact that k-means is an NP-hard problem and all the approximated solutions can be arbitrarily bad compared to the optimal solution. With k-means++ we are sure that the solution is $O(\log k)$ competitive with the optimal one [5]. To understand the right value of k also different strategies exist. A very frequent scenario while using k-means clustering is to start without knowing the right value of k and then consider it as a hyper-parameter to be tuned. A well-known heuristic used to estimate the right k is to run the algorithm for an increasing number of k and then plot the average diameters of the clusters for each run. Eventually, we will find two values of k with a very little decrease in the average diameter, and we restrict our search interval between those two new values until we find the right value of k [24].

### 3.4.2 Feature Selection

The term feature selection means a series of procedures aimed at eliminating non-relevant features from the data being analyzed. Eliminating non-relevant data allows us to reduce the dimensionality of the data in question and allows us to have a more accurate model, which does not learn from features that are not relevant for the purpose of the clustering. To remove not relevant features from our analysis we have employed a standard statistical independence test, called Pearson chi-squared test.

In statistics, the Pearson's chi-squared[19] test is used to determine the independence of a variable. Applied to feature selection for clustering, it can be used to determine the dependency of a feature on the output of the clustering algorithm. The lower the dependency between feature and output the greater the probability that this feature will not be selected for clustering purposes.

### 3.4.3 Performance Evaluation

To evaluate the goodness of the clustering result there are several metrics. Some require that the ground truth labels are known, others are more generic are not based on knowledge of ground truth labeling. In this work three different scores are used: Silhouette Coefficient, Calinski-Harabasz Index, and Davies-Bouldin Index. These indexes are all suitable for clustering where ground truth labels are not known and they measure clustering performances based on the notion that a good clustering should produce groups that are dense and well separated. In detail:

- *Silhouette Coefficient*[39]: is composed of two different outcome **a** and **b**. **a**

is the mean distance between a sample and all other points in the same class while **b** is the mean distance between a sample and all other points in the next nearest cluster. It is computed as

$$s_1 = \frac{b - a}{max(a, b)} \tag{3.15}$$

The score ranges between -1 for incorrect clustering and +1 for highly dense clustering.

- *Calinski-Harabasz Index*[8]: we call dispersion the sum of squared distances between points in the clusters. The index is then computed as the ratio of the sum of between-clusters dispersion and of inter-cluster dispersion for all clusters. For a set of data $E$ of size $n_E$ , clustered into $k$ clusters, the Calinski-Harabasz score $s$ is defined as the ratio of the between-clusters dispersion mean and the within-cluster dispersion: :

$$s_2 = \frac{tr(B_k)}{tr(W_k)} \times \frac{n_E - k}{k - 1} \tag{3.16}$$

where $tr(B_k)$ is trace of the between group dispersion matrix and $tr(W_k)$ is the trace of the within-cluster dispersion matrix defined as:

$$W_k = \sum_{q=1}^{k} \sum_{x \in C_q} (x - c_q)(x - c_q)^T \tag{3.17}$$

$$B_k = \sum_{q=1}^{k} n_q (c_q - c_E)(c_q - c_E)^T \tag{3.18}$$

- *Davies-Bouldin Index*[9] defines the average similarity between each cluster and its most similar one. It has zero as the lowest possible values and it indicates a better partition. As value goes away from zero they indicate worse clusters. Here similarity is defined as a measure $R_{ij}$ that trades off: $s_i$ and $d_{ij}$, that are respectively the average distance between each point of cluster $i$ and the diameter of that cluster. We can compute similarity $R_{ij}$ as:

$$R_{ij} = \frac{s_i + s_j}{d_{ij}} \tag{3.19}$$

and then Davies-Bouldin index is defined as:

$$s_3 = \frac{1}{k} \sum_{i=1}^{k} max_{i \neq j} R_{ij} \tag{3.20}$$

## 3.5 Time Series Analysis

In descriptive statistics, a historical (or temporal) series is defined as a set of random variables ordered over time and expresses the dynamics of a certain phenomenon over time. The historical series are studied both to interpret a phenomenon, identifying components of trend, cyclicality, seasonality, and/or accidentality, and to predict its future trend (forecasting).

### 3.5.1 Sliding Window Forecasting

The sliding window forecasting methodology has been defined in [38] and is defined specifically in the context of T1DM CGM data forecasting. The method aims to provide a framework based exclusively on data collected by sensors (CGM) in order to avoid contamination of the dataset with spurious and error-prone data, as often happens in the creation of datasets for patients with diabetes, in where most of the measurements are hand-written by medical personnel or are personal descriptions of the patient's habits. Indeed they compare different regression models and discuss the effectiveness of those forecasting techniques on T1DM patients CGM data. The methodology defines a *past sliding window (PSW)* that collects training samples from 3 to 36h hours and it controls the volume on data used from the model. It also introduce the notion of *predictive horizon (PH)* telling how far in the future the model should predict (values used in[38] are 15, 30, 45, and 60 minutes in the future). The model is called a sliding window because, at each forecast, the most remote value within the PSW is removed and a new recent sample, never seen before, is inserted preserving the order of observation of samples, creating a forecasting model based on samples sliding. It works iteratively by taking in input PSW, predict the value at PH, sliding left the samples in the PSW of one position, adding a new sample, predicting at PH+1 and iterating, till data are available.

### 3.5.2 Novelty CGM Forecasting Model

One of the main contributions of this thesis is the definition of a new forecasting methodology. It has been defined specifically to address diabetic patient and CGM data but it can easily be generalized to work with any time-series data.

In the proposed approach, unlike those analyzed in the state of the art, no attempt is made to predict the exact outcome of the blood glucose measurement sensor. The method is based on glucose reference ranges, widely used in literature:

- Time in range (TIR): indicate the percentage of time the measured blood glucose is between 70 mg/dL and 180 mg/dL.
- Time above range (TAR): indicate the percentage of time the measured blood glucose is above 180 mg/dL. It has two levels of severity, low between 180 mg/dL and 250 mg/dL and high above 250 mg/dL.
- TBR: Time below range (TBR): indicate the percentage of time the measured blood glucose is below 70 mg/dL. It has two levels of severity, low between 54 mg/dL and 70 mg/dL and high below 54 mg/dL.

and tries to predict in a certain instant of time in which reference range the value measured by the sensor will fall.

Since the dataset is composed of time series data, we modeled the features according to date-time information: we define the date-time space as a six-dimensional space: $\mathcal{DT}$ where $\mathbf{x} \in \mathcal{DT} : \mathbf{x} =< yyyy, mm, dd, HH, MM, SS >$[8] meaning that it includes all point where each dimension represent a date-time component.

[8] yyyy: year, mm: month, dd: day, HH hour, MM minutes, SS seconds.

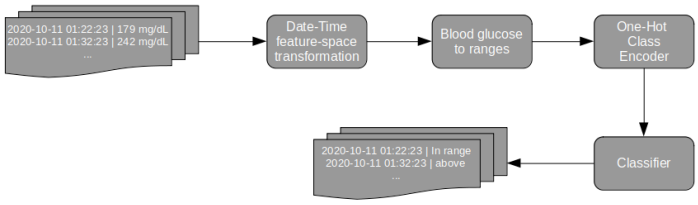



**Figure 3.2.** CGM Signal Forecasting Proposed Framework.

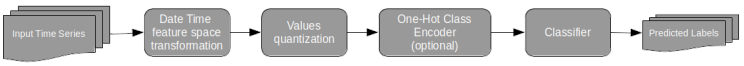



**Figure 3.3.** General Time Series Forecasting Proposed Framework.

Once the CGM series has been transformed into DateTime-space, each data is labeled with its reference interval, according to the value retrieved from the sensor. For example, if a measurement made on 2020-12-10 10:15:00 is of 145 mg/dL, then the value is *"in range"* and we assign to it the in-range label. Then labels values are 1-Hot encoded. The labels considered are 5 and correspond to the five reference ranges defined previously (note that TAR and TBR have two intervals each based on severity).

*Once we have transformed the input dataset as explained, the overall problem left is a labeling problem i.e. a classification problem, where instant of time are classified according to the most probable glucose range.*

In fig. 3.2 the full framework is depicted, the used classification algorithm is up to the user, in this work Random Forest classification algorithm has been used.

Although this technique was specifically designed for forecasting on CGM data, it can easily be generalized for any type of forecast. In particular, we can think of this technique as a quantization of the ordinate axis, as happens for the digitization of analog signals[9]. The general framework is showed in fig. 3.3.

We can formalize the method of converting a forecasting problem to a date-time labeling one (classification) as follows: we define a generic univariate time series $TS$ as a set of ordered $(\mathbf{x}, y)$ tuples, where $\mathbf{x} \in DT$ represent a date-time and it is responsible of ordering data in the series[10], while $y \in R$ represent the series value:

$$TS = \{(\mathbf{x_1}, y_1), (\mathbf{x_2}, y_2), ..., (\mathbf{x_n}, y_n)\}, \quad n < \infty \tag{3.21}$$

Then we assume a feature space function $FS$, that maps our time variable data into features and we also assume a quantization function $Q$ that is able to map our data

[9] https://en.wikipedia.org/wiki/Quantization_(signal_processing)

[10] We defined $\mathbf{x}$ as a datetime value for convenience but it can be generalized to be any number in $R^n$.

to a finite interval. Our transformed dataset will be:

$$TS' = \{(FS(\mathbf{x_1}), Q(y_1)),\ (FS(\mathbf{x_2}), Q(y_2)),\ ...,\ (FS(\mathbf{x_n}), Q(y_n)\ )\},\quad n < \infty \tag{3.22}$$

The $FS$ and $Q$ function vary and can be designed by the user.

In this work the $FS$ function is defined above as the transformation from timestamp to date-time space $DT$ while $Q$ is defined as:

$$Q(y) = \begin{cases} In - Range & \text{if } 70 < y \leq 180 \text{ (mg/dL)} \\ Above - Range - Level - 1 & \text{if } 180 < y \leq 250 \text{ (mg/dL)} \\ Above - Range - Level - 2 & \text{if } y > 250 \text{ (mg/dL)} \\ Below - Range - Level - 1 & \text{if } 54 < y \leq 70 \text{ (mg/dL)} \\ below - Range - Level - 2 & \text{if } y \leq 54 \text{ (mg/dL)} \end{cases} \tag{3.23}$$

Now that we have transformed sensors outcomes into a small and finite set of possible values, we can proceed with classification (or further transformations like 1-hot encoding of class labels).

### 3.5.3 Support Vector Regressor

Support vector regressors[44] are defined as all those registration algorithms that use support vector machines. In this work, it has been used as a regressor in the sliding window forecasting model, since it was the best performing in [38].

Given training vectors $x_i \in R^p$, $i = 1, ..., n$ and a vector $y \in R^n$, then SVR solves the primal problem[6]:

$$\min_{w,b,\zeta,\zeta^*} \frac{1}{2} w^T w + X \sum_{i=1}^{n} (\zeta_i + \zeta_i *) \tag{3.24}$$

subject to:

$$y_i - w^T \phi(x_i) - b \leq \epsilon + \zeta_i, \tag{3.25}$$

$$w^T \phi(x_i) + b - y_i \leq \epsilon + \zeta_i^*, \tag{3.26}$$

$$\zeta_i, \zeta_i^* \geq 0, i = 1, ..., n \tag{3.27}$$

The problem can be easily solved in its dual formulation:

$$\min_{\alpha,\alpha^*} \frac{1}{2} (\alpha - \alpha^*)^T Q (\alpha - \alpha^*) + \epsilon e^T (\alpha - \alpha^*) - y^T (\alpha - \alpha^*) \tag{3.28}$$

subject to:

$$e^T (\alpha - \alpha^*) = 0, \tag{3.29}$$

$$0 \leq \alpha, \alpha^* \leq C, \quad i = 1, ..., n \tag{3.30}$$

where $e$ is an identity vector, $Q$ is an n by n positive semidefinite matrix made of kernel entry: $Q_{ij} \equiv K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$. The solution to the regression problem (i.e. the predictions) is:

$$\sum_{i \in SV} (\alpha - \alpha^*) K(x_i, x) + b \tag{3.31}$$

and is often called *Support Vector expansion*[43].

### 3.5.4 Random Forest Classifier

Random Forest methods belong to the ensemble methods class since they combine multiple learners to get the final outcome. Indeed it takes the name from the fact that it is a combination of decision tree classifiers, and a huge set of trees form a "forest". Learners are combine using *bagging*[46], even if other techniques exists[6].

If we call $x_i \in R^n,\ i = 1, ..., n$ our training vectors and $y \in R^l$ the label vector, then we can say that a decision tree is an algorithm that recursively partitions the input space to group together samples with the same label.

