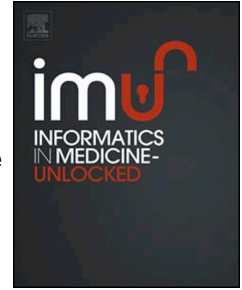


# Accepted Manuscript



The open D1NAMO dataset: A multi-modal dataset for research on non-invasive type 1 diabetes management

Fabien Dubosson, Jean-Eudes Ranvier, Stefano Bromuri, Jean-Paul Calbimonte, Juan Ruiz, Michael Schumacher

PII: S2352-9148(18)30105-9

DOI: [10.1016/j.imu.2018.09.003](https://doi.org/10.1016/j.imu.2018.09.003)

Reference: IMU 132

To appear in: *Informatics in Medicine Unlocked*

Received Date: 16 May 2018

Revised Date: 19 September 2018

Accepted Date: 26 September 2018

Please cite this article as: Dubosson F, Ranvier J-E, Bromuri S, Calbimonte J-P, Ruiz J, Schumacher M, The open D1NAMO dataset: A multi-modal dataset for research on non-invasive type 1 diabetes management, *Informatics in Medicine Unlocked* (2018), doi: <https://doi.org/10.1016/j.imu.2018.09.003>.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# The Open D1NAMO Dataset: a Multi-modal Dataset for Research on Non-invasive Type 1 Diabetes Management

Fabien Dubosson<sup>a,\*</sup>, Jean-Eudes Ranvier<sup>b</sup>, Stefano Bromuri<sup>c,d</sup>, Jean-Paul  
Calbimonte<sup>a</sup>, Juan Ruiz<sup>e</sup>, Michael Schumacher<sup>a</sup>

<sup>a</sup>*University of Applied Sciences and Arts Western Switzerland, Sierre, Switzerland*

<sup>b</sup>*EPFL, Lausanne, Switzerland*

<sup>c</sup>*Open University of the Netherlands, Heerlen, Netherlands*

<sup>d</sup>*BISS institute, Smart Service Campus, Heerlen, Netherlands*

<sup>e</sup>*Hôpital Riviera-Chablais, Vevey, Switzerland*

---

## Abstract

The usage of wearable devices has gained popularity in the latest years, especially for health-care and well being. Recently there has been an increasing interest in using these devices to improve the management of chronic diseases such as diabetes. The quality of data acquired through wearable sensors is generally lower than what medical-grade devices provide, and existing datasets have mainly been acquired in highly controlled clinical conditions. In the context of the *DINAMO* project — aiming to detect glycemic events through non-invasive ECG pattern analysis — we elaborated a dataset that can be used to help developing health-care systems based on wearable devices in non-clinical conditions. This paper describes this dataset, which was acquired on 20 healthy subjects and 9 patients with type-1 diabetes. The acquisition has been made in real-life conditions with the *Zephyr BioHarness 3* wearable device. The dataset consists of *ECG*, *breathing*, and *accelerometer* signals, as well as *glucose* measurements and annotated *food pictures*. We open this dataset to the scientific community in order to allow the development and evaluation of diabetes management algorithms.

*Keywords:* Diabetes Management; Wearable Devices; Glucose; ECG; Accelerometers; Annotated food pictures

---

## 1. Introduction

Data science and data-driven algorithms are transforming the way health-care and chronic diseases are managed, including the use of monitoring, diag-

---

\*Corresponding author

*Email address:* [fabien.dubosson@hevs.ch](mailto:fabien.dubosson@hevs.ch) (Fabien Dubosson)

*URL:* [www.fabiendubosson.ch](http://www.fabiendubosson.ch) (Fabien Dubosson)

5 nosis and treatment technologies. This trend is supported by the increasing adoption of wearable devices with sensing capabilities, such as *Fitbit One*<sup>1</sup>, *Jawbone Up*<sup>2</sup> or *Garmin Vivofit*<sup>3</sup> [1]. Thanks to such devices, it is now possible to provide live analytics on personal health data, leveraging state-of-the-art research on quantified-self and health-tracking [2].

10 Diabetes management is a particularly interesting use-case in this context, given the importance of providing accurate and timely feedback, monitoring and diagnosis tools to both patient and health-care providers [3, 4]. In fact, several research initiatives have addressed the challenges of managing the different aspects of diabetes, using wearable devices as sources of data [5, 6]. However, the development of data-driven approaches depend on the existence of datasets that  
15 can be used for training, validation or automated learning. These datasets are needed not only in clinical and controlled environments (e.g. monitoring during a hospital stay), but also in everyday-life conditions, which are closer to what a patient with diabetes experiences in her daily routine.

The goal of this paper is to contribute to the development of data-centric algorithms and diabetes monitoring technologies, by providing an openly available  
20 multi-modal dataset acquired on real patients in non-clinical conditions. This dataset has been collected in the context of the *DINAMO* project, and consists of a set of *Electrocardiogram (ECG)* signals, *Breathing* signals, *Accelerometers* outputs, information about *glucose levels* and *annotated food pictures*. The  
25 *DINAMO* project aims at providing non-invasive diabetes management through machine learning data analytics performed on wearable sensor data signals [7, 8]. This heterogeneous dataset was acquired on a study that included 29 patients — 20 healthy people and 9 patients with diabetes — in real-life conditions, using the *Zephyr BioHarness 3* wearable chest-belt device.

30 As explained in Section 5, the availability of datasets such as this, offers interesting possibilities for the development of healthcare monitoring systems, especially in the context of diabetes management through wearable devices. For example, one possible use case is for detection of hypoglycemic events, which can be potentially dangerous for the patient. Changes in glucose level in the  
35 blood have been shown to correlate with changes in the shape of the ECG signals, opening the door for non-intrusive detection. Nevertheless, this is not the only parameter to consider, given that the activity levels, or food intake, may have direct and indirect consequences on glyceic events on the patient. Added to the complexity of dealing with these different parameters, it is important  
40 to take into account the potentially noisy data that is commonly acquired under non-clinical conditions. Many algorithms that were originally developed using medical-grade signals require to be tested and adapted for everyday-life wearable devices signals, in order to be used effectively for health monitoring. Therefore, the purpose of this paper is to open this dataset to the scientific com-

---

<sup>1</sup><http://www.fitbit.com/>

<sup>2</sup><http://www.jawbone.com/>

<sup>3</sup><http://www.garmin.com/>

45 munity, providing a valuable source of real patient information, which can be  
 used for automated learning, training, experimentation and validation of algo-  
 rithms and techniques in the realm of diabetes management. The publication of  
 this dataset constitutes an important milestone and contribution in the context  
 of the *DINAMO* project, following our previous efforts regarding data analyt-  
 50 ics [8] and mobile health developments [7] for Type 1 diabetes management.

The remaining of this paper is structured as follows: Section 2 compares  
 our dataset to relevant related works. Section 3 presents the methodology used  
 for the data acquisition. Statistics of the dataset are described in Section 4.  
 Different opportunities offered by this dataset are proposed in Section 5, and  
 55 finally we provide final remarks in Section 6.

