Prediction of hyperglycaemia and hypoglycaemia events using longitudinal data
Natalia Mordvanyuk, Ferran Torrent-Fontbona, Beatriz López
University of Girona, eXiT research group
natalia.mordvanyuk@udg.edu, ferran.torrent@udg.edu, beatriz.lopez@udg.edu

Abstract
In type 1 diabetes management, mobile health applications are becoming a cornerstone to empower people to self-manage their disease. There are many applications addressed to calculate insulin doses based on the current information (e.g., carbohydrates intake) and a few of them are accompanied by modules able to supervise postprandial conditions and recommend corrective actions if the user falls in an abnormal state (i.e., hyperglycaemia or hypoglycaemia). On the other hand, mobile apps favour the gathering of historical data from which machine learning techniques can be used to predict if user conditions will worsen.

This work presents the application of k-nearest neighbour on the historical data gathered on patients, so that given the information related to a sequence of meals, the model is able to predict if the patient will fall in an abnormal condition. The experimentation has been carried out with the UVA-Paola type I diabetes simulator over eleven adult profiles. Results corroborate that the use of sequential data improves significantly the prediction outcome when forecasts distinguish the type of meal (breakfast, lunch and dinner).

Introduction
Type 1 Diabetes Mellitus (T1DM) is a chronic disease that demands a strict control of the Blood Glucose (BG) level of the patient. This BG control is required to avoid hypoglycaemia or hyperglycaemia events, which are associated to serious short-term and long-term complications, e.g. coma, blindness, severe kidney failure or even death. This paper studies how sequences of data (recorded by T1DM people) can be used to predict hypoglycaemia and hyperglycaemia events using the k-nearest neighbours (k-NN) method.

Methodology
In the problem faced in this paper, the following information is available: Time (T) in minutes; Carbohydrate (CH) intake (mg); Bolus insulin dose (B); CGM readings (mg/dL). This information is used to create an event for each meal.

Day 1 07:16 3866.7 3.36 139.52 Hypo

Table 1: Structure of a meal event

Given the set of meal events, these are sorted according to the time attribute and processed to create sequences of meals where each sequence contains all the ordered meals of a time window of x hours.

The class of each sequence of meals is labelled according to the postprandial status (PS) of the last meal of the sequence, following two procedures:

- PS is labelled as hypoglycaemia if a CGM reading between 2 and 6 hours after the bolus administration is below 70 mg/dL, otherwise it is labelled as non-hypoglycaemia.
- PS is labelled as hyperglycaemia if CGM readings are above 180 mg/dL during at least 60 minutes between 2 and 6 hours after the bolus administration, otherwise it is labelled as non-hyperglycaemia.

These sequences are then used to predict the class of a given sequence of meals using KNN, where $k < \sqrt{n}$ (n is the number of examples).

Results

Conclusions
1. Recommendations based on meals increases the overall performance (TPRs up to 0.88 and accuracies up to 83% have been achieved, while FNR and FPR are significantly lower, from 0.25 to 0.11). Furthermore, the use of sequential data clearly outperforms the one-shot data.
2. Recommendations should be based on type of meal (do not merge all the meals together).
3. Personalized databases with less amount of registers, does not favour nor the use of sequences, neither the distinction of different types of meals.

Forthcoming Research
- Eager mechanisms or its hybridization for sequence learning
- Fuzzy approach in the labelling

Acknowledgements
This work has received funding from the EU Horizon 2020 research and innovation programme under grant agreement No 689810 (PEPPER), and from the University of Girona under the grant MFCug2016-1 (Ajust per a la millora de la productivitat científica dels grups de recerca), and the Spanish MINECO under the grant number DIP2013-47458-C21-R. This work has been developed with the support of the research group SITES awarded with distinction by the Generalitat de Catalunya (SGR 2014-2016).