If we call $Q$ the data at node $m$, a candidate split for that node is defined as $\theta = (j, t_m)$ made of features $j$ and split threshold $t_m$. The split partitions the node into $Q_{left}(\theta)$ and $Q_{right}(\theta)$, where:

$$Q_{left}(\theta) = (x, y)|x_j \leq t_m \tag{3.32}$$

and

$$Q_{right}(\theta) = Q \setminus Q_{left}(\theta) \tag{3.33}$$

An impurity function can be defined $H()$ to indicate all the samples in the split that are misclassified. This function depends on whether we are classifying or regressing. Since in this work we used Random Forest for classification purposes, it can be defined as entropy: assume a we have K classes, node $m$ representing a region $R_m$ with $N_m$ samples in it, then we can define the proportion of classes in that region as

$$p_{mk} = \frac{1}{N_m} \sum_{x_i \in R_m} I(y_i = k) \tag{3.34}$$

and the impurity became ($X_n$ is the training data in node $m$):

$$H(X_m) = \sum_k p_{mk}(1 - p_{mk}) \tag{3.35}$$

By joining the impurity of left and right node we obtain the objective impurity function, to be minimized:

$$G(Q, \theta) = \frac{n_{left}}{N_m} H(Q_{left}(\theta)) + \frac{n_{right}}{N_m} H(Q_{right}(\theta)) \tag{3.36}$$

the solution is:

$$\theta^* = argmin_\theta G(Q, \theta) \tag{3.37}$$

And we recursively apply it for all the subset $Q_{left}(\theta^*)$ and $Q_{right}(\theta^*)$ until we cannot split anymore.

## 3.6 Implementation

To implement the work we have used python 3 and some popular libraries for scientific analysis, data manipulation and machine learning, like Scikit-Learn[35], Pandas[34, 48], Numpy[16], Plotly[17]. The script are available either as python scripts or as jupyter notebooks depending on whether a visual output is fundamental to comment results.

# Chapter 4

# Results

In this chapter, the main results achieved in this study are presented. The first part focus on visual analytics outcomes that confirm the validity of CGM as a support tool for retrospective analysis into an educational program for T2DM patients. The second part of the results focus on the CGM signal as a time series and forecasting of its values.

## 4.1 Overall Population Analysis

As a pre-processing step for the analytics, a table with 75 records is created (25 *patient* $\times$ 3 *visits*). Those records are an aggregated form of the original sensor data representing the percentage of time each subject is below range (TBR%), above range (TAR%), in range (TIR%). For the hypoglycemia and hyperglycemia, it has differentiated between events of level 1 and level 2, denoting respectively low and high severity.

From this aggregation, when plotting the distributions, we are sure that we are plotting according to our population and not according to the sensors data, and hence the median in the boxplot (see next referenced figures) divides population (not sensor measures) into low 50% and high 50%. From Figure 4.1, the first evidence of the effectiveness of the educational path emerges. In the figure, in red, are represented subjects with only SMBG feedback between visits, while in blue the group with the double feedback SMBG + CGM. We observe how 50% of the CGM population have an effective response to the cure after V1, increasing its median TIR from 54% up to 63.75% while the other 50% drastically decrease the percentage of time spent in range, hence do not respond well to the educational path. From this first graphical analysis, it emerges that, despite of the 50% of the population that respond well to the cure, the other 50% did not, even by worsening the percentage of time spent in range. These changes are reflected in an increase in the standard deviation which therefore tends to make the distribution wider. As a consequence of the increase in both positive and negative values, the average remains constant between visits V1 and V2 and is equal to 54%. On the other side, no improvement can be found in the SMBG group. Indeed we see that the avg. TIR percentage decreases from 57% at V0 to 51% at V2 and also the median goes from 64% to 48%.

From this first analysis, it can be concluded that the CGM group shows im-

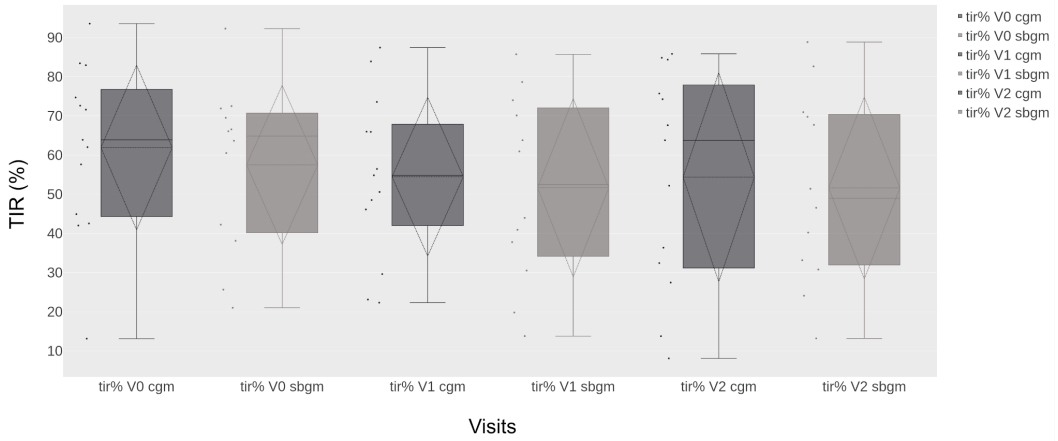



**Figure 4.1.** Time In Range (%) Distribution - CGM in blue vs. SMBG in red. We can see that cgm group at V2 has a subgroup of people improving the percentage of time they stay in range.

provements in the percentage of time spent in range i.e. they spent more time in euglycemia. Since below range (hypoglycemic) events are very rare in type 2 patients, the TAR% distribution is complementary to what has been said about TIR% distribution.

It is important to note that even if we see a positive trend in terms of increasing the percentage of time spent in range (euglycemia), the same trend is visible in the increase of TAR% of level 2 (high severity) as shown in Figure 4.2. The major contribution to worsening is given by subjects that do not show improvements in TIR since 71.43% of the improved TIR% patients do not show TAR% worsening.

## 4.2 Clustering

In this section clustering techniques explained in the previous chapter are applied. Recall that the number of clusters (k) is an input parameter for the k-means algorithm. A very frequent scenario while using k-means clustering is to start without knowing the right value of k and then consider it as a hyper-parameter to be tuned. A well-known heuristic used to estimate the right k is to run the algorithm for an increasing number of k and then plot the average diameters of the clusters for each run. Eventually, we will find two values of k with a very little decrease in the average diameter, and we restrict our search interval between those two new values until we find the right value of k. By examining Figure 4.3, we see that the suggested optimal k is between 5 and 8.

Despite the heuristic, the groups generated starting from k = 3 are always formed by 2 main and highly populated sets while the remaining k-2 sets consisting of 1 or 2 subjects with a trend not very different from that of the two main groups in terms of TIR% and TAR%. Thus the decision of running the k-means algorithm with $k = 2$ and hence to cluster the subjects only into two sets called Group A and Group B.

After the feature selection process, using the chi-squared test as described in the Methodology chapter, we have selected 28 over 132 initial features to perform the

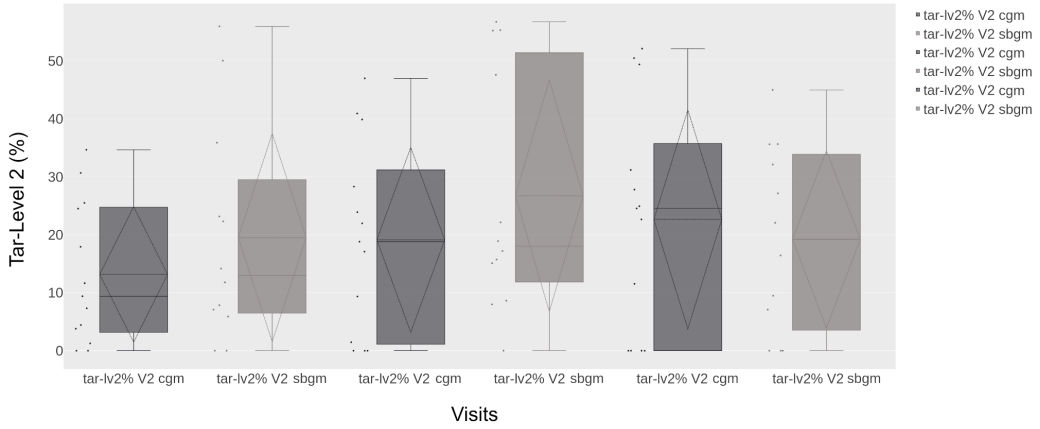



**Figure 4.2.** Time Above Range of Lv.2 [>250 mg/dL] (%) - CGM vs. SMBG - between the three visits. This percentage is intended with respect to TAR(%).

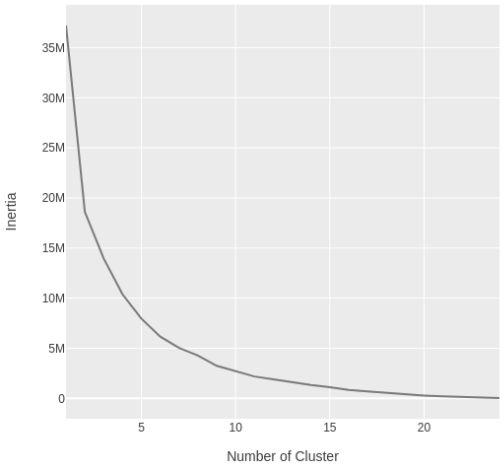



**Figure 4.3.** Average diameters of the clusters (inertia) for increasing number of k in k-means algorithm

k-means algorithm. One of the most relevant feature used is the quantity of insulin administered each day.

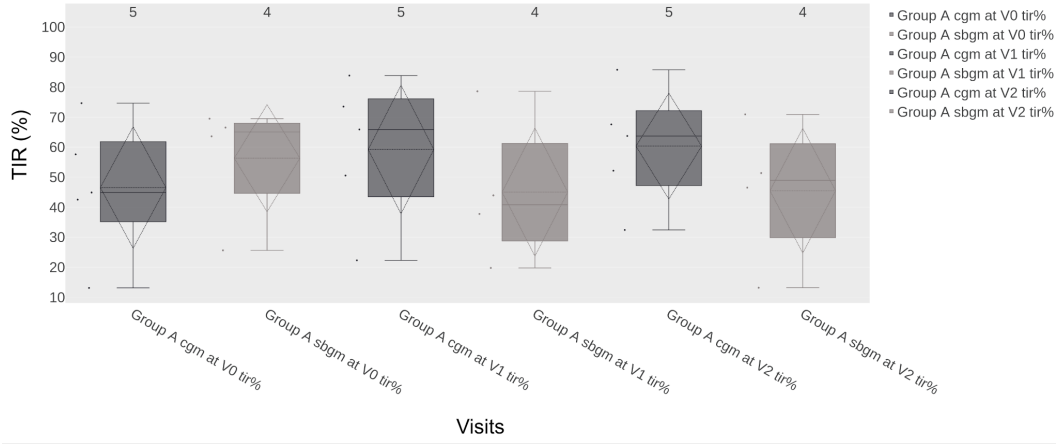
We have divided all available features of our dataset into 7 different groups: generics, dataset feature about V0, about V1, about V2, and features computed using CGM signal statistics (like mean TIR%, std TAR% ...) at V0, V1, and V2. Then we have used all the possible combinations of these groups to produce input features for different k-means execution. Then we selected the best features by filtering out all those features that produce mono-labeled clusters. As the second step, we computed the index explained previously and filtered features with values indicating inadequate clustering. In the end, we remained with three different features to compare with the chi-squared selection.

Result are presented as triple denoting the three indexes to evaluate clusters ($s1,\ s2,\ s3$), denoting respectively the silhouette coefficient, Calinski-Harabasz index, and Davied-Boulind index:

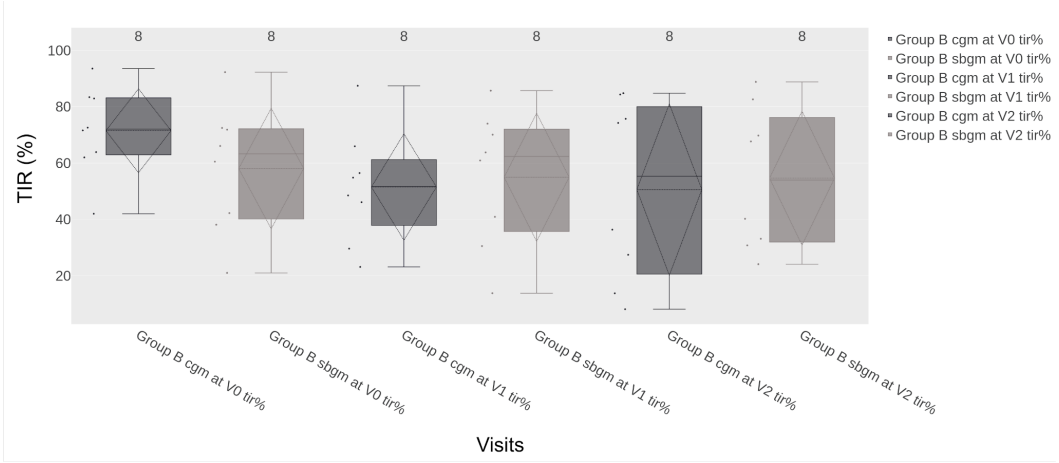
- feature obtained with Chi-Square selection: (0.63, 63.21, 0.44)
- generic feature + dataset feature about V0, V1 and V2: (0.45, 22.84, 0.88)
- features computed using CGM signal statistic at V0, V1 and V2: (0.61, 70.05, 0.51)

In the end, it has been concluded that, despite quite similar between them in term of indexes performances, but with produced clusters with at most 50% of the element in common between different run of k-means using different features, feature obtained with the Chi-Square selection are the preferred ones because they both have better clusters separation indexes and they provide the additional characteristic of being statistical tested and relevant.