## 2. Related work

Research in diabetes has gathered considerable attention, especially regard-  
 ing the development of techniques and algorithms relying on monitoring pa-  
 rameters using sensing devices. To support this research line, datasets have  
 60 been released to allow researchers to advance the state of the art with respect  
 to diabetes monitoring and treatment. While a lot of these publicly released  
 datasets contain ECG waveforms, we have found only one that also contains  
 glycemic data: the MIMIC II [9] dataset. This dataset is available on the Phys-  
 ioNet platform [10] and provides data from patients who were in Intensive Care  
 65 Unit (ICU). The database has two parts: first, the clinical database containing  
 information about the patients, their diseases, and other ICU-related data. The  
 second part is the waveform database: it contains the signals that were recorded  
 during the patients' stays at ICU using medical-grade devices, while patients  
 were in a static position (i.e. lying in bed, with very limited movement). Despite  
 70 the good signal quality and the additional type of data available, the dataset is  
 not representative of everyday signals recorded with wearable devices, because  
 of the ICU-related condition of the patients and of the quality of the sensing  
 devices. Also, the glucose levels are not available for all patients but have only  
 been acquired when the patient's condition required it.

75 The PhysioNet platform also provides a long list of physiological databases.  
 Some of them are ECG databases, like for instance the MIT-BIH arrhythmia  
 dataset [11] which offers 48 half-hour of ambulatory acquired ECG. Others  
 are multi-parametric databases that include ECG along with other signals, but  
 among all the databases listed on the webpage<sup>4</sup>, none of them includes glucose  
 80 levels as time series. Another known resource for Machine Learning databases  
 is the *UCI Machine Learning Repository* [12]. At least two of its datasets are  
 related to this work. First, the *MHEALTH* dataset, which offers body motion  
 data — including acceleration — as well as 2-leads ECG measurements [13].  
 Second, the *Diabetes* dataset that consists of glucose levels acquired both from

---

<sup>4</sup><http://www.physionet.org/physiobank/database/> - November 2017

85 a Continuous Glucose Monitoring (CGM) and from manual measurements<sup>5</sup>.  
 These two datasets do not offer all the type of data needed for working on  
 diabetes management through wearable devices, as they either lack glucose in-  
 formation in the first dataset, and ECG in the second.

The *Diabetic Cardiac Neuropathy Diagnostic and Modeling* dataset [14, 15]  
 90 (DICARDIA<sup>6</sup>) provides ECG from patients with diabetes, both with and with-  
 out cardiac complications. Additional laboratory data such as glucose, hemoglobin  
 or cholesterol are available, but only with a unique spot measurement per pa-  
 tient. Therefore, predictions of glucose levels or identification of glycemic events  
 cannot be done with this dataset. Other research has also been made on topics  
 95 requiring both glucose levels and ECG signals, but these datasets have not been  
 made publicly available. Some examples of such works includes [16, 17, 18, 19].

The D1NAMO dataset focuses on diabetes management through ECG, but  
 the other types of data it provides have also been treated by the scientific com-  
 munity. The domain of activity recognition is mainly built upon Accelerometers  
 and have released many datasets, for instance [20, 21]. There exist picture data-  
 bases that are oriented towards food recognition, such as *UEC FOOD-100* [22],  
 100 *UEC FOOD-256* [23], *Food-101* [24] and *Food-pics* [25]. The D1NAMO dataset  
 has less images in comparison, but the annotations by a dietitian are valuable  
 in the context of calories estimation as it impacts glucose levels. Finally, up to  
 105 our knowledge there is no other publicly accessible dataset offering a collection  
 of signals coming from the *Zephyr BioHarness 3*.

Dataset	ECG	Glucose	Activity	Food Intake	Measurement conditions	Target
MIMIC II [9]	✓	limited	✗	✗	UCI conditions	Monitoring UCI patients
MIT-NIH [11]	✓	✗	✗	✗	Ambulatory	Monitoring ambulatory patients
ECG Physionet datasets	✓	✗	✗	✗	Heterogeneous conditions	Various
UCI MHEALTH [13]	✓	✗	✓	✗	Everyday life	Mobile healht
UCI Diabetes	✗	✓	✗	✗	Everyday life	diabetes man-agement
DICARDIA [14]	✓	limited	✗	✗	Everyday life	Cardiac pa-tients
D1NAMO	✓	✓	✓	✓	Everyday life	Diabetes man-agement

Table 1: Existing research datasets related to Diabetes management, including ECG, and glucose measurements.)

<sup>5</sup><http://archive.ics.uci.edu/ml/datasets/Diabetes>

<sup>6</sup><http://gbbanet.labc.usb.ve/DICARDIA.php>

### 3. Methodology

The D1NAMO dataset is composed of two distinct subsets: a first one that has been acquired on 20 healthy people, and a second one that has been acquired on 9 patients with diabetes. Having data from both populations offers possibilities in terms of comparison, exploration and impact of Type 1 diabetes with a reference group. The methodologies to acquire both subsets were quite similar, the same wearable device has been used, but a few differences exist though, notably concerning the acquisition of glucose levels.

Both subsets contain the following data:

- ECG signals
- Breathing signals
- Accelerometer outputs
- Glucose measurements (the method differs between subsets)
- Food pictures and annotations by a dietitian

Subsection 3.1 details the setup that has been used to acquire these signals and data and the subsection 3.2 describes the protocol that has been followed for the acquisition in both groups.

#### 3.1. Acquisition setup

The acquisition of physiological signals: ECG, Breathing and Accelerometers outputs has been made using the *Zephyr Bioharness 3* device for both subsets. This device looks like a sport chest belt as shown in Figure 1. It has 3 electronic sensors: one measuring ECG with two electrodes; one measuring breathing through chest expansion; and one measuring accelerations in the three dimensions. These raw signals can be used directly and the device also computes additional aggregated metrics such as Heart Rate (HR), Breathing Rate (BR), activity level, posture, etc.



Figure 1: The Zephyr BioHarness 3 used for the data acquisition phase for the D1NAMO dataset [7].

The electrocardiogram of the *Zephyr BioHarness 3* is taken with a I-lead sensor, consisting of two silver-coated nylon electrodes working on skin contact.

135 It measures the ECG at a rate of 250 Hz and up to 54.89 mV. The breathing is measured with an Ethylene-vinyl acetate (EVA) foam pressure sensor at 18 Hz. This measurement has no unit, it should be used as a relative value. Accelerometers measure the accelerations in vertical, lateral and sagittal directions at a rate of 100 Hz. These output values are given within the range of  $\pm 16g$ .

140 The measurement of glucose levels for healthy people has been made 6 times per day using the *Bayer Contour XT* glucose meter with *Bayer Next strips*. To get a drop of blood, participants used the *Microlet 2 Lancing Device* with their related *Microlet Coloured Lancets*. Glucose levels of patients with diabetes have been recorded with an *iPro2 Professional CGM* sensor, allowing for a precision  
145 of 5 minutes between two glucose measurements.

For the food information, people were asked to take pictures of all their meals with their smartphone. We asked to put a 1 Swiss Franc coin on the picture in order to have a notion of scale (size of the coin: 23.20 mm). The pictures have then been annotated by a dietitian with a description of the food, an estimation  
150 of the number of calories, and an indication about the *Food Balance* and *Food Quality*. The later concepts are described by the dietitian as follows:

- **Food Balance** corresponds to the percentage distribution of macronutrients (carbohydrates, proteins and lipids) in a meal according to the Nutritional composition table from the Swiss Nutrition Society. A “Balanced meal” respects the optimal distribution of macronutrients according  
155 to this document and an “Unbalanced meal” does not.
- **Food Quality** is related to the amount of food items that should be ideally sporadically consumed in a meal: processed food, alcohol, saturated and trans fats, sweet or salty snacks. A “High food quality meal” contains mostly unprocessed food items and/or includes low quantities of the  
160 aforementioned food items. A “Low food quality meal” contains processed and/or the aforementioned food items.