Results are shown in Figure 4.4 and 4.5. It can quickly be noticed that Group A is the one where the whole population is sensitive to improvements and shows a marked improvement over Group A SMBG subjects. Group B, instead, has a group of virtuous 4, that perform even better than Group A CGM, and another group of 4 that instead do not respond well. In Group B it is more difficult to distinguish between CGM and SMBG in terms of improvements and response to the educational path. Those results reflect to TAR%, see Figure 4.6. From Figure 4.7 it can be seen that, despite group A increases its TIR%, whenever a patient experiences a hyperglycemic event, it becomes more severe. Even if the maximum TAR% of level 2 is reduced, and hence a small improvement is achieved, the amount of severity 2 hyperglycemic events settle down to 20% of total hyperglycemic events with a very narrow distribution. This means that an increase of *high severity hyperglycemic events* is shown on Group A CGM subjects, despite the increase of TIR%. It might be concluded that some subjects, as a response to the training, spend more time in range but whenever their glucose value goes above range, they do it with high severity. On the opposite, TAR%-Level 2 distribution of Group B CGM shows a general worsening of the response to the training in terms of an increase of percentage of level 2 TAR. But, by looking carefully at the data, we see that 3 of the 4 virtuous subjects do not experience at all TAR%-Level 2 events, while the bad response 4 makes the distribution look worse.

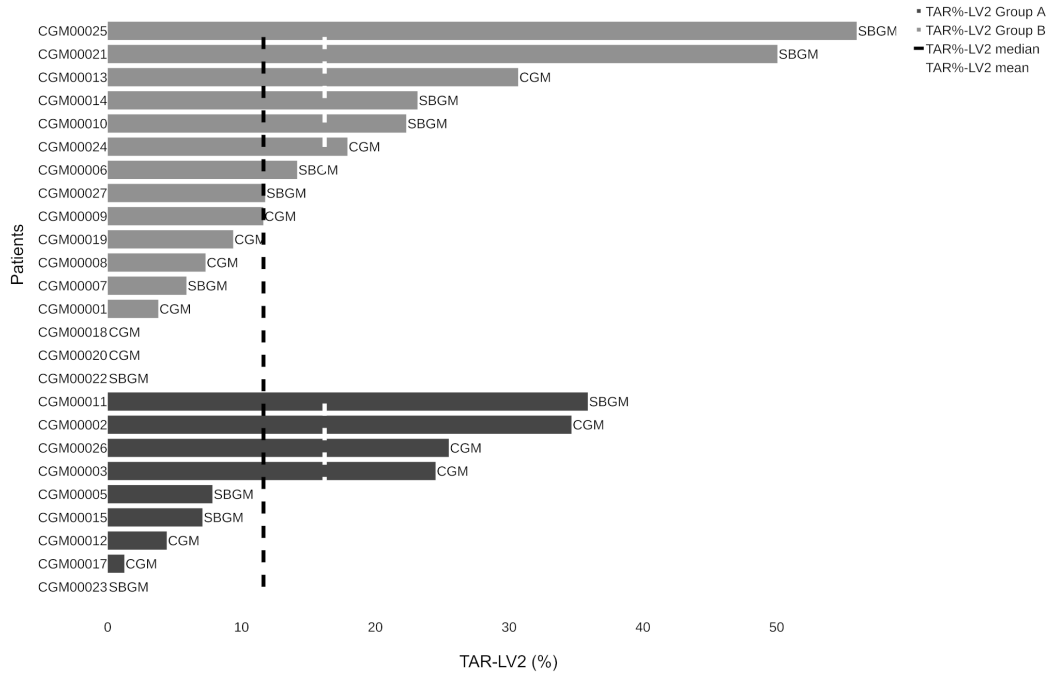



**(a)** K-means Group A TIR(%)

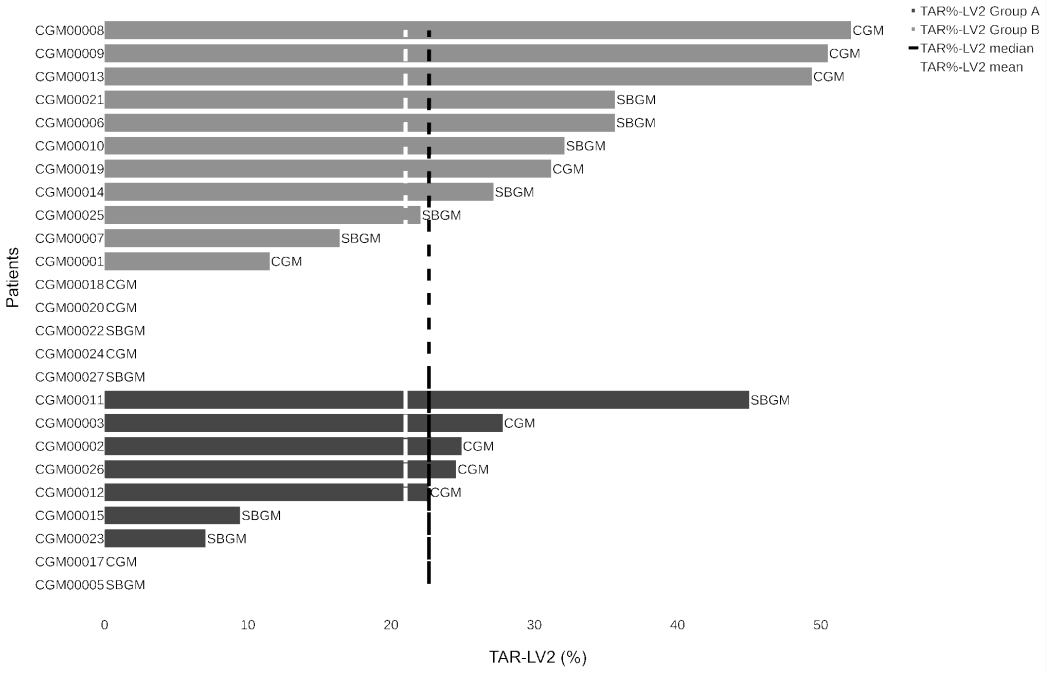



**(b)** K-means Group B TIR(%)

**Figure 4.4.** Time in Range [70–180 mg/dL] (%) between visits for K-means Groups. Group A CGM show a general increase in percentage of time spent in range, with a positive trend also for the whisker denoting minimum. On the contrary, Group A SMBG do worsen its TIR% and shows an evident negative trend on the minimum whisker. Group B has a less linear behavior. Group B CGM starts as but then all but one decrease more than 20 points the TIR% between V0 and V1. Then the subjects split into two subgroups, one with an increase of TIR%, one opposite. The same observation is true for SMBG. For Group B we can consider insufficient the response to the CGM as an educational path since their performances are compared with the ones of Group B SMBG subjects.

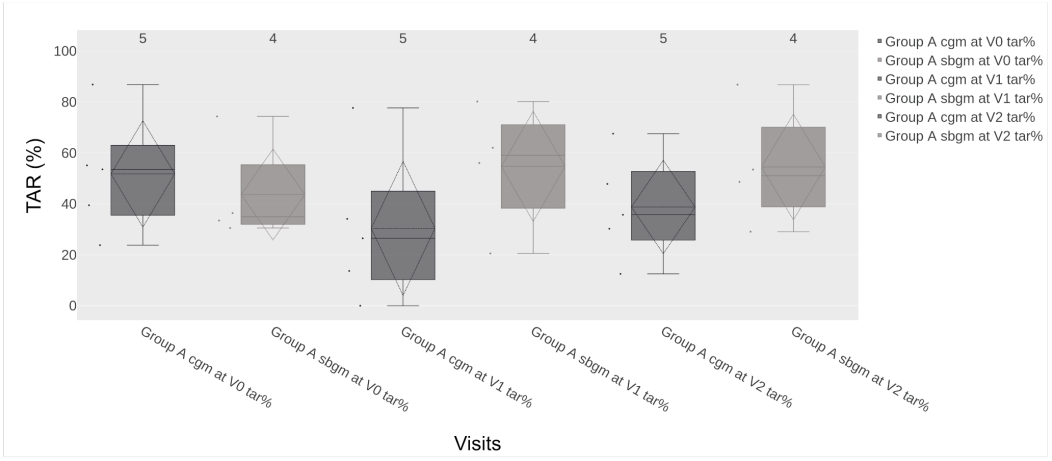



**(a)** TAR% of Level 2 between k-means group at V0. Median and mean are showed as dashed lines and are computed above the whole data-set.

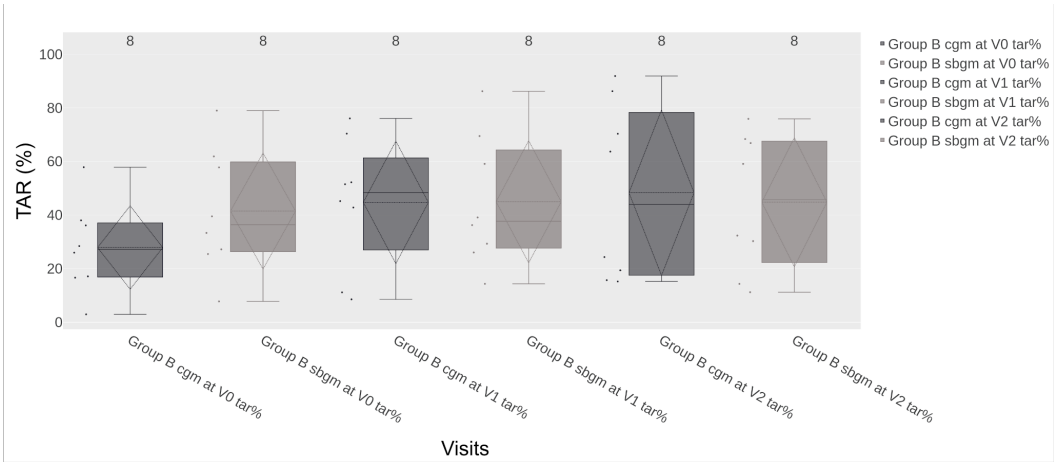



**(b)** TAR% of Level 2 between k-means group at V2. Median and mean are showed as dashed lines and are computed above the whole data-set.

**Figure 4.5.** Comparison of Time Above Range of Level 2 [>250 mg/dL] (%) between visits V0 and V2 for K-means Groups. Those pictures gives a different perspective respect to box-plot but make us drive the same conclusions.

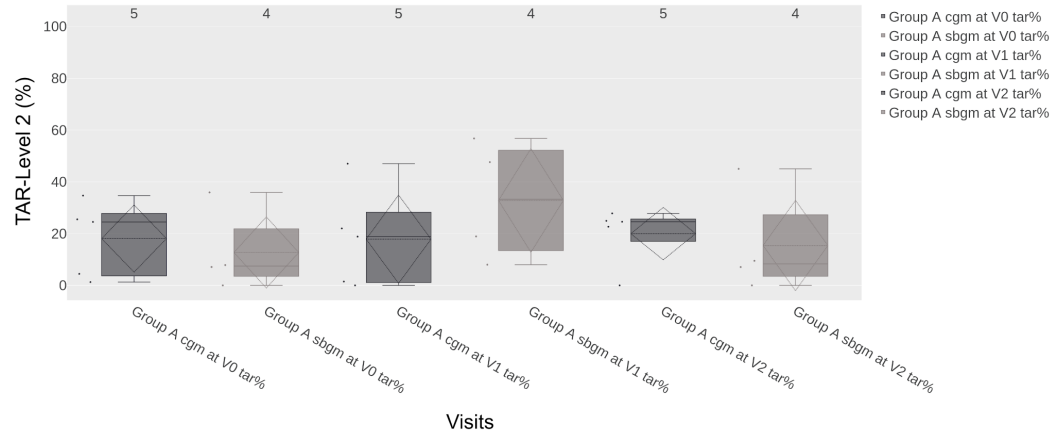



**(a)** K-means Group A TAR(%)

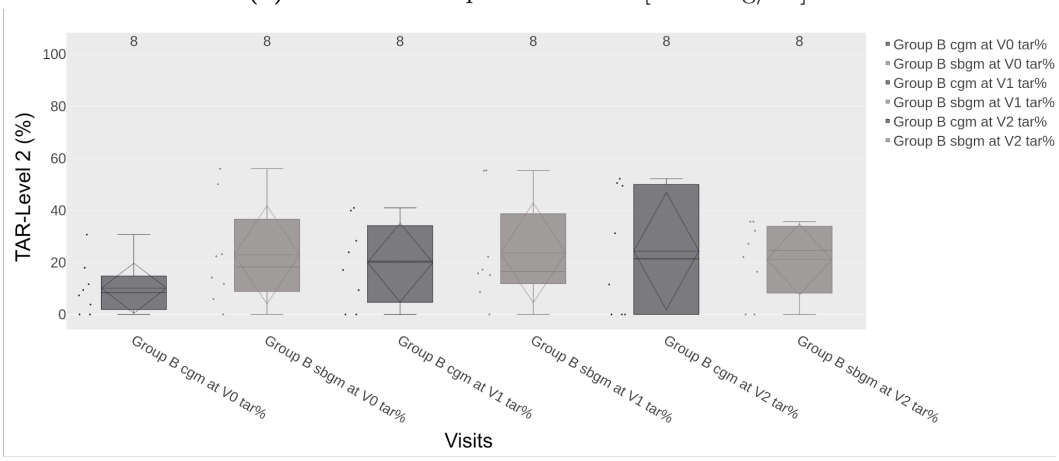



**(b)** K-means Group B TAR(%)

**Figure 4.6.** Time Above Range [>180 mg/dL] (%) between visits for K-means Groups. Since TBR% is almost zero, this plot is complementary to Figure 4.4

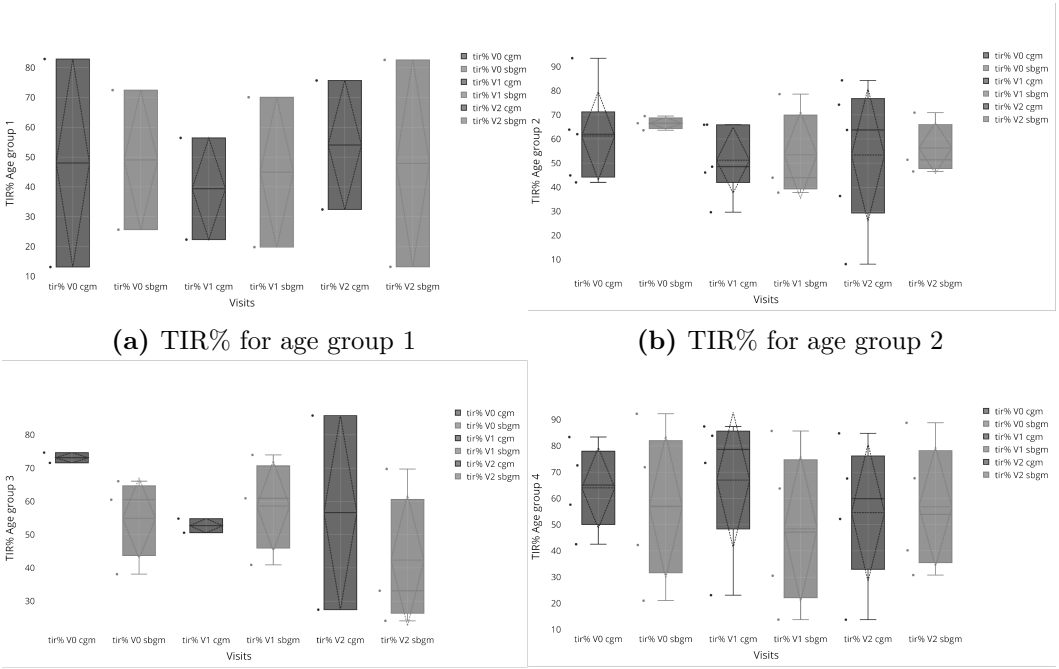



**(a)** K-means Group A TAR Lv.2 [>250 mg/dL]




**(b)** K-means Group B TAR Lv.2 [>250 mg/dL]

**Figure 4.7.** Comparison of Time Above Range of level 2 [>250 mg/dL] (%) between visits for K-means Groups. Those percentages are intended with respect to TAR(%). A problem emerges from this plot: the improvement in TIR% of Group A CGM is translated into an increase of severity 2 hyperglycemic events, even if the maximum whisker is reduces, and hence a small improvement is achieved. The amount of severity 2 hyperglycemic events settle down to 20% of total hyperglycemic event and has a very narrow distribution. On the opposite, Group B CGM distribution is almost the same as Group B SMBG.




**(a)** TIR% for age group 1

**(b)** TIR% for age group 2

**(c)** TIR% for age group 3

**(d)** TIR% for age group 4

**Figure 4.8.** Time In Range (TIR) percentage between visit for age groups: age group 1 between 48-60 years old, group 2 between 61-64 years old, group 3 between 65-66 years old,group 4 between 67-78 years old.

## 4.3 Age Groups Analysis

In this section, we are going to explore any significant behavior in terms of time in ranges percentages of the various age groups described in the *Population* section.

As shown in Figure 4.8 it is difficult to find a clear and distinct pattern between age groups, however, plots confirm a general better response to the cure in term of TIR% from patient in the CGM group, despite some outlier that is not able to respond to the cure. Between those outliers, they are quite heterogeneous and there is not a common factor between those patients that allows us to categorize them. Since hypoglycemic episodes are quite rare for T2DM patients, TAR% plots are complementary to TIR% ones and hence omitted because they do not add any value to the discussion.

## 4.4 Glycemic Variability Analysis

In this section, we are going to characterize our population hyperglycemia and hypoglycemia risk according to blood glucose variability indexes defined previously for both overall population, then distinguishing by age groups.

### 4.4.1 Overall Population

Recall from the previous chapter that the J-index is considered ideal with values less than 20, good less than 30, inadequate if greater than 40 [14]. In our test case, 71% of the subject are out of control between the three visits. Only one subject is in ideal control, but we have CGM values for him/her only at V0. Only 3 subjects in good control.

We have very high MAGE values, meaning that our subjects have a very high risk of hyperglycemia. More than 85% of measured values have a risk higher than 150. The first quartiles of the MAGE index is at 174, the third is 197.

The GVI index is always greater than or equal to 4 [36]. Its 1st quartiles is 11.67 while its 3rd is 19.85, meaning, also here, a very high variability of data.

The LGI is less than 1.1 for 89% of subjects meaning that they have a minimum risk of hypoglycemic events. Only one subject has a high risk at V1, all the others are between minimum and low-risk zone. On the contrary, the HGI index is in the minimum-risk zone ($HGI < 5.5$) only for 8 subjects over 28, meaning around 28% of the total. The remaining part are 35% in low-risk ($5.5 < HGI < 10$), 17% in the medium risk zone ($10 < HGI < 15$), and all the others (17%) in the high-risk zone. Those observations confirm that T2DM subjects are more prone to hyperglycemic events and also show different behaviors in terms of hyperglycemic variability in our subjects. It is interesting to observe that 77.7% (7 of 9) patients belongs to k-means Group B, that is the one with worse performances in term of TIR% (see Figure 4.4b).

Detailed results are presented in Table 4.1

| | | Sensor Glucose (mg/dL) | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | std | cv | J_index | MAGE | GVI | LGI | HGI |
| patient | visit | | | | | | | |
| CGM00001 | V0 | 33.48 | 0.22 | 32.25 | 162.24 | 19.33 | 0.00 | 3.81 |
| | V1 | 38.93 | 0.23 | 47.35 | 179.00 | 24.58 | 0.00 | 8.34 |
| | V2 | 32.69 | 0.22 | 32.18 | 181.53 | 18.23 | 0.00 | 4.17 |
| CGM00002 | V0 | 51.33 | 0.33 | 57.15 | 191.36 | 22.19 | 0.03 | 11.08 |
| | V1 | 33.45 | 0.22 | 45.25 | 183.18 | 11.05 | 0.04 | 7.30 |
| | V2 | 44.80 | 0.26 | 41.66 | 172.40 | 14.52 | 0.29 | 6.36 |
| CGM00003 | V0 | 43.29 | 0.24 | 49.96 | 185.99 | 23.92 | 0.01 | 9.03 |
| | V1 | 30.71 | 0.20 | 30.55 | 159.40 | 16.73 | 0.02 | 3.41 |
| | V2 | 51.45 | 0.33 | 41.59 | 166.98 | 29.47 | 0.50 | 5.72 |
| CGM00004 | V0 | 37.71 | 0.23 | 49.00 | 178.57 | 19.73 | 0.00 | 8.35 |
| CGM00005 | V0 | 29.78 | 0.17 | 34.81 | 166.88 | 10.13 | 0.00 | 5.09 |
| | V1 | 27.11 | 0.19 | 30.14 | 155.16 | 12.59 | 0.01 | 3.58 |
| | V2 | 30.33 | 0.19 | 38.07 | 164.42 | 9.34 | 0.00 | 5.32 |
| CGM00006 | V0 | 41.14 | 0.30 | 38.02 | 149.17 | 8.35 | 0.57 | 4.98 |
| | V1 | 35.41 | 0.23 | 38.64 | 165.19 | 10.03 | 0.01 | 6.06 |
| | V2 | 52.01 | 0.27 | 58.53 | 233.61 | 21.92 | 0.15 | 14.51 |
| CGM00007 | V0 | 39.42 | 0.26 | 36.53 | 159.23 | 15.88 | 0.12 | 4.76 |
| | V1 | 34.25 | 0.20 | 33.26 | 180.19 | 14.86 | 0.00 | 5.70 |
| | V2 | 41.21 | 0.28 | 33.25 | 148.03 | 13.09 | 0.34 | 3.66 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| CGM00008 | V0 | 32.25 | 0.18 | 38.93 | 160.67 | 35.86 | 0.08 | 5.35 |
| | V1 | 33.33 | 0.15 | 58.37 | 201.24 | 27.18 | 0.00 | 14.34 |
| | V2 | 38.92 | 0.16 | 69.54 | 231.56 | 17.68 | 0.00 | 17.95 |
| CGM00009 | V0 | 29.26 | 0.19 | 34.19 | 153.10 | 5.18 | 0.02 | 4.94 |
| | V1 | 33.85 | 0.17 | 50.89 | 198.91 | 6.91 | 0.00 | 10.81 |
| | V2 | 48.87 | 0.22 | 74.13 | 206.15 | 22.98 | 0.00 | 17.92 |
| CGM00010 | V0 | 38.55 | 0.24 | 40.29 | 173.10 | 16.64 | 0.17 | 6.02 |
| | V1 | 39.11 | 0.25 | 36.25 | 150.82 | 13.97 | 0.11 | 4.81 |
| | V2 | 48.78 | 0.22 | 66.88 | 206.50 | 15.71 | 0.00 | 14.56 |
| CGM00011 | V0 | 55.52 | 0.24 | 66.71 | 206.78 | 21.94 | 0.00 | 15.05 |
| | V1 | 35.77 | 0.15 | 85.08 | 240.16 | 13.05 | 0.00 | 21.87 |
| | V2 | 46.48 | 0.18 | 79.72 | 232.15 | 17.42 | 0.00 | 18.88 |
| CGM00012 | V0 | 40.69 | 0.24 | 40.39 | 158.18 | 13.30 | 0.19 | 7.48 |
| | V1 | 38.25 | 0.25 | 43.65 | 170.81 | 7.10 | 0.11 | 7.10 |
| | V2 | 42.75 | 0.22 | 44.91 | 174.50 | 12.13 | 0.00 | 7.72 |
| CGM00013 | V0 | 43.08 | 0.24 | 64.83 | 203.72 | 16.51 | 0.00 | 13.88 |
| | V1 | 48.13 | 0.20 | 70.45 | 215.01 | 19.29 | 0.00 | 15.01 |
| | V2 | 33.44 | 0.14 | 84.59 | 247.53 | 18.46 | 0.00 | 21.90 |
| CGM00014 | V0 | 38.24 | 0.20 | 54.16 | 180.77 | 15.29 | 0.00 | 10.96 |
| | V1 | 41.58 | 0.21 | 57.15 | 199.08 | 14.22 | 0.00 | 12.70 |
| | V2 | 36.10 | 0.17 | 58.73 | 209.78 | 13.10 | 0.00 | 13.38 |
| CGM00015 | V0 | 31.78 | 0.17 | 40.84 | 178.95 | 8.57 | 0.00 | 6.62 |
| | V1 | 59.50 | 0.28 | 73.38 | 208.36 | 20.38 | 0.00 | 14.44 |
| | V2 | 32.13 | 0.17 | 46.46 | 181.25 | 7.56 | 0.00 | 8.71 |
| CGM00016 | V0 | 17.24 | 0.17 | 15.08 | 112.27 | 3.98 | 1.18 | 0.13 |
| CGM00017 | V0 | 25.26 | 0.15 | 29.54 | 160.32 | 10.04 | 0.01 | 2.87 |
| | V1 | 18.74 | 0.26 | 14.95 | 89.71 | 5.12 | 4.63 | 0.08 |
| | V2 | 38.91 | 0.30 | 28.05 | 130.51 | 14.07 | 1.17 | 2.60 |
| CGM00018 | V0 | 15.35 | 0.13 | 19.47 | 128.89 | 10.65 | 1.47 | 0.93 |
| | V1 | 30.56 | 0.29 | 22.67 | 110.53 | 17.97 | 6.14 | 2.42 |
| | V2 | 37.17 | 0.26 | 27.74 | 125.30 | 13.83 | 0.58 | 2.44 |
| CGM00019 | V0 | 33.31 | 0.20 | 42.44 | 189.86 | 12.79 | 0.00 | 7.22 |
| | V1 | 60.23 | 0.34 | 63.15 | 204.20 | 30.26 | 0.24 | 11.96 |
| | V2 | 33.58 | 0.16 | 61.26 | 195.94 | 17.69 | 0.00 | 12.79 |
| CGM00020 | V0 | 36.30 | 0.26 | 32.48 | 145.12 | 17.36 | 0.24 | 3.59 |
| | V1 | 30.22 | 0.20 | 30.45 | 140.93 | 11.06 | 0.38 | 2.91 |
| | V2 | 27.01 | 0.17 | 28.65 | 164.65 | 6.43 | 0.05 | 2.95 |
| CGM00021 | V0 | 46.28 | 0.20 | 82.22 | 238.84 | 30.02 | 0.00 | 18.73 |
| | V1 | 57.30 | 0.23 | 94.57 | 269.59 | 18.42 | 0.00 | 23.51 |
| | V2 | 42.84 | 0.26 | 42.92 | 199.29 | 22.42 | 0.05 | 6.44 |
| CGM00022 | V0 | 15.23 | 0.11 | 25.75 | 150.00 | 6.64 | 0.00 | 2.46 |
| | V1 | 22.83 | 0.14 | 30.17 | 153.91 | 8.37 | 0.00 | 3.58 |
| | V2 | 17.77 | 0.12 | 30.46 | 158.25 | 5.85 | 0.00 | 3.81 |
| CGM00023 | V0 | 25.72 | 0.18 | 32.32 | 155.04 | 9.52 | 0.02 | 4.05 |
| | V1 | 40.90 | 0.24 | 47.94 | 188.36 | 14.48 | 0.00 | 9.32 |
| | V2 | 27.69 | 0.16 | 41.47 | 189.87 | 10.29 | 0.00 | 9.22 |
| CGM00024 | V0 | 35.01 | 0.19 | 47.92 | 194.63 | 14.18 | 0.00 | 8.46 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | V1 | 31.29 | 0.16 | 44.71 | 197.25 | 11.92 | 0.00 | 9.19 |
| | V2 | 26.78 | 0.18 | 31.46 | 152.25 | 10.19 | 0.02 | 3.64 |
| CGM00025 | V0 | 58.86 | 0.26 | 74.25 | 233.09 | 28.89 | 0.00 | 14.87 |
| | V1 | 59.88 | 0.28 | 83.03 | 197.87 | 24.38 | 0.00 | 17.62 |
| | V2 | 33.58 | 0.18 | 56.12 | 189.98 | 15.99 | 0.00 | 13.20 |
| CGM00026 | V0 | 30.82 | 0.13 | 66.63 | 219.57 | 12.42 | 0.00 | 16.22 |
| | V1 | 39.73 | 0.20 | 73.51 | 217.00 | 11.71 | 0.00 | 17.51 |
| | V2 | 36.91 | 0.18 | 50.63 | 209.44 | 12.05 | 0.00 | 10.80 |
| CGM00027 | V0 | 30.38 | 0.17 | 44.19 | 178.54 | 10.09 | 0.00 | 8.17 |
| | V1 | 36.68 | 0.22 | 42.56 | 178.61 | 22.69 | 0.00 | 6.63 |
| | V2 | 25.02 | 0.16 | 37.19 | 173.56 | 8.93 | 0.00 | 6.36 |
| CGM00028 | V0 | 46.50 | 0.15 | 131.24 | 298.08 | 33.20 | 0.00 | 37.71 |