### 3.2. Acquisition protocol

165 The protocol has been divided in two main arms, one for patients with diabetes and another for healthy ones. In the following we provide details about enrollment, inclusion and exclusion criteria, and the protocol description.

*Data collection and equipment:*. Information collected at patient enrollment is anthropometrical (weight, height, sex, seated office blood pressure). Patients were provided with the following devices after enrollment: BioHarness sensor,  
170 BioHarness belt, BioHarness charger and cable. Healthy patients have been provided with a glucose meter. For patients with diabetes, glucose levels were measured using their continuous glucose monitoring (CGM) device. Participants were instructed to use their mobile phones for taking a picture of their meals.



*Recruitment of patients with diabetes:* Patients were recruited at routine clinic visits at the diabetes outpatient clinic of Hpital Riviera-Chablais, Vevey. All patients were screened for data acquisition enrollment. Exhaustive and detailed information about the data acquisition, its scope and objectives, its duration, its risks and benefits and its execution were provided to each and every patient. Patients willing to participate, signed a consent form. Eligibility for the diabetes arm was established according to the following inclusion criteria: patients should have a definite diagnosis of diabetes mellitus type 1, followed at the outpatient clinic. Patients must be able to speak French and be at least 18 years old. Furthermore, patients must provide written informed consent and be willing to participate in follow-up. Excluded from the study are patients who do not fulfill the diagnostic criteria for diabetes mellitus type 1. Patients that do not speak French and those not possessing the cognitive capacities to handle a smart phone are equally not eligible for study inclusion. All patients unwilling to provide written informed consent or to participate in follow-up were excluded from study enrollment. Furthermore, patients with significant psycho-social impairment that impedes on their ability to participate in follow-up such as homelessness, psychotic episodes, etc. were excluded.

All consent forms, information sheets, and the entire protocol was validated and approved by the ethcical committee of the Canton of Vaud in Switzerland (Commission cantonale d'éthique de la recherche sur l'être humain).

*Protocol:* Patients received instructions to start wearing the BioHarness after waking up in the morning, and shutting it down before going to sleep. For the first day they started to wear it directly after the instructions. People were asked to stop the data acquisition after the fourth day and to bring back the devices and forms the day after. The usage of the glucose meter was explained to healthy patients, and they were asked to do it once by themselves to be sure they understood correctly. The protocol defined 6 glucose measures per day, one before each meal — breakfast, lunch, diner — and one 2 hours after the meal. Patients with diabetes were asked for participation in the study by their diabetologist, and upon agreement received instructions for the data acquisition directly from the nurses in diabetology, who had been trained beforehand. The protocol was almost identical to the healthy people subset: wearing the BioHarness during 4 days and taking pictures of their meals. However, and differently to the healthy group, the glucose levels have been acquired with CGM devices, which sample one glucose measurement every 5 minutes.

#### 4. Statistics

This section first describes the statistics of the full dataset, then presents the statistics individually for each of the subsets in the following subsections.

The full dataset has been acquired for 29 participants. Table 2 presents the statistics of the full dataset and of both subsets. The full dataset contains around 1550 hours of BioHarness measurements (for each kind of signals: 1550 hours of ECG, 1550 hours of Breathing and 1550 hours of Accelerometers).



	Total	Average per participant
<b>Full dataset</b>		
Participants	29	
Signals recording	~ 1550 h	~ 53 h
Glucose measurements	8884*	306.3*
Food annotations	473	16.3
Food pictures	352	12.1
<b>Healthy people subset</b>		
Participants	20	
Signals recording	~ 1100 h	~ 55 h
Glucose measurements	470	23.5
Food annotations	358	17.9
Food pictures	246	12.3
<b>Subset of patients with diabetes</b>		
Participants	9	
Signals recording	~ 450 h	~ 50 h
Glucose measurements	8414	934.9
Food annotations	115	12.8
Food pictures	106	11.8

\* Not representative because of the difference in glucose measurements between the two subsets

Table 2: Statistics of the dataset

The number of glucose measurements for the full dataset is not representative because of the difference in measurements methods between the two subsets. In addition, 352 food pictures and 473 food annotations complete the dataset.

220 Figure 2 shows a sample of each kind of data available in the dataset, except for the glycemic levels that are presented later. The samples shown here are from clean parts of the signals, probably acquired when people were sitting. Due to the fact that time series have been acquired with a chest belt, every movement made by the people affects the measurements, whether it is of ECG or Breathing.  
 225 This results in signals with both clean and noisy parts, and we also suspect that some people have worn the belt in a way that make the measurements mostly noisy. The dataset also has missing values: some people did not include pictures for all of their meals, or forgot to add a 1 Swiss Franc coin on some pictures. In other cases they forgot to write down measured glycemic values, or had problems  
 230 with the belt. We decided to keep such signals in the dataset, because these problems naturally occur in every healthcare system including wearable devices and people, so these issues would have to be dealt with.

#### 4.1. Healthy people subset

235 The participants agreed to share their data in an *anonymized* format with the scientific community. Therefore, we do not link their clinical information with the signals in the dataset in order to preserve their anonymity.

The population of the dataset consists of 20 healthy people from different countries: Switzerland(9), Italy(3), Spain(3), Russia(1), Austria(1), Mexico(1),

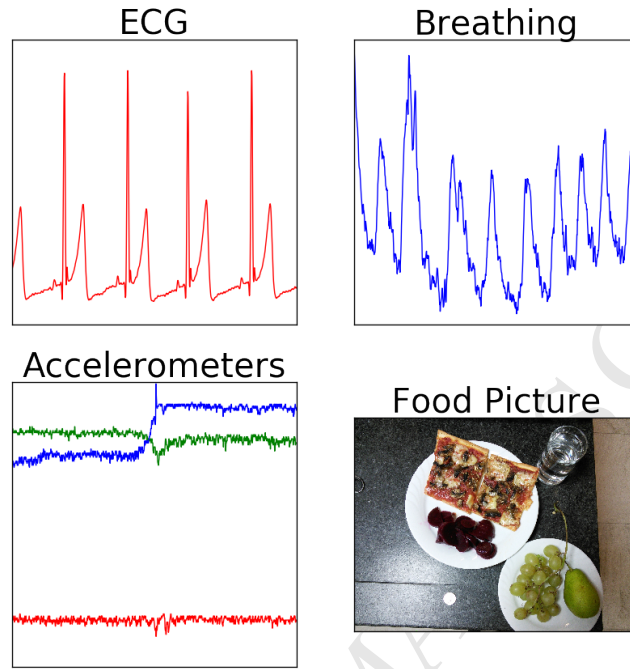


Figure 2: Data samples

Mauritius(1), and Britain(1). The participants clinical data are shown in Table 3. The dataset contains the following bias: All people are working in academic research, and are hence highly qualified people. We believe that this bias does not render the dataset less interesting, despite having a broader population would have been more valuable. Also, there is only 4 women for 16 men, and the ages range goes from 26 to 45. One person in this subset has diabetes type 1: he has been put in the healthy subset because he has not worn a CGM like other patients with diabetes, so the protocol he followed is the one of healthy subset. This person is not included in the glucose plots of this section, but his data are available in the dataset.