**Table 4.1.** Variability indexes computed daily, grouped by visit and aggregated using the median. Presented indexes, beyond standard deviation (std) and coefficient of variation (cv) are: J Index (denoting glycemic control qualities), MAGE (mean amplitude of glycemic excursion), GVI (glucose variability index expressed as percentage), LGI and HGI (Low and High Glucose Index) denoting risk of hypo/hyperglycemic events.

### 4.4.2 Age Groups

In Table 4.2 all the GVI indicators are presented for all the age groups defined in previous sections. From the tables it is possible to observe an increase of the risk of hyperglycemic events as patients get older from the HGI index: we start with age group 1, from 48 to 60 years old, with 50 % of the subject at minimal risk (HGI $<$ 5.5) and only 16.67% at medium risk and we end with age group 4, from 67 to 78 years old, where patient at minimal risk are only 20% of the total and patient with medium risk are the 40%.

The same is not true for hypoglycemic events, which are at minimal risk for all age groups of patients.

J-index also shows a lack of control of glycemic variability as patients get older: Inadequate control patients (J index $>$ 40) goes from 40% of age group 1 to 66.67% at age group 4. Patients in ideal control at age group 1 are 16.6% of the total, at age group 4 became 0%.

| | Age Group 1 | | | | | | |
|---|---|---|---|---|---|---|---|
| | std | cv | J_index | MAGE | GVI | LGI | HGI |
| patient | | | | | | | |
| CGM00001 | 36.60 | 0.22 | 36.76 | 173.35 | 19.77 | 0.00 | 4.94 |
| CGM00004 | 37.71 | 0.23 | 49.00 | 178.57 | 19.73 | 0.00 | 8.35 |
| CGM00007 | 37.98 | 0.25 | 35.35 | 159.74 | 14.56 | 0.14 | 4.41 |
| CGM00011 | 39.78 | 0.18 | 78.02 | 229.42 | 16.55 | 0.00 | 17.68 |
| CGM00016 | 17.24 | 0.17 | 15.08 | 112.27 | 3.98 | 1.18 | 0.13 |
| CGM00026 | 36.91 | 0.18 | 62.94 | 217.00 | 12.02 | 0.00 | 14.52 |
| | Age Group 2 | | | | | | |
| | std | cv | J_index | MAGE | GVI | LGI | HGI |

| patient | | | | | | | |
|---|---|---|---|---|---|---|---|
| CGM00002 | 44.80 | 0.26 | 51.73 | 181.75 | 15.95 | 0.04 | 9.42 |
| CGM00005 | 29.27 | 0.18 | 34.68 | 162.29 | 9.82 | 0.00 | 4.99 |
| CGM00013 | 42.58 | 0.20 | 70.91 | 216.65 | 17.14 | 0.00 | 15.47 |
| CGM00015 | 32.37 | 0.19 | 46.97 | 186.21 | 12.72 | 0.00 | 9.24 |
| CGM00018 | 32.17 | 0.23 | 24.12 | 125.62 | 13.48 | 0.65 | 1.64 |
| CGM00019 | 38.35 | 0.22 | 51.13 | 194.82 | 19.71 | 0.01 | 10.01 |
| CGM00023 | 30.70 | 0.18 | 43.94 | 181.11 | 11.22 | 0.00 | 8.16 |
| CGM00024 | 31.58 | 0.18 | 43.19 | 185.79 | 13.60 | 0.00 | 7.58 |
| | Age Group 3 | | | | | | |
| patient | std | cv | J_index | MAGE | GVI | LGI | HGI |
| CGM00009 | 37.52 | 0.21 | 53.86 | 200.25 | 9.03 | 0.00 | 11.64 |
| CGM00010 | 40.26 | 0.24 | 43.64 | 170.62 | 15.57 | 0.09 | 6.75 |
| CGM00014 | 38.35 | 0.19 | 58.22 | 198.52 | 14.05 | 0.00 | 12.82 |
| CGM00017 | 31.15 | 0.28 | 26.22 | 135.93 | 10.19 | 1.16 | 2.20 |
| CGM00027 | 30.25 | 0.18 | 39.47 | 176.57 | 10.12 | 0.00 | 6.68 |
| | Age Group 4 | | | | | | |
| patient | std | cv | J_index | MAGE | GVI | LGI | HGI |
| CGM00003 | 41.77 | 0.27 | 39.27 | 172.38 | 22.84 | 0.07 | 5.49 |
| CGM00006 | 45.66 | 0.26 | 51.98 | 165.59 | 11.62 | 0.11 | 9.66 |
| CGM00008 | 36.01 | 0.16 | 57.15 | 201.24 | 19.61 | 0.00 | 13.67 |
| CGM00012 | 41.77 | 0.24 | 42.64 | 163.24 | 9.06 | 0.08 | 7.37 |
| CGM00020 | 31.18 | 0.24 | 30.45 | 145.12 | 11.06 | 0.19 | 2.95 |
| CGM00021 | 46.94 | 0.23 | 75.33 | 237.74 | 27.73 | 0.00 | 14.39 |
| CGM00022 | 17.68 | 0.12 | 29.34 | 153.95 | 7.23 | 0.00 | 3.58 |
| CGM00025 | 49.14 | 0.23 | 66.19 | 197.87 | 24.38 | 0.00 | 14.68 |
| CGM00028 | 46.50 | 0.15 | 131.24 | 298.08 | 33.20 | 0.00 | 37.71 |

**Table 4.2.** Table showing all the glycemic index presented for each age group: age group 1 between 48-60 years old, group 2 between 61-64 years old, group 3 between 65-66 years old,group 4 between 67-78 years old.

## 4.5 Part Of The Day Analysis

According to [13] we divide day hours into seven periods (PoD) and refer to them as: *morning* from 6 am to 10 am, *late morning* from 10 am to 1 pm, textitearly noon from 1 pm to 4 pm, *afternoon* from 4 pm to 7 pm, *early evening* from 7 pm to 9 pm, *evening* 9 pm to 00 am and *night* 00 am to 6 am.

In this analysis, instead of considering exact sensor values, we only check if patients are within the ranges explained in the previous sections. We consider how many sensor readings are recorded in the range, above range, and below range.

In Figure 4.9 it is possible to understand the daily class distribution: on the

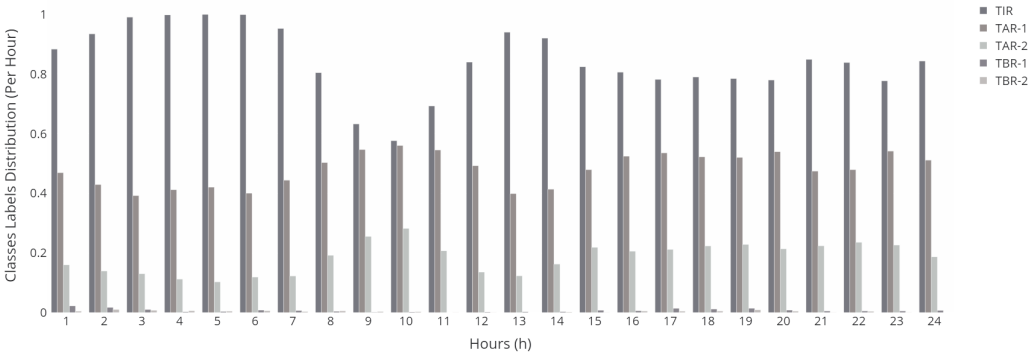



**Figure 4.9.** Number of class labels at given day hour for all the measurements periods.

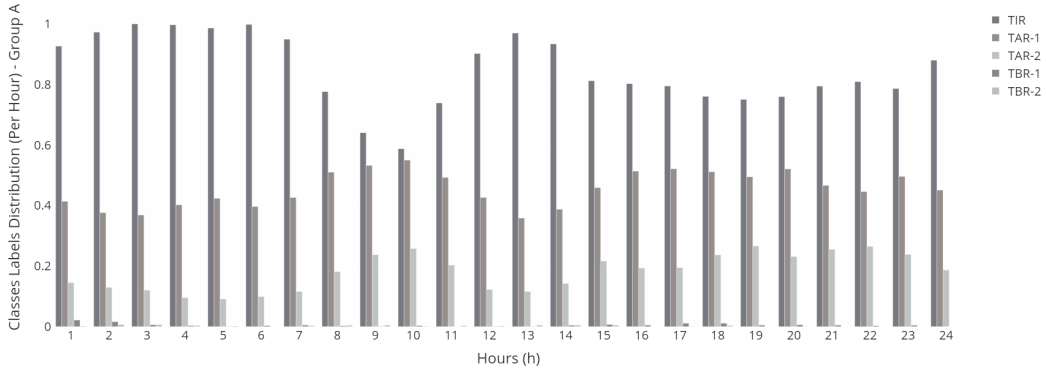
x-axis, we find the hour of the day, while on the y-axes probability of getting a given class. Bars are grouped by label, namely, blue bars denote TIR probability at a given hour, red TAR of level 1, green TAR of level 2, purple TBR of level 1, and orange TBR of level 2. It is very unlikely to have TBR values, in line with T2DM blood glucose behavior. Given that TBR events are quite rare, we can argue that TIR and TAR distribution are inverse proportional since the increase of TIR implies a decrease of TAR and vice versa. Looking at the plot, it is possible to note that daily values of TIR and TAR (both level 1 and 2) follows a sinusoidal shape with a phase shift of 90 degrees between them, while TAR of level 1 and level 2 seems to follow the same trend but with TAR of level 2 events with smaller amplitude than ones of level 1.

A significant characteristic shown from the plot is that during the morning and the late morning, from 8 am to 12 am, we see a decrease in TIR labels of around 40%, with a negative peak around 10 am, where the TIR and TAR-1 became equally likely and, summing up TAR-2 and TAR-1, we can say that it is more likely to have TAR events during the morning before 12 am than TIR. This decrease is the steepest during the day, especially if considering that during night, until the morning, TIR labels are at their maximum values. We see a positive trend for TIR labels at the beginning of early noon and then the TIR trend stabilizes to a horizontal line till night where we see the best TIR performances.

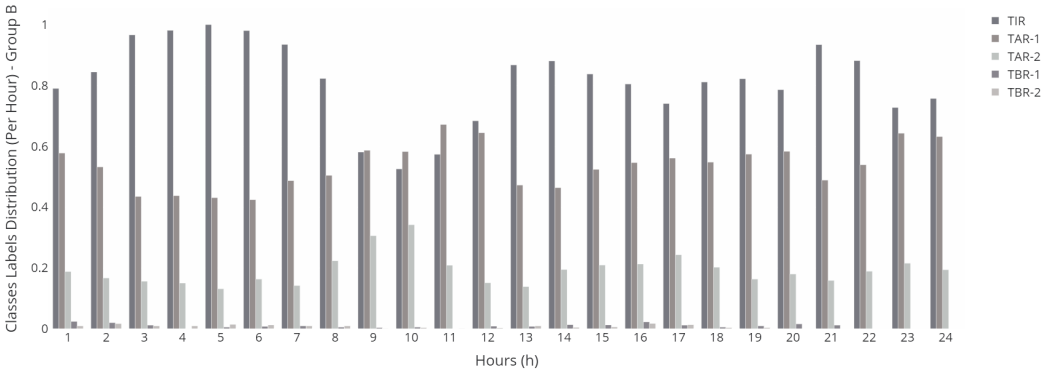
Using the same kind of plot we can again confirm the ability of k-means to capture the bad and good response of patients to the educational path. Indeed, as shown in Figure 4.10, data from both groups follow the sinusoidal pattern described above, but in Group B patients, the TAR peaks between morning and late morning are more marked, denoting a propensity for hyperglycemic phenomenon. These pictures give a different perspective with respect to the box-plot of the previous chapter but make us reach the same conclusions: Group A has a better response to the educational path than Group B.

With the same grouped bar-chart we also want to highlight the importance and efficiency of the educational process, which has led to a net increase in TIR classes at lunchtime (12 am to 3 pm) from Visit 0 to Visit 2, as shown in Figure 4.11b

In Table 4.3 it is possible to confirm the trend observed in *part-of-the-day bar-chart* in terms of high and low blood glucose risk indexes. Definition for these




**(a)** Group A class labels distribution

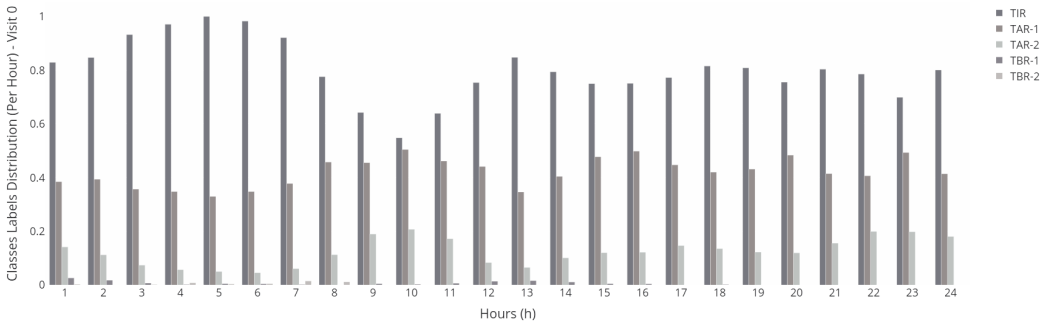



**(b)** Group B class labels distribution

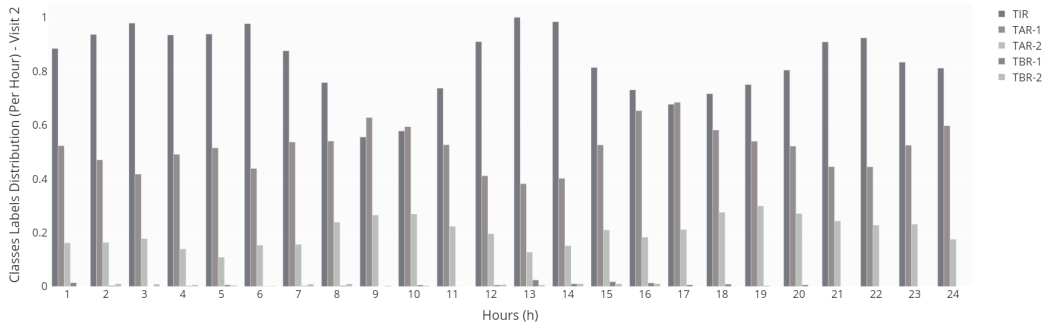
**Figure 4.10.** Comparison of class labels distribution for K-means Groups.

| risk-level | M | LM | EN | AN | EE | E | N |
|---|---|---|---|---|---|---|---|
| Hyperglycemia | | | | | | | |
| Minimal | 20 | 32 | 28 | 24 | 28 | 24 | 44 |
| Low | 36 | 24 | 32 | 24 | 32 | 28 | 28 |
| Moderate | 24 | 28 | 16 | 28 | 12 | 32 | 16 |
| High | 20 | 16 | 24 | 24 | 28 | 16 | 12 |
| Hypoglycemia | | | | | | | |
| Minimal | 96 | 100 | 96 | 92 | 92 | 92 | 92 |
| Low | 4 | 0 | 4 | 4 | 4 | 4 | 4 |
| Moderate | 0 | 0 | 0 | 4 | 4 | 4 | 4 |
| High | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Table 4.3.** Percentage of patients at different risk levels of hypo- and hyperglycemia. Abbreviations: M, Morning; LM, Late-morning; EN, Early-noon; AN, Afternoon; EE, Early-evening; E, Evening; N, Night. (1) Risk levels are defined by the Low Glucose Index (LGI): minimal risk for hypoglycemia (LGI less than 1.1); low-risk (LGI between 1.1 and 2.5); moderate-risk (LGI between 2.5 and 10); high-risk (LGI greater than 10). (2) Risk levels are defined by the High Glucose Index (HGI): minimal-risk (HGI less than 5.5); low-risk (HGI between 5.5 and 10); moderate-risk (HGI between 10 and 15); high-risk (HGI greater than 15)




**(a)** V0 class labels distribution




**(b)** V2 class labels distribution

**Figure 4.11.** Educational path response in term of class labels distribution between visits.

measures are available in section *Glycemic Variability Analysis*. From the table, we see that, in general, patients have a low risk of hypoglycemic events during the day. Results also confirm that the hyperglycemia risk deeply decreases at night, while it is not so evident the increase of hyperglycemic events between morning and late morning.

Also for age groups, the negative peak between the morning and the late morning is present and it is common to all the age groups even if it is slightly different for each group. After this negative peak, only age groups 1 and 3 show an evident recovery of TIR at launch-time with a positive peak around 1 pm., all the others recover from the negative peak but then the distribution flattens and stays more or less flat till night. TIR distribution is above 0.8 at night for all groups but only age groups 2 and 4 have very good TIR performances at night (>0.9). We observe that, as age increases (group 3 and 4) also does TAR events in the early evening. In younger groups (1 and 2) these events are less frequent and they happen in the evening, between 10 pm and 11.30 pm. We might argue that this time shift is given from different meal-time habits of age groups: older people eat earlier at night, but there is no evidence since we have not meal-time data.

## 4.6 Forecasting

In this section are described all the results obtained using sliding window forecasting on a type 2 diabetes mellitus population and using the date-time labeling methodology

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| tar1 | 0.929061 | 0.905401 | 0.917079 | 13626.0 |
| tar2 | 0.942621 | 0.884899 | 0.912849 | 5291.0 |
| tbr1 | 0.878205 | 0.674877 | 0.763231 | 203.0 |
| tbr2 | 0.919192 | 0.771186 | 0.838710 | 118.0 |
| tir | 0.967962 | 0.968488 | 0.968225 | 23896.0 |
| micro avg | 0.952370 | 0.936384 | 0.944309 | 43134.0 |
| macro avg | 0.927408 | 0.840970 | 0.880019 | 43134.0 |
| weighted avg | 0.952009 | 0.936384 | 0.943956 | 43134.0 |
| samples avg | 0.936384 | 0.936384 | 0.936384 | 43134.0 |

**Table 4.4.** Results of forecasting as classification.

(or foresting as classification) defined in the previous section.

### 4.6.1 Timestamp Labeling - Forecasting As A Classification Problem

Looking at the CGM forecasting problem as a date-time labeling problem, we can decouple the problem from the notion of time-series and we are no more tight to the concept of time windows. Indeed, now date-time data can be considered coordinates of a point in a multi-dimensional space. We can query classes by giving any date-time input to our algorithm.

By losing the concept of time we can build a single dataset with time-series data from different patients at different visits. It is possible since our data, now, are only point in a space and there is not the need of maintaining a sequential relationship between them as in the time-series sense.

In this way, we are also trying to associate specific glycemic events to specific date-time event. For example, with a balanced dataset, the algorithm might learn that at breakfast or dinner time glycemic events are more frequent due to a bad diet, according to what we have learned from PoD analysis in the previous section. As a classification algorithm, we used a Random Forest classifier. The biggest advantages of this algorithm, without considering performances, are that we can export all the decision trees generated, analyze them, we can get the decision path for any classification task. This provides not only a powerful method for classification but also a useful tool to improve human knowledge of the process.

Proceeding as described in the methodology chapter, led us to promising results shown in Table 4.4.

The Random Forest classifiers reach an accuracy of 93.6% and from Table 4.4 it is visible that the algorithm has good performances even into classifying TBR events, which have a very small support in this dataset, since they are rare events for this kind of patients. As a further experiment, we have compared classification performances between group A and group B described in the previous chapter.

Using the same method we have also tried to train and test only using visit-specific data (i.e train test only with V0 data or only with V1 data etc.) we obtain even better performances than on full dataset, always around 93% of accuracy. The

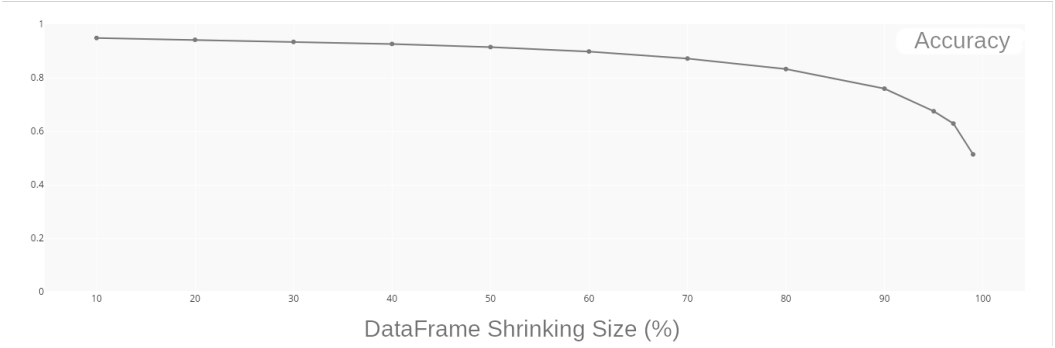



**Figure 4.12.** Accuracy on DataFrame Percentage Shrinking Size

| Features | Accuracy |
|---|---|
| month, day, hour, minute, second | 0.94 |
| year, month, day, hour, minute, second | 0.93 |
| month, day, hour, second | 0.91 |
| year, month, day, hour, second | 0.91 |
| year, day, hour, minute, second | 0.88 |
| year, day, hour, second | 0.88 |
| day, hour, second | 0.85 |
| day, hour, minute, second | 0.85 |
| year, month, hour, second | 0.72 |
| month, hour, second | 0.71 |
| year, month, day, hour | 0.68 |

**Table 4.5.** Top 10 Date-time Space features in terms of classification accuracy running Random Forest classification algorithm.

same is true using group-specific data from KM clustering with similar accuracy.

It is important to point out also that, despite our reported results assume a train-test split of 67-33% of the total dataset (130000 samples), we can reach good performances also if we remarkably shrink our dataset size. This is visible in Figure 4.12 where we can reach accuracy higher than 90% even shrinking the test size of 60% (52000 samples more or less). We have an accuracy of more than 80% with a shrinking size of 85% (only 20000 samples as training data).

Since the space we are using is not a standard feature transformation it is also provided in Table 4.5 a top-10 of features in terms of classification accuracy. The table is computed by running the algorithm on all the feature combinations, and then only the top 10 are presented, the others are not relevant. Note that the key couple of features is `(day, hour)`, that combined with another feature led almost always to satisfying accuracy results.

**Feature Space Insights** This section is functional to explain and better understand the feature space used in the "Forecasting As Classification Problem" methodology. Here we provide some insights into data in date-time space.

Recall that date-time space is a six-dimensional space where our axes are date-time coordinates $\mathbf{x} =< yy, mm, dd, HH, MM, SS >$, namely *year* ($yy$), *month* ($mm$), *day* ($dd$), *hour* ($HH$), *minute* ($MM$), *second* ($SS$), and our target values are class labels, namely $\mathbf{y} = y_i$ with $y_i = tir, tar_1, tar_2, tbr_1, tbr_2$, denoting the blood glucose values ranges widely described in this work.

As the first step, we try to understand which is the predominant class in our dataset. We do it by counting the number of occurrences. We have at first place the $tir$ class with 72000 row, followed by $tar_1$ with 41000 labels, then $tar_2$ with 15000 row and $tbr_1$ and $tbr_2$ with less than 1000 rows together. In previous discussion, it has been showed that, despite the lack of TBR values, classification is still accurate even for them, by reaching accuracy above 87%.