All measured glucose levels have timestamps indicating at which time of the day they were taken. Table 4 shows the per-patient statistics of glucose measurements. Figure 3 shows all the measures in one plot. In this plot the left-pointing green triangle represent the fasting glucose (just before eating) and the right-pointing blue triangle represent the 2 hours postload glucose levels (2 hours after eating). The dashed red lines represents the interpolated mean value of all measurements, and the blue area is the standard deviation around this mean. The dashed blue line represent a commonly used threshold for hypoglycemia [26, 27], showing that a few hypoglycemia events appear in this subset. Figure 4 and Figure 5 show the same data with the respective thresholds of the American Diabetes Association for fasting glucose and 2 hours postload glucose [28]. The

Age	Gender	Height (cm)	Weight (kg)
26	Woman	174	64
27	Man	184	75
28	Woman	170	54
29	Man	170	62
29	Man	190	83
31	Man	171	65
32	Woman	162	58
32	Man	170	72
32	Man	173	89
33	Man	170	78
33	Man	177	72
33	Man	185	90
33	Man	186	94
34	Woman	164	63
34	Man	185	87
34	Man	192	81
36	Man	170	73
36	Man	176	68
43	Man	176	74
45	Man	178	82

Table 3: Population of the Healthy subset (sorted by age)

260 green areas represent values that are considered as normal for fasting glucose, respectively for 2-h postload glucose. The orange area contains values that are considered as “impaired fasting glucose”, respectively “impaired glucose tolerance”, that are now referred as “pre-diabetes”. The red zones (not visible in Figure 5) correspond to levels that are subject to provisional diagnosis of

265 diabetes. Except for the patient with diabetes type 1 that is not shown in these plots, the other values are mostly in the normal range for the 2-h postload glucose. On the other side around 25% of fasting glucose are above the normal threshold. We suppose this is due to the fact that people are eating snacks in-between meals so it is not real fasting glucose.

270 Most of the glucose values are labelled with the moment they were taken: the 6 labels are: {Before, After}  $\times$  {Breakfast, Lunch, Diner}. This has been done in order to compensate the fact that people have different eating timing habits. The Figure 6 shows the distribution of glucose level for each label.

#### 4.2. Patients with diabetes subset

275 The patients agreed to share their data in an *anonymized* format with the scientific community. Therefore, we do not link their clinical information with the signals in the dataset in order to preserve their anonymity. Also, due to the

Patient ID	Mean	Std	Median	Count
001	5.39	0.73	5.70	18.00
002	5.45	0.38	5.40	24.00
003	5.24	0.48	5.30	21.00
004	6.21	0.86	5.90	24.00
005	5.42	0.70	5.20	20.00
006	5.97	0.72	5.95	24.00
007	5.80	0.93	5.60	31.00
008	4.93	0.76	4.80	23.00
009	5.35	0.63	5.30	24.00
010	6.11	0.60	5.90	23.00
011	4.86	0.70	4.80	23.00
012	12.31	4.40	11.75	30.00
013	5.37	0.70	5.20	22.00
014	5.47	1.04	5.00	23.00
015	5.32	0.43	5.30	23.00
016	5.30	0.57	5.20	24.00
017	6.03	1.11	5.40	24.00
018	5.63	0.57	5.60	24.00
019	5.53	0.66	5.55	22.00
020	5.63	0.73	5.50	23.00

Table 4: Individual glucose statistics of the Healthy Subset

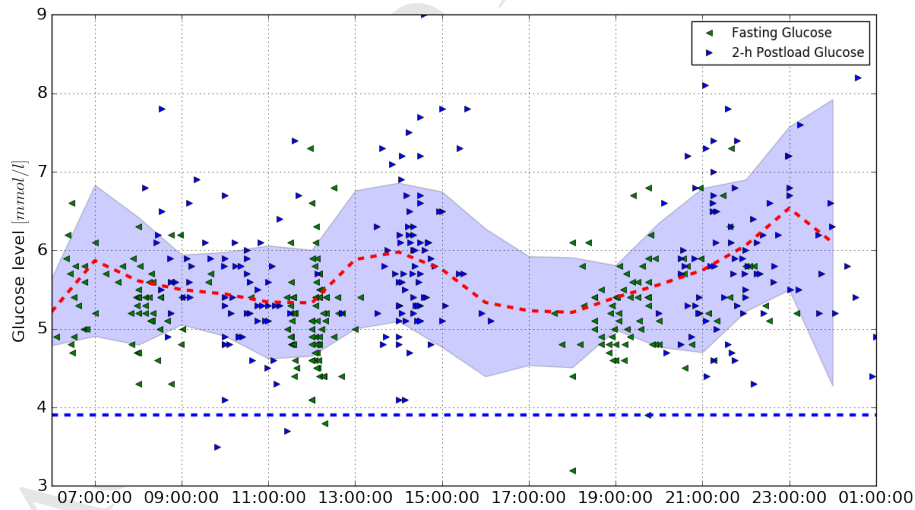


Figure 3: Glucose measures of the Healthy subset

relatively small size of this population, only the range of their clinical data are presented in Table 5.

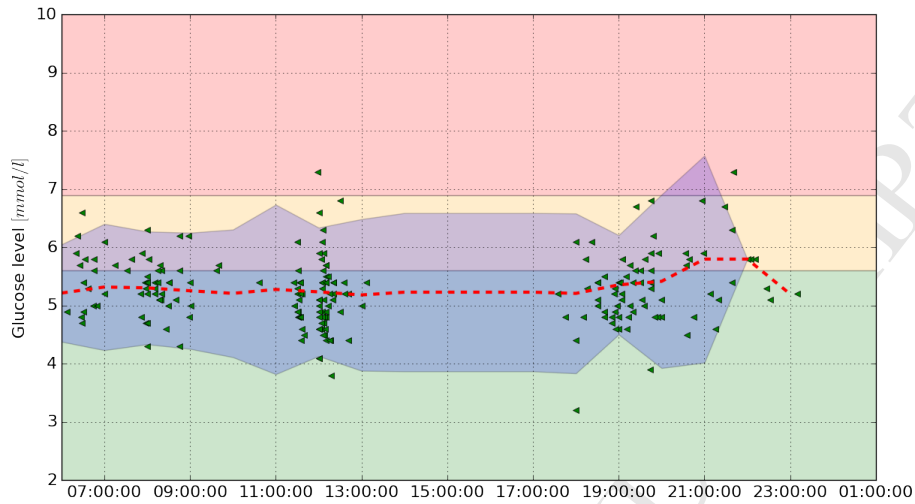


Figure 4: Fasting glucose measures of the Healthy Subset

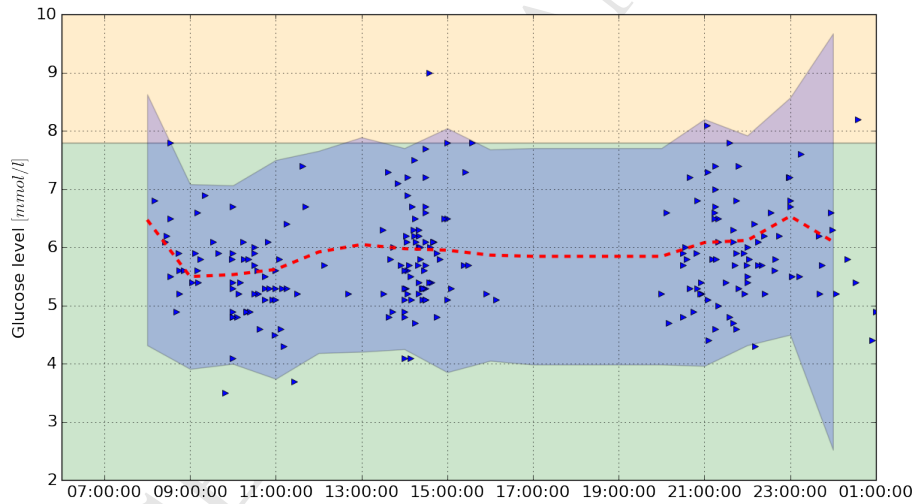


Figure 5: 2-hours postload glucose measures of the Healthy Subset

280 As shown in Table 2, the subset is composed of 9 patients with type 1 di-  
 abetes. 8414 measurements of glycemic levels have been made with the CGM,  
 but this covers the 24 hours of a day: only a part of these measurements can be  
 mapped to the daytime-only signals acquired with the sensors. Table 6 shows  
 the per-patient statistics of glucose measurements. Although the number of  
 285 patients with diabetes is limited, we computed the required number of hypo-  
 glycemia episodes needed to get statistically relevant results. In order to do  
 so, a calculation was made considering 3 main ECG parameters: ST-Segment