Then it is shown how, aggregating by `[day, hour, minute, class]`, `[day, hour, class]`, `[hour, minute, class]` and `[day, minute, class]` modifies the size of the dataset (i.e it is showed how many records have the same 'hour-minute-class' or the same 'day-hour-class'). Results are summarized as feature aggregated, row count, max and median aggregation size, dataset size reduction by aggregating:

- `[day, hour, minute, class]`: 76561 rows, max aggregation of 7 rows, median aggregation of 2 rows, reduction of 41.4%.
- `[day, minute, class]`: 5627 rows, max aggregation of 117 rows, median aggregation of 17 rows, reduction of 95.7%.
- `[hour, minute, class]`: 5160 rows, max aggregation of 78 rows, median aggregation of 20 rows, reduction of 96.0%.
- `[day, hour, class]`: 2296 rows, max aggregation of 201 rows, median aggregation of 51 rows, reduction of 98.2%.

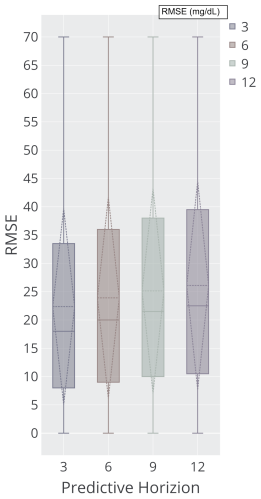
More than 40% of the dataset has the same day-time information (seconds are ignored since not relevant, the granularity of sensors is at minute) and more than 90% share the same class at hour-minute granularity.

### 4.6.2 Sliding Window Forecasting

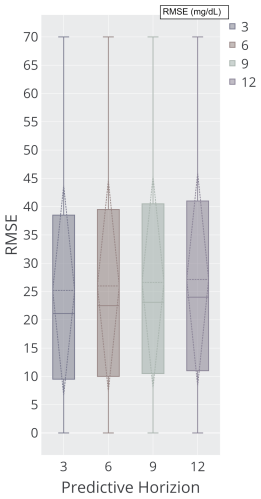
Recall from the previous chapter that the aim of this method is to replicate the work in [38] but on a different kind of population. Indeed, we will try to verify the effectiveness of the proposed methods on a population of type 2 diabetics with complex management. We verify the effectiveness of the Support Vector Regressor (SVR) on this new type of diabetic population. The error function used is the root mean squared error (RMSE). We try to optimize algorithm performances by scaling input features since the Support Vector Machine algorithm is not scale-invariant.

The results of the sliding window forecasting without any optimization are summarized in Figure 4.13. Accordingly to [**?**], as the window size increase so does the measured forecasting error. Unlike in [38], where we can see higher error values as we increase the predictive horizon, here we observe a wider error distribution that is almost constant between the analyzed predictive horizons, especially for large window sizes.

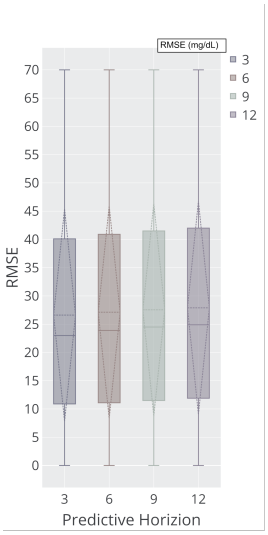
Instead in Figure 4.14 it is possible to evaluate the performance of the algorithm after having scaled the features since Support Vector Machine algorithms are not scale-invariant. In general, it can be observed that scaled features are a benefit since the error is always 2 to 5 points less than non-scaled feature runs. Performances with low predictive horizon also increase for large window sizes. Despite the small improvement, this method is far from the performances this work was trying to replicate.



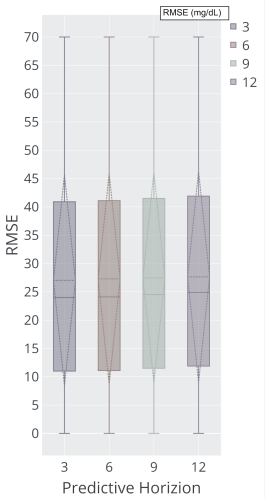

**(a)** RMSE with a sliding window size of 36 samples



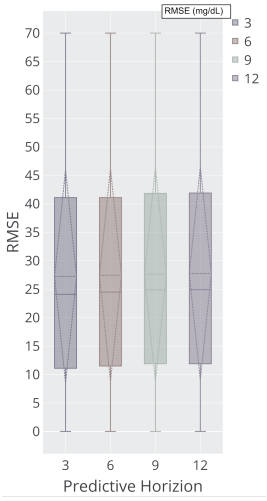

**(b)** RMSE with a sliding window size of 72 samples




**(c)** RMSE with a sliding window size of 144 samples

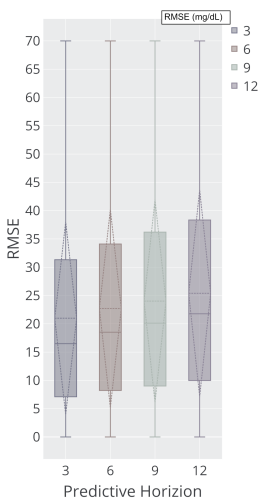



**(d)** RMSE with a sliding window size of 288 samples

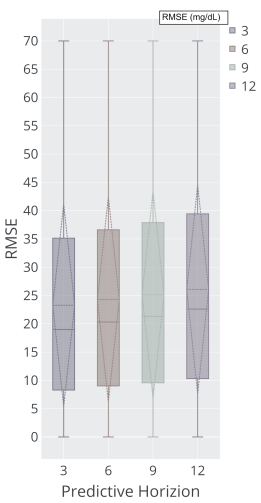



**(e)** RMSE with a sliding window size of 432 samples

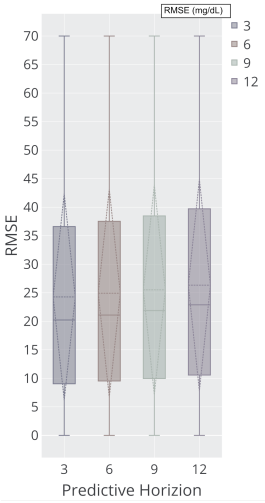
**Figure 4.13.** Root mean squared error (RMSE) of sliding window forecasting using SVR.



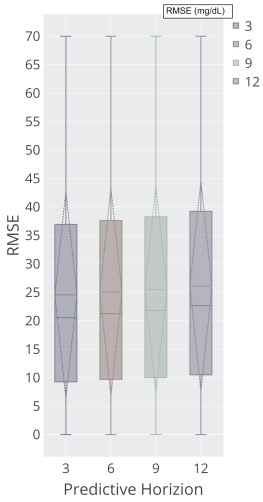

**(a)** RMSE with a sliding window size of 36 samples



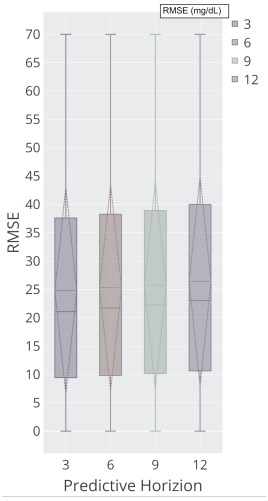

**(b)** RMSE with a sliding window size of 72 samples




**(c)** RMSE with a sliding window size of 144 samples




**(d)** RMSE with a sliding window size of 288 samples




**(e)** RMSE with a sliding window size of 432 samples

**Figure 4.14.** Root mean squared error (RMSE) of sliding window forecasting using SVR with scaled features to boost performances. Despite scaling the improvements in terms of error are minimal.

# Chapter 5

# Conclusions

This works has shown how CGM data can be analyzed using well-known visual analytics techniques to reveal important results.

It has be shown that 50% of the CGM population have a positive response to the cure after the assessment visit, increasing its median TIR percentage from 54% up to 63.75%. The other 50% decrease the percentage of time spent in range, hence did not respond well to the educational path. On the contrary, people with only SMBG feedback has not show any improvement during the training period. Their TIR percentage decreases from 57% at V0 to 51% at V2 and the median goes from 64% to 48%. From graphical analysis, it emerges an increase of TAR% of level 2 events. The major contribution to this increase is given by subjects that do not show improvements from the educational journey discussed.

Using k-means clustering over patients screening data and medical history data showed the ability to automatically catch those patients that had good or bad response to the educational path on T2DM treatment: clustering Group A CGM is the one where the whole population is sensitive to improvements and shows a marked improvement over Group A SMBG subjects; Group B has a group of *virtuous 4 subjects*, that perform even better than Group A CGM, and another group of 4 that instead do not respond well to the training. Both clustering groups experience the same correlation between TIR% improvement and TAR level 2 % increase. Also, age groups confirm a good response to educational path from patient with CGM feedback.

Glycemic variability analysis showed in general an high risk of hyperglycemia as expected in T2DM patients. Traditional statistics indicator follows typical behavior for this type of population: $\sigma$ is never higher than $\frac{\mu}{3}$. Using glycemic-specific indexes it has been discovered that 71% of the subject are out of control, according to J index reference values. The GVI index is always greater than or equal to 4. Its 1st quartiles is 11.67 while its 3rd is 19.85, meaning a very high variability of data. The LGI is less than 1.1 for 89% of subjects meaning that they have a minimum risk of hypoglycemic events. On the contrary, the HGI index is in the minimum-risk zone for 8 subjects over 28 (28%). The remaining 35% are in low-risk, 17% in

the medium risk zone, and all the others (17%) in the high-risk zone. Glycemic variability indexes analysis for age groups showed an increase of glycemic variability risk and hyperglycemia events risk in older patients. Glycemic variability risk indexes computed for part of the day showed that patients have a low risk of hypoglycemic events during the day, hyperglycemia risk deeply decreases at night, while it is not evident the increase of hyperglycemic events between morning and late morning.

By analyzing the glycemic ranges (TIR, TAR, TBR) during part of the day it can be argued that, since TBR events are quite rare, TIR and TAR distribution can be considered inverse proportional. TIR and TAR daily values follow a sinusoidal shape with a phase shift of 90 degrees between them. Sinusoidal pattern is also common to CGM and SMBG groups of patients. Between morning and late morning, from 8 am to 12 am, we see a decrease in TIR labels of around 40%, with a negative peak around 10 am. In clustering Group B patients, TAR peaks between morning and late morning are more marked, denoting a propensity for hyperglycemic phenomenon. The educational process has led to a net increase in TIR classes at lunchtime (12 am to 3 pm) from Visit 0 to Visit 2. Despite there is an increase of TAR% at morning risk indexes are not able to capture it. In early evening, older patients have a worse TIR distribution with a second but less evident decrease of percentage of time spent in range.

By using the technique of timestamp labeling (transforming forecasting problem to a classification one) it is achieved 93% of accuracy into predicting at a given date-time if CGM values will be in TIR, TAR Level 1, TAR Level 2, TBR Level 1, TBR Level 2. Same results are achieved also if training and testing only on V0, V1, or V2 data. Accuracy higher than 90% can be reached with a train-test split of 40-60% (train size of 52000 samples more or less) and an accuracy of more than 80% can be achieved with a train-test split of 15-85% (training with only 20000 samples).

The Sliding window forecasting model error increase as window size increase. Higher error values have been found as the predictive horizon increase. Scaled features boost SVR performances but the error distribution is still wider (higher variance) than the reference one on T1DM study [38].

## 5.1 Future Work

In future work, it could be interesting to deepen the method of time-stamp labeling and find a much wider range of applications for the technique. It would be advisable to verify with the sliding window model whether other regression algorithms are able to obtain less error or a narrow distribution. Finally, it is important to build a dataset made up of a much larger population and made available to the scientific community to remove barriers from any kind of research in this field.

# Bibliography

[1] Consensus statement on self-monitor ing of blood glucose. *Diabetes Care*, **10** (1987), 95. Available from: https://care.diabetesjournals.org/content/10/1/95, arXiv:https://care.diabetesjournals.org/content/10/1/95.full.pdf, doi:10.2337/diacare.10.1.95.

[2] Implications of the diabetes control and complications trial. *Diabetes Care*, **25** (2002), s25. Available from: https://care.diabetesjournals.org/content/25/suppl_1/s25, arXiv:https://care.diabetesjournals.org/content/25/suppl_1/s25.full.pdf, doi:10.2337/diacare.25.2007.S25.

[3] Standards of medical care in diabetes—2010. *Diabetes Care*, **33** (2010), S11. Available from: https://care.diabetesjournals.org/content/33/Supplement_1/S11, arXiv:https://care.diabetesjournals.org/content/33/Supplement_1/S11.full.pdf, doi:10.2337/dc10-S011.

[4] Allen, N. A., Fain, J. A., Braun, B., and Chipkin, S. R. Continuous glucose monitoring in non–insulin-using individuals with type 2 diabetes: Acceptability, feasibility, and teaching opportunities. *Diabetes Technology & Therapeutics*, **11** (2009), 151. PMID: 19216684. Available from: https://doi.org/10.1089/dia.2008.0053, arXiv:https://doi.org/10.1089/dia.2008.0053, doi:10.1089/dia.2008.0053.

[5] Arthur, D. and Vassilvitskii, S. K-means++: The advantages of careful seeding. *Society for Industrial and Applied Mathematics*, (2007), 1027–1035.