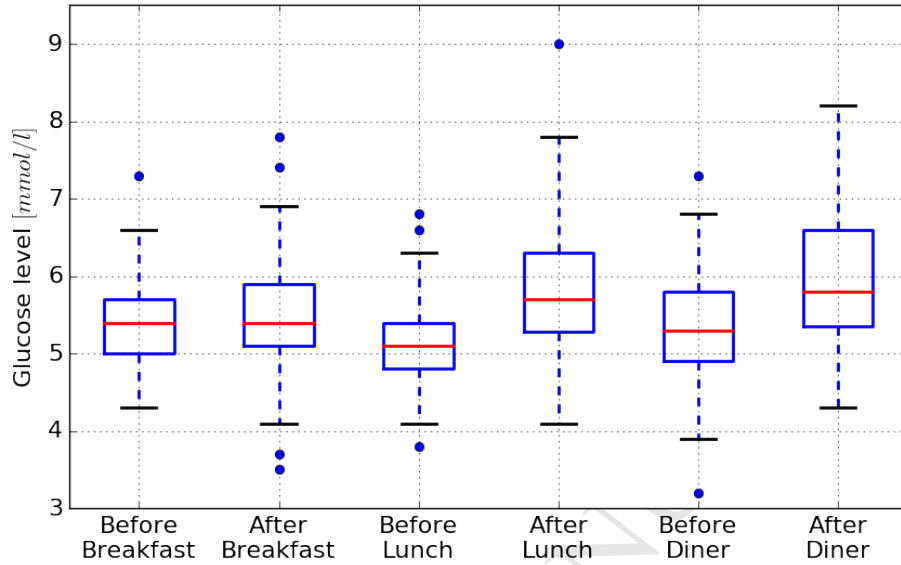


Figure 6: Glucose measures of the Healthy subset grouped by categories

changes, T-Wave amplitude changes, and length of the QT-interval. According to this calculation the recording of 18 hypoglycemic episodes would be sufficient to detect the differences published by [29] (with 80% statistical power).

Not all the patients have taken pictures of their meals, but people who have not done this have instead annotated what they have eaten. This is what has been used by the dietitian to annotate the food intake in such cases.

Age	Gender	Height (cm)	Weight (kg)
NA	Man	180-189	80-89
20-29	Man	170-179	60-69
20-29	Man	180-189	70-79
20-29	Man	180-189	80-89
30-39	Man	180-189	80-89
30-39	Man	190-199	70-79
30-39	Woman	160-169	70-79
60-69	Woman	150-159	50-59
70-79	Woman	160-169	50-59

Table 5: Patients of the Type-1 Diabetes subset (sorted by age)

Glucose measures acquired with the CGM are shown in Figure 7. The same thresholds as for Figures 4 and 5 are shown for comparison. The background on the half-left side corresponds to American Diabetes Association thresholds for fasting glucose, while the background on the half-right side correspond to

2-h postload glucose. The same blue dashed line as before shows that a few hypoglycemia also appear in this subset.

Patient ID	Mean	Std	Median	Count
001	10.19	4.87	9.60	1438.00
002	10.16	5.56	9.80	1071.00
003	7.13	1.85	6.95	186.00
004	11.20	4.98	10.80	984.00
005	8.55	2.49	8.30	928.00
006	9.44	3.19	9.20	1298.00
007	8.29	2.50	8.20	1011.00
008	7.18	3.19	7.00	1175.00
009	4.86	2.35	4.80	306.00

Table 6: Individual glucose statistics of the Type-1 Diabetes Subset

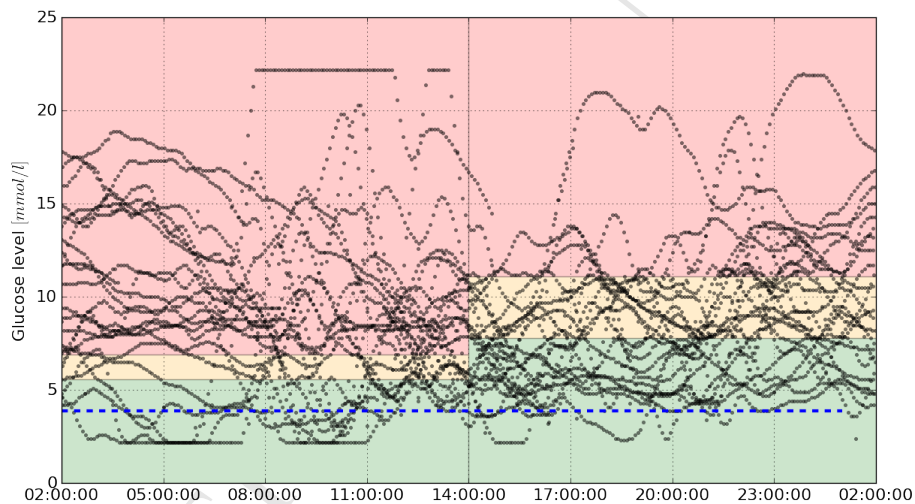


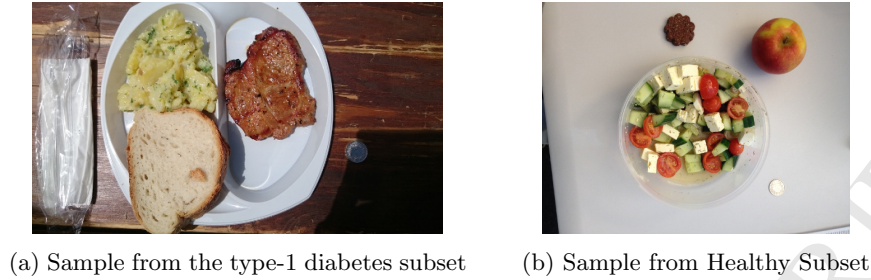
Figure 7: Glucose measures of the Type-1 Diabetes subset

### 300 4.3. Food Pictures

Food annotations are provided for pictures of the meals for both healthy and type-1 diabetes subsets. The pictures have been taken from patients' personal devices, meaning they have been taken from different devices in different places and conditions, as it can be expected in real-life datasets. The Figure 8 shows a sample of food pictures and their related remarks, as annotated by a dietitian.