[6] Bishop, C. M. *Pattern recognition and machine learning.* Springer (2006). ISBN 978-03-87310-73-2. Available from: https://login.ezproxy.uniroma1.it/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=cat05438a&AN=sapienza.RMS1332495&lang=it&site=eds-live&scope=site.

[7] Bogardus, C., Lillioja, S., Mott, D., Hollenbeck, C., and Reaven, G. Relationship between degree of obesity and in vivo insulin action in man. *The American Journal of Physiology*, **248** (1985). doi:10.1152/ajpendo.1985.248.3.e286.

[8] Caliński, T. and Harabasz, J. A dendrite method for cluster analysis. *Communications in Statistics*, **3** (1974), 1. Available from: https://www.tandfonline.com/doi/abs/10.1080/03610927408827101, arXiv:

https://www.tandfonline.com/doi/pdf/10.1080/03610927408827101, doi:10.1080/03610927408827101.

[9] Davies, D. L. and Bouldin, D. W. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-1** (1979). doi:10.1109/TPAMI.1979.4766909.

[10] di Statistica [ISTAT], I. N. (2016).

[11] Fisher, L., Polonsky, W., Parkin, C. G., Jelsovsky, Z., Amstutz, L., and Wagner, R. S. The impact of blood glucose monitoring on depression and distress in insulin-naïve patients with type 2 diabetes. *Current Medical Research and Opinion*, **27** (2011), 39 . Available from: https://login.ezproxy.uniroma1.it/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2011-27853-006&lang=it&site=eds-live&scope=site.

[12] Fox, I., Ang, L., Jaiswal, M., Pop-Busui, R., and Wiens, J. Deep multi-output forecasting: Learning to accurately predict blood glucose trajectories. (2018). Available from: https://login.ezproxy.uniroma1.it/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edsarx&AN=edsarx.1806.05357&lang=it&site=eds-live&scope=site.

[13] Frandes, M., Timar, B., Timar, R., and Lungeanu, D. Chaotic time series prediction for glucose dynamics in type 1 diabetes mellitus using regime-switching models. *Scientific Reports*, **7** (2017). doi:10.1038/s41598-017-06478-4.

[14] Frontoni, S., Di Bartolo, P., Avogaro, A., Bosi, E., Paolisso, G., and Ceriello, A. Glucose variability: An emerging target for the treatment of diabetes mellitus. *Diabetes Research and Clinical Practice*, **102** (2013), 86 . Available from: http://www.sciencedirect.com/science/article/pii/S0168822713003227, doi:https://doi.org/10.1016/j.diabres.2013.09.007.

[15] Group, U. P. D. S. U.k. prospective diabetes study 16: Overview of 6 years therapy of type ii diabetes: A progressive disease. *Diabetes*, **44** (1995), 1249. Available from: https://diabetes.diabetesjournals.org/content/44/11/1249, arXiv:https://diabetes.diabetesjournals.org/content/44/11/1249.full.pdf, doi:10.2337/diab.44.11.1249.

[16] Harris, C. R., et al. Array programming with NumPy. *Nature*, **585** (2020), 357. Available from: https://doi.org/10.1038/s41586-020-2649-2, doi:10.1038/s41586-020-2649-2.

[17] Inc., P. T. Collaborative data science (2015). Available from: https://plot.ly.

[18] Irace, C., Fiorentino, R., Carallo, C., Scavelli, F., and Gnasso, A. Exenatide improves glycemic variability assessed by continuous glucose

monitoring in subjects with type 2 diabetes. *Diabetes Technology & Therapeutics*, **13** (2011), 1261. PMID: 21751893. Available from: https://doi.org/10.1089/dia.2011.0096, arXiv:https://doi.org/10.1089/dia.2011.0096, doi:10.1089/dia.2011.0096.

[19] Karl, P. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Breakthroughs in Statistics : Methodology and Distribution*, (2011), 11. Available from: https://login.ezproxy.uniroma1.it/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edssjb&AN=edssjb.978.1.4612.4380.9.2&lang=it&site=eds-live&scope=site.

[20] Klonoff, D., Buckingham, B., Christiansen, J., Montori, V., Tamborlane, W., Vigersky, R., and Wolpert, H. Continuous glucose monitoring: An endocrine society clinical practice guideline. *Journal of Clinical Endocrinology and Metabolism*, **96** (2011), 2968. doi:10.1210/jc.2010-2756.

[21] Kovatchev, B., Breton, M., and Clarke, W. Chapter 3 analytical methods for the retrieval and interpretation of continuous glucose monitoring data in diabetes. *Methods in Enzymology*, **454** (2009), 69 . Available from: https://login.ezproxy.uniroma1.it/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edselp&AN=S0076687908038032&lang=it&site=eds-live&scope=site.

[22] Kovatchev, B. P., Cox, D. J., Gonder-Frederick, L. A., and Clarke, W. Symmetrization of the blood glucose measurement scale and its applications. *Diabetes care*, **20** (1997), 1655 . Available from: https://login.ezproxy.uniroma1.it/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=cmedm&AN=9353603&lang=it&site=eds-live&scope=site.

[23] Lahouar, A. and Ben Hadj Slama, J. Day-ahead load forecast using random forest and expert input selection. *Energy Conversion and Management*, **103** (2015), 1040 . Available from: https://login.ezproxy.uniroma1.it/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edselp&AN=S0196890415006925&lang=it&site=eds-live&scope=site.

[24] Leskovec, J., Ullman, J. D., and Rajaraman, A. *Mining of massive datasets.* Cambridge University press (2014). ISBN 978-11-07-07723-2. Available from: https://login.ezproxy.uniroma1.it/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=cat05438a&AN=sapienza.RMS3204217&lang=it&site=eds-live&scope=site.

[25] Li, K., Liu, C., Zhu, T., Herrero, P., and Georgiou, P. Glunet: A deep learning framework for accurate glucose forecasting. *IEEE Journal of Biomedical and Health Informatics, Biomedical and Health Informatics, IEEE Journal of, IEEE J. Biomed. Health Inform*, **24** (2020), 414 . Available from: https://login.ezproxy.uniroma1.it/login?url=https:

//search.ebscohost.com/login.aspx?direct=true&db=edseee&AN=edseee.8779644&lang=it&site=eds-live&scope=site.

[26] MacQueen, J. Some methods for classification and analysis of multivariate observations. *University of California Press*, (1967), 281. Available from: https://projecteuclid.org/euclid.bsmsp/1200512992.

[27] Marfella, R., Barbieri, M., Grella, R., Rizzo, M., Nicoletti, G., and Paolisso, G. Effects of vildagliptin twice daily vs. sitagliptin once daily on 24-hour acute glucose fluctuations. *Journal of Diabetes and its Complications*, **24** (2010), 79. doi:10.1016/j.jdiacomp.2009.01.004.

[28] Mori, Y., Taniguchi, Y., Matsuura, K., Sezaki, K., Yokoyama, J., and Utsunomiya, K. Effects of sitagliptin on 24-h glycemic changes in japanese patients with type 2 diabetes assessed using continuous glucose monitoring. *Diabetes Technology & Therapeutics*, **13** (2011), 699. PMID: 21504334. Available from: https://doi.org/10.1089/dia.2011.0025, arXiv:https://doi.org/10.1089/dia.2011.0025, doi:10.1089/dia.2011.0025.

[29] Mori, Y., Taniguchi, Y., Sezaki, K., Yokoyama, J., and Utsunomiya, K. Liraglutide narrows the range of circadian glycemic variations in japanese type 2 diabetes patients and nearly flattens these variations in drug-naive type 2 diabetes patients: A continuous glucose monitoring–based study. *Diabetes Technology & Therapeutics*, **13** (2011), 1139. PMID: 21877924. Available from: https://doi.org/10.1089/dia.2011.0137, arXiv:https://doi.org/10.1089/dia.2011.0137, doi:10.1089/dia.2011.0137.

[30] NCD Management-Screening, D. and Treatment. Global report on diabetes. (2016).

[31] Norris, S. L., et al. Increasing diabetes self-management education in community settings: A systematic review. *American Journal of Preventive Medicine*, **22** (2002), 39 . doi:https://doi.org/10.1016/S0749-3797(02)00424-5.

[32] Organizaiton, W. H. Classification of diabetes mellitus 2019. (2019).

[33] Oviedo, S., Contreras, I., Quirós, C., Giménez, M., Conget, I., and Vehi, J. Risk-based postprandial hypoglycemia forecasting using supervised learning. *International Journal of Medical Informatics*, **126** (2019), 1 . doi:https://doi.org/10.1016/j.ijmedinf.2019.03.008.

[34] pandas development team, T. pandas-dev/pandas: Pandas (2020). Available from: https://doi.org/10.5281/zenodo.3509134, doi:10.5281/zenodo.3509134.

[35] Pedregosa, F., et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, **12** (2011), 2825.

[36] Peyser, T., Balo, A., Buckingham, B., Hirsch, I., and Garcia, A. Glycemic variability percentage: A novel method for assessing glycemic variability from continuous glucose monitor data. *Diabetes Technology and Therapeutics*, **20** (2017). doi:10.1089/dia.2017.0187.

[37] Rodbard, D. Continuous glucose monitoring: A review of successes, challenges, and opportunities. *DIABETES TECHNOLOGY And THERAPEUTICS*, **18** (n.d.), 3 . Available from: https://login.ezproxy.uniroma1.it/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edswsc&AN=000368602900006&lang=it&site=eds-live&scope=site.

[38] Rodríguez Ignacio, R. J.-V. M. M. G. M. Z.-I. M., Chatzigiannakis Ioannis. Utility of big data in predicting short-term blood glucose levels in type 1 diabetes mellitus through machine learning techniques. *Sensors*, **19** (2019), 4482. doi:10.3390/s19204482.

[39] Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, **20** (1987), 53 . Available from: http://www.sciencedirect.com/science/article/pii/0377042787901257, doi:https://doi.org/10.1016/0377-0427(87)90125-7.

[40] Service, F. J. Glucose variability. *Diabetes*, **62** (2013), 1398. Available from: https://diabetes.diabetesjournals.org/content/62/5/1398, arXiv:https://diabetes.diabetesjournals.org/content/62/5/1398.full.pdf, doi:10.2337/db12-1396.

[41] Service, F. J., Molnar, G. D., Rosevear, J. W., Ackerman, E., Gatewood, L. C., and Taylor, W. F. Mean amplitude of glycemic excursions, a measure of diabetic instability. *Diabetes*, **19** (1970), 644. Available from: https://diabetes.diabetesjournals.org/content/19/9/644, arXiv:https://diabetes.diabetesjournals.org/content/19/9/644.full.pdf, doi:10.2337/diab.19.9.644.

[42] Skinner, T. Psychological barriers. *European Journal of Endocrinology Eur J Endocrinol*, **151** (01 Oct. 2004), T13 . Available from: https://eje.bioscientifica.com/view/journals/eje/151/Suppl_2/T13.xml.

[43] Smola, A. J. and Schölkopf, B. A tutorial on support vector regression (2004).

[44] Smola, S. B., A.J. A tutorial on support vector regression. *Statistics and Computing*, (2004). doi:https://doi.org/10.1023/B:STCO.0000035301.49549.88.

[45] Stumvoll, M., Goldstein, B., and van Haeften, T. Type 2 diabetes: principles of pathogenesis and therapy. *Lancet*, **365** (2005), 1333. doi:10.1016/s0140-6736(05)61032-x.

[46] Tin Kam, H. Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition, Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*, **1** (1995), 278. Available from: https://login.ezproxy.uniroma1.it/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edseee&AN=edseee.598994&lang=it&site=eds-live&scope=site.

[47] Wang, J.-S., Lin, S.-D., Lee, W.-J., Su, S.-L., Lee, I.-T., Tu, S.-T., Tseng, Y.-H., Lin, S.-Y., and Sheu, W. Effects of acarbose versus glibenclamide on glycemic excursion and oxidative stress in type 2 diabetic patients inadequately controlled by metformin: A 24-week, randomized, open-label, parallel-group comparison. *Clinical therapeutics*, **33** (2011), 1932. doi:10.1016/j.clinthera.2011.10.014.

[48] Wes McKinney. Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference* (edited by Stéfan van der Walt and Jarrod Millman), pp. 56 – 61 (2010). doi:10.25080/Majora-92bf1922-00a.

[49] Wójcicki, J. M. "j"-index. a new proposition of the assessment of current glucose control in diabetic patients. *Hormone and metabolic research = Hormon- und Stoffwechselforschung = Hormones et metabolisme*, **27** (1995), 41 . Available from: https://login.ezproxy.uniroma1.it/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=cmedm&AN=7729793&lang=it&site=eds-live&scope=site.

[50] Zick, R., Petersen, B., Richter, M., and Haug, C. a. Comparison of continuous blood glucose measurement with conventional documentation of hypoglycemia in patients with type 2 diabetes on multiple daily insulin injection therapy. *Diabetes Technology & Therapeutics*, **9** (2007), 483. PMID: 18034602. Available from: https://doi.org/10.1089/dia.2007.0230, arXiv:https://doi.org/10.1089/dia.2007.0230, doi:10.1089/dia.2007.0230.

[51] Zimmet, P., Alberti, G., and Shaw, J. Global and societal implications of the diabetes epidemic. *Nature*, **414** (2002), 782. doi:10.1038/414782a.