305





(a) Sample from the type-1 diabetes subset (b) Sample from Healthy Subset

	Description	Calories	Balance*	Quality*
a	Pork potatoes salad and white bread	634	Not balanced	Medium quality
b	Greek salad, apple and chocolate cookie	677	Not balanced	Good quality

\* as defined in Section 3.1

Figure 8: Sample of the food pictures with the associated annotations

## 5. Discussion & Opportunities

For the treatment of diabetes, glycemia is one of, if not the most important parameter to monitor. Therefore, in addition to tracking activities, gathering information about glycemetic events is required in order to produce useful services for the patient. Signals acquired on the devices mentioned above, such as *Heart Rate*, *Skin Temperature* or *Steps Counter*, have already been addressed in the scientific literature. However, medical devices are still used for acquiring health-related signals such as ECG or glycemetic levels. The design of these sensors, for instance glued patches ECG or subcutaneous sample acquisition for glucose, make their usage constraining as opposed to the ease-of-use of wearable devices. Technological progress makes now possible to capture some health-related signals such as ECG on embedded wearable devices. However, the gain in comfort has been made in detriment of the signals quality: wearable devices do not offer signals with quality as good as medical-grade devices. Therefore, it is important to count with relevant datasets that include this type of conditions, which are more representative of realistic scenarios of patients during their daily activities. Using these datasets, algorithms can be designed, developed, trained, validated, etc. in order to cope with uses related to diabetes monitoring and management.

The DINAMO dataset provides interesting information that could be used in this context. Even if the data is reduced in terms of number of patients, it contains a considerable number of hours of recordings, which can be used in different scenarios. Most of these scenarios can be set up focusing on the different relationships and interactions among different parameters: activity levels, breathing and heart activity, food intake, and particularly, glycemetic events. Among glycemetic events, hypoglycemia episodes are especially relevant and potentially dangerous, as they have been shown to predict sudden death [30] and have been linked to the dead-in-bed [31] syndrome. It has been shown that these hypoglycemic events can be detected from the shape of ECG beats [32, 33, 34],

and despite the decrease in signals quality, some preliminary results show that  
 335 it is possible to detect glycemic events from wearable devices (e.g. such such  
 as the *Zephyr Bioharness 3*<sup>7</sup> [35, 8] in D1NAMO), opening a way to improve  
 diabetes management through wearable devices without the need of medical  
 sensors. The *D1NAMO* project has focused on studying this topic, relying on  
 machine learning algorithms to infer hypoglycemic events from the sensor data  
 340 signals [8, 7].

Although nowadays new alternative technologies are developed for continu-  
 ous and less invasive glycemic monitoring, there is still room for further devel-  
 opments and the adoption of these technologies by a wider public. Examples  
 345 of these devices include electrochemical impedance spectroscopy [36], Raman  
 spectroscopy [37], transdermal monitoring [38], among others [39]. Even if these  
 technologies are promising, direct glycemic measurements need to be combined  
 with other information sources in order to provide a holistic approach to diabetes  
 management. Food intake, activity levels, as well as heartbeat and breathing  
 350 monitoring can substantially improve prevention and early detection of hypo-  
 glycemic situations, as well as enable personalized recommendations targeting  
 behavioral change.

This dataset offers several opportunities for use in different scientific ac-  
 tivities and research directions, and permits to develop applications in several  
 355 domains. Next, we briefly mention some of these opportunities.

- The multimodal nature of the dataset allows to explore predictions of one  
 of the signals from the other ones. For instance, trying to predict the  
 HR from the *Breathing* signal or vice-versa. In the context of diabetes  
 management it could be used to explore the prediction of glucose levels  
 360 from ECG, Breathing and Accelerometers signals.
- The availability of multiple signals could permit to study relationships  
 between them, for instance the ECG variations under different types of  
 exercises or activities, or the impact of glucose level on breathing rate, if  
 any. Different algorithms based on machine learning or pattern mining  
 365 can be used for this purpose, leading to relevant correlations among the  
 sensor observations.
- The fact that the dataset has been acquired in real-life conditions, with a  
 wearable device, opens opportunities for the development and evaluation  
 of pre-processing algorithms such as for cleaning or denoising signals. Dif-  
 370 ferent filtering and pre-processing techniques may be required, depending  
 on the use-case. In some cases, excessive cleaning may lead to the elimina-  
 tion of apparently noisy signal features that may be relevant for further  
 processing tasks.

---

<sup>7</sup><http://www.zephyranywhere.com/>

- 375 • Another approach for working with noisy signals would be to discard parts that are not usable and to keep only the clean/usable portions. Such works can also be applied on this dataset, e.g. focusing only on time windows relevant to an activity, or a portion of the data where noise does not increase over a certain threshold.
- 380 • This dataset contains data from healthy people and patients with diabetes, therefore it permits to explore the differences between these two populations.
- 385 • The large number of hours of signals in the dataset allows to imagine its usages in the world of Deep Learning. Possible ideas include the usage of these data to create a model to recognize healthy heart beat shape, to identify normal breathing behaviors, or to predicate heart rate and breathing rate from Accelerometers. Usages in unsupervised learning are also possible, such as building networks for learning features on one-dimensional time series.
- 390 • Having food information for each patient as well as their physiological signals — especially glucose — could be used for diet advice. Advice can either be in regard to the global health condition or in regard to glucose levels for decreasing risks for type 1 diabetes patients.
- 395 • Glucose together with food information may permit to explore the impact of Calories, Food Balance or Food Quality on the Glucose levels.
- 400 • The food annotations could also be used in conjunction with pictures in food recognition algorithms. The number of pictures is not sufficient on itself to fully train models but it can be used in context of transfer learning. It can also be used as a test set, or just used to increase the number of samples in an existing dataset. Since the pictures have been annotated by a dietitian, they offer a reliable ground truth for verification or calibration.

The availability of this dataset, inspired by the current trends on Open Science, also has limitations, related both to the acquisition process and the data itself. First, given that the goal is to have data measurements for everyday-life conditions, the dataset inevitably includes noise and data artifacts that make it more challenging to perform analytics on it. This is both a feature and a limitation, linked to both the sensing conditions (e.g. arbitrary movements, heterogeneous activities), as well as the device limitations (e.g. accuracy). Second, the recruitment process had a certain degree of complexity, considering the limited number of diabetes type 1 patients willing, or able to accept the study conditions. In many cases, the co-morbidities of patients would make them ineligible, given the complications of their health status. In other cases, physical and cognitive impairments would place certain patients clearly outside the inclusion group. Third, and related to the previous point, the dataset has a limited number of participants, although this is compensated by the diversity of the data, as well as the number of hours or registered observations. Finally,

the dataset also has limitations considering the bias regarding age, background and geographical/socio-economic situation of the participants.

Another aspect that need to be considered is related to the usability, convenience and general feasibility of the approach, even beyond technical and algorithmic considerations. The fact of wearing the equipment, or using a mobile application that supports the data collection, monitoring and recommendations, needs to be explored from the point of view of the potential users. To that end, and in order to evaluate and improve these aspects, we conducted a qualitative evaluation on the ease of use of the Android Application developed for the *DINAMO* project, with 18 participants. The study main goal was to evaluate the ease of use of the Android application, but we were also interested in getting general feedback and comments about all aspects of the approach. This study provided us precious knowledge on what the application should look like, an how the application should behave from the user's perspective. Although this study is out of the scope of this dataset paper, it provided positive evidence on the feasibility of non-invasive hypoglycemic detection.

## 6. Conclusions

This work presents the *DINAMO* dataset that focuses on monitoring health, activity and food intake measurements in patients with and without diabetes type 1, under everyday-life conditions. This dataset been acquired in the context of the *DINAMO* project, which studied the feasibility of non-invasive monitoring of hypoglycemic events. The dataset consists of *ECG*, *Breathing* and *Accelerometers* signals, as well as *glucose* measurements and annotated *food pictures*. These signals have been acquired with the *Zephyr BioHarness 3* wearable device in real-life conditions. The dataset is composed of two distinct subsets that have been acquired on different populations: one on healthy people and one on patients with diabetes. The full dataset is composed of roughly 1550 hours of both *ECG*, *Breathing* and *Accelerometers* signals. The data acquisition took place on everyday-life conditions with a chest belt, meaning the dataset has noisy parts as well as missing data. This makes the dataset ideal for the development and evaluation of algorithms for wearable healthcare systems.

The emergence of datasets of this kind is expected to help and facilitate the development of technologies and algorithms that are able to pre-process, filter, clean, analyze, and extract information, mine, and exploit raw data measurements resulting from continuous monitoring of health related information, typically acquired through wearable sensing devices. Although, as noted in Section 5, other alternative methods for glycemic monitoring are currently under development, the inclusion of multi-modal data for diabetes management is still necessary, considering the possibility of taking into account food intake or activity levels for prediction or personalized recommendation tasks.

Data acquired in real-life conditions as in this dataset, offers interesting opportunities in different directions. First, in the context of diabetes management with wearable devices: there is not any other publicly accessible dataset offering

460 *ECG* and *glucose* time series acquired both on wearable devices and in real-life conditions. Second, for the improvement of healthcare management algorithms through better processing of signals coming from wearable devices. Other potential use-cases of this dataset are also proposed in Section 5, even if reuse in other contexts is not excluded.

465 The dataset can be obtained through the Zenodo<sup>8</sup> platform. The platform provides a permanent URL and address for the dataset, as well as a DOI and versioning. The access to the data will be granted free of charge, subject to certain conditions specified by a data license distributed with the dataset. The main conditions refer to attribution and non-commercial use, as well as the  
470 provision of: a short description of the context in which the dataset will be used, and a signed form assessing that the dataset will only be used for scientific research and that no attempt will be made to identify patients. The license also includes other terms, such as the deletion of data upon owner request.

### Acknowledgment

475 This research project *DINAMO* and the data collection protocol has been approved by the "Commission cantonale d'éthique de la recherche sur l'être humain"<sup>9</sup> of canton de Vaud in Switzerland, under the number CER-VD 53/15.

The authors would like to thank:

- The Nano-Tera initiative<sup>10</sup> that financed the *DINAMO* research project  
480 in which this data acquisition took place,
- Alejandra Ruiz Torres, the dietitian who annotated the pictures,
- The Hôpital-Riviera Chablais, in particular the nurses, for the valuable collaboration,
- Sandrine Ding and Karl Aberer for their work on the *DINAMO* project.

### 485 References

- [1] J. Clawson, J. A. Pater, A. D. Miller, E. D. Mynatt, L. Mamykina, No longer wearing: investigating the abandonment of personal health-tracking technologies on craigslist, in: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM, 2015, pp. 647–658.  
490
- [2] D. Calvaresi, D. Cesarini, P. Sernani, M. Marinoni, A. F. Dragoni, A. Sturm, Exploring the ambient assisted living domain: a systematic review, *Journal of Ambient Intelligence and Humanized Computing* 8 (2) (2017) 239–257.

<sup>8</sup><https://zenodo.org/> Note to reviewers: Dataset will be released before publication

<sup>9</sup><http://www.cer-vd.ch/>

<sup>10</sup><http://www.nano-tera.ch/index.php>

- 495 [3] J. Connelly, A. Kirk, J. Masthoff, S. MacRury, The use of technology to promote physical activity in type 2 diabetes management: a systematic review, *Diabetic Medicine* 30 (12) (2013) 1420–1432.
- [4] J. Vietri, E. Witt, T. Meyer, Physical activity and health-related quality of life among adults with type 2 diabetes: Results from wearable fitness  
500 trackers, *Value in Health* 19 (7) (2016) A678.
- [5] B. Lopez, P. Herrero, C. Martin, Proceedings of the 1st ecai workshop on artificial intelligence for diabetes, in: 22nd European Conference on Artificial Intelligence (ECAI), 2016.  
URL <http://www.pepper.eu.com/ApplicationFiles/Download/325>
- 505 [6] N. Falcionelli, P. Sernani, A. Brugués, D. N. Mekuria, D. Calvaresi, M. Schumacher, A. F. Dragoni, S. Bromuri, Event calculus agent minds applied to diabetes monitoring, in: *International Conference on Autonomous Agents and Multiagent Systems*, Springer, 2017, pp. 258–274.
- [7] F. Dubosson, S. Bromuri, J.-E. Ranvier, M. Schumacher, D1namo, a personal health system for glycemic events detection, in: *Workshop Artificial Intelligence for Diabetes, 22nd European Conference on Artificial Intelligence (ECAI 2016)*, ECAI, 2016.  
510
- [8] J.-E. Ranvier, J.-P. Calbimonte, F. Dubosson, K. Aberer, Detection of hypoglycemic events through wearable sensors, in: *Proceedings of the 1st Workshop on Semantic Web Technologies for Mobile and Pervasive Environments co-located with the 13th Extended Semantic Web Conference (ESWC 2016)*, Heraklion, Greece, May 29, 2016., 2016, pp. 21–26.  
515 URL <http://ceur-ws.org/Vol-1588/paper2.pdf>
- [9] M. Saeed, M. Villarroel, A. T. Reisner, G. Clifford, L.-W. Lehman, G. Moody, T. Heldt, T. H. Kyaw, B. Moody, R. G. Mark, Multiparameter intelligent monitoring in intensive care ii: A public-access intensive care unit database\*, *Critical Care Medicine* 39 (5) (2011) 952960.  
520 doi:10.1097/ccm.0b013e31820a92c6.  
URL <http://dx.doi.org/10.1097/CCM.0b013e31820a92c6>
- 525 [10] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, H. E. Stanley, Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals, *Circulation* 101 (23) (2000) e215e220. doi:10.1161/01.cir.101.23.e215.  
530 URL <http://dx.doi.org/10.1161/01.CIR.101.23.e215>
- [11] G. Moody, R. Mark, The impact of the mit-bih arrhythmia database, *IEEE M EMB* 20 (3) (2001) 45–50. doi:10.1109/51.932724.  
URL <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=932724>

- 535 [12] M. Lichman, UCI machine learning repository (2013).  
URL <http://archive.ics.uci.edu/ml>
- [13] O. Banos, R. Garcia, J. A. Holgado-Terriza, M. Damas, H. Pomares, I. Rojas, A. Saez, C. Villalonga, mhealthdroid: a novel framework for agile development of mobile health applications, in: International Workshop on  
540 Ambient Assisted Living, Springer, 2014, pp. 91–98.
- [14] L. Quintero, Desarrollo de una base de datos para la evaluación de la neuropatía autonómica cardíaca de origen diabético, Ph.D. thesis, Tesis de Maestría, Universidad Simón Bolívar, Coord. de Ing. Electrónica (2007).
- [15] C. A. Ledezma, E. Severeyn, G. Perpignan, M. Altuve, S. Wong, A new  
545 on-line electrocardiographic records database and computer routines for data analysis, in: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2014, pp. 2738–2741.
- [16] H. Nyongesa, C. Alexakis, R. Saatchi, M. Rodrigues, N. Harris, S. Heller, C. Davies, C. Emery, R. Ireland, Classification of saecg by autoregressive modelling and neural networks, in: AFRICON, 2004. 7th AFRICON Conference in Africa, Vol. 2, IEEE, 2004, pp. 841–844.  
550
- [17] T. F. Christensen, L. Tarnow, J. Randløv, L. Kristensen, J. Struijk, E. Elstrup, O. K. Hejlesen, Qt interval prolongation during spontaneous episodes of hypoglycaemia in type 1 diabetes: the impact of heart rate correction,  
555 *Diabetologia* 53 (9) (2010) 2036–2041.
- [18] C. Alexakis, M. Rodrigues, R. Saatchi, H. Nyongesa, C. Davies, R. Ireland, N. Harris, S. Heller, A knowledge-based electrocardiogram-monitoring system for detection of the onset of nocturnal hypoglycaemia in type 1 diabetic patients, in: 2006 Computers in Cardiology, IEEE, 2006, pp. 5–8.
- [19] N. Nuryani, S. S. Ling, H. Nguyen, Electrocardiographic signals and swarm-based support vector machine for hypoglycemia detection, *Annals of biomedical engineering* 40 (4) (2012) 934–945.  
560
- [20] D. Anguita, A. Ghio, L. Oneto, X. Parra, J. L. Reyes-Ortiz, A public domain dataset for human activity recognition using smartphones., in: ESANN, 2013.  
565
- [21] L. B. Sherar, P. Griew, D. W. Esliger, A. R. Cooper, U. Ekelund, K. Judge, C. Riddoch, International children’s accelerometry database (icad): design and methods, *BMC Public Health* 11 (1) (2011) 485.
- [22] Y. Matsuda, H. Hoashi, K. Yanai, Recognition of multiple-food images by detecting candidate regions, in: Proc. of IEEE International Conference on  
570 Multimedia and Expo (ICME), 2012.



- [23] Y. Kawano, K. Yanai, Automatic expansion of a food image dataset leveraging existing categories with domain adaptation, in: Proc. of ECCV Workshop on Transferring and Adapting Source Knowledge in Computer Vision (TASK-CV), 2014.
- 575 [24] L. Bossard, M. Guillaumin, L. Van Gool, Food-101 – mining discriminative components with random forests, in: European Conference on Computer Vision, 2014.
- [25] J. Blechert, A. Meule, N. A. Busch, K. Ohla, Food-pics: an image database for experimental research on eating and appetite, *Food cravings* 26 (2015) 40.
- 580 [26] K. R. Pitzer, S. Desai, T. Dunn, S. Edelman, Y. Jayalakshmi, J. Kennedy, J. A. Tamada, R. O. Potts, Detection of hypoglycemia with the glucowatch biographer, *Diabetes Care* 24 (5) (2001) 881–885. arXiv:<http://care.diabetesjournals.org/content/24/5/881.full.pdf>, doi:10.2337/diacare.24.5.881.  
URL <http://care.diabetesjournals.org/content/24/5/881>
- [27] G. Reach, Which threshold to detect hypoglycemia?, *Diabetes Care* 24 (5) (2001) 803–804. arXiv:<http://care.diabetesjournals.org/content/24/5/803.full.pdf>, doi:10.2337/diacare.24.5.803.  
590 URL <http://care.diabetesjournals.org/content/24/5/803>
- [28] A. D. Association, Diagnosis and classification of diabetes mellitus, *Diabetes Care* 29 (suppl 1) (2006) s43–s48. arXiv:[http://care.diabetesjournals.org/content/29/suppl\\_1/s43.full.pdf](http://care.diabetesjournals.org/content/29/suppl_1/s43.full.pdf).  
595 URL [http://care.diabetesjournals.org/content/29/suppl\\_1/s43](http://care.diabetesjournals.org/content/29/suppl_1/s43)
- [29] T. Laitinen, T. Lyyra-Laitinen, H. Huopio, I. Vauhkonen, T. Halonen, J. Hartikainen, L. Niskanen, M. Laakso, Electrocardiographic alterations during hyperinsulinemic hypoglycemia in healthy subjects, *Annals of Non-invasive Electrocardiology* 13 (2) (2008) 97–105.
- 600 [30] S. P. Lee, L. Yeoh, N. D. Harris, C. M. Davies, R. T. Robinson, A. Leathard, C. Newman, I. A. Macdonald, S. R. Heller, Influence of autonomic neuropathy on qtc interval lengthening during hypoglycemia in type 1 diabetes, *Diabetes* 53 (6) (2004) 1535–1542.
- [31] R. Tanenberg, C. Newton, I. Almond Drake, Confirmation of hypoglycemia in the “dead-in-bed” syndrome, as captured by a retrospective continuous glucose monitoring system, *Endocrine practice*.
- 605 [32] B. Eckert, C.-D. Agardh, Hypoglycaemia leads to an increased qt interval in normal men, *Clinical Physiology* 18 (6) (1998) 570–575.
- [33] J. Marques, E. George, S. Peacey, N. Harris, I. Macdonald, T. Cochrane, S. Heller, Altered ventricular repolarization during hypoglycaemia in patients with diabetes, *Diabetic Medicine* 14 (8) (1997) 648–654.
- 610

- [34] J. Meinhold, T. Heise, K. Rave, L. Heinemann, Electrocardiographic changes during insulin-induced hypoglycemia in healthy subjects., *Hormone and metabolic research= Hormon-und Stoffwechselforschung= Hormones et metabolisme* 30 (11) (1998) 694–697.
- 615 [35] B. Cvetkovic, U. Pangerc, A. Gradišek, M. Luštrek, Monitoring patients with diabetes using wearable sensors: Predicting glycaemias using ecg and respiration rate, 1st ECAI Workshop on Artificial intelligence for Diabetes at the 22nd European Conference on Artificial Intelligence (ECAI).
- 620 [36] A. Malkoc, D. Probst, C. Lin, M. Khanwalker, C. Beck, C. B. Cook, J. T. La Belle, Enhancing glycemic control via detection of insulin using electrochemical impedance spectroscopy, *Journal of diabetes science and technology* 11 (5) (2017) 930–935.
- 625 [37] S. M. Lundsgaard-Nielsen, A. Pors, S. O. Banke, J. E. Henriksen, D. K. Hepp, A. Weber, Critical-depth raman spectroscopy enables home-use non-invasive glucose monitoring, *PloS one* 13 (5) (2018) e0197134.
- 630 [38] L. Lipani, B. G. Dupont, F. Doungmene, F. Marken, R. M. Tyrrell, R. H. Guy, A. Ilie, Non-invasive, transdermal, path-selective and specific glucose monitoring via a graphene-based platform, *Nature nanotechnology* (2018) 1.
- [39] C. Chen, X.-L. Zhao, Z.-H. Li, Z.-G. Zhu, S.-H. Qian, A. J. Flewitt, Current and emerging technology for continuous glucose monitoring, *Sensors* 17 (1) (2017) 182.

# Conflict of Interest

The following people have been related to the D1NAMO project and thus can not be called as reviewers:

- Karl Aberer
- Sandring Ding
- Matteo Vasirani

# Ethical Statement

This research project D1NAMO and the data collection protocol has been approved by the "Commission cantonale d'éthique de la recherche sur l'être humain" [1] of canton de Vaud in Switzerland, under the number CER-VD 53/15.

[1] <http://www.cer-vd.ch/>