

Automatic classification of glycaemia measurements to enhance data interpretation in an expert system for gestational diabetes

Estefanía Caballero-Ruiz , Gema García-Sáez , Mercedes Rigla , María Villaplana , Belén Pons , M. Elena Hernando

A B S T R A C T

Expert systems for diabetes care need to automatically evaluate glycaemia measurements in relationship to meals to correctly determine patients' metabolic condition and generate recommendations about therapy adjustments. Most glucose meters allow patients to manually label each measurement with a meal tag, but as this utility is not always used, a completion procedure is needed. Classification methods are usually based on predefined mealtimes and present insufficient accuracy that might affect the automatic data analysis. Expert systems in diabetes require a reliable method to manage incomplete glycaemia data so that they can determine if patients' metabolic condition is altered due to a specific meal or due to an extended fasting period.

This paper presents the design and application of a classification module to automatically assign the appropriate meal and 'moment of measurement' to incomplete glycaemia data. Different machine learning techniques were studied in order to design the best classification algorithm in terms of accuracy. The selected classifier was implemented with a C4.5 decision tree with 7 input features selected with a wrapper evaluator and the genetic search algorithm, which achieved 95.45% of accuracy with the training set on cross-validation. The classification module was integrated in the *Sinedie* expert system for gestational diabetes care and was evaluated in a clinical environment for 8 months with 42 patients. A total of 7,113 glycaemia measurements were uploaded by patients into the *Sinedie* system and were completed by the "classification module". The 98.79% of the measurements were correctly classified, while patients modified the automatic classification of 1.21% of them. Classification results were improved by 21.04% compared to a classification based on predefined mealtimes. The automatic classification of glycaemia measurements minimizes the patient's intervention, allows structuring measurements in relationship to meals and makes automatic data interpretation by expert systems more reliable.

1. Introduction

As in other types of diabetes, the prevalence of Gestational diabetes mellitus (GDM) is increasing throughout the world (IDF, 2015). If the new International Association of Diabetes Study Group diagnosis criteria (IADPSG, 2010) –recently proven to be cost effective (Duran, Saenz, & Torrejon, 2014) – are adopted, the prevalence could be doubled. Several adverse outcomes are associated with hyperglycaemia in pregnancy, as foetal macrosomia,

shoulder dystocia or caesarean section (Metzger, Lynn, & Lowe, 2008). Although most cases resolve with delivery, both mother and foetus are at a higher risk of developing type 2 diabetes in the future (Boney, 2005; Franks, Looker, & Kobes, 2006).

Maternal glycemic control reduces adverse GDM outcomes (Hartling, Dryden, & Guthrie, 2013) so patients are prescribed to self-monitor their blood glucose (BG) levels with a glucose meter around main meals. Although measurements are stored in the glucose meter memory file, patients usually note down their results in a paper logbook, structuring measurements in relationship to meals. They indicate the specific meal which the measurement is related to (breakfast, lunch or dinner) and whether it was made before (preprandial) or after the meal (postprandial). Clinicians evaluate the patients' measurements each week or every other

week to determine the appropriate treatment, which consists of nutritional prescription, physical activity and, if necessary, insulin therapy. It has been observed that patients commit errors when manually reporting their BG levels, being the mean values significantly higher than the logbooks' ones (Given, O' Kane, Bunting, & Coates, 2013). Although this could mask a bad glycemic control and make clinicians establish a wrong therapy, they still prefer to examine logbooks instead of meter memory files (Polonsky, Jelsovsky, & Panzera, 2009). The reason might be that logbooks are easier to be reviewed, as they provide structured information that glucose meter memory lacks, like associations of measurements to meals, which are essential to make therapy adjustments. Logbooks also provide additional information such as food intakes, insulin doses or exercise.

Telemedicine allows patients to send their BG data to the system to be remotely evaluated, which avoids unnecessary displacements (Carral, Ayala, del, & Fernández, 2015; Pérez-Ferre, Galindo, & Fernández, 2010) and improves access to specialized care in rural areas (Mohan & Pradeepa, 2014). Furthermore, by a more exhaustive and frequent evaluation of accurate data, telemedicine is capable of improving glycemic control (Wojcicki, Ladyzynski, & Krzymien, 2001) and reducing GDM adverse outcomes (Dalfra, Nicolucci, & Lapolla, 2009; Ferrara, Hedderson, & Ching, 2012). Monitoring data in telemedicine systems should be presented to clinicians organized as they appear in paper logbooks to facilitate their interpretation.

The use of telemedicine could increase clinician's workload as it favors the generation of a greater amount of data to be evaluated by clinicians. Expert systems can solve the potential increment of clinicians' workload (Klonoff & True, 2009) by automatically analysing patients' monitoring data according to expert specifications (Hernando, Gómez, Corcoy, & del Pozo, 2000). The automatic analysis of monitoring data could optimize clinician's time by notifying which patients are evolving satisfactorily and which ones need a deeper examination. Expert systems, like clinicians, need to analyze glycaemia data in relation to meals to be able to determine patients' condition and to generate specific recommendations about therapy adjustments.

Glycaemia data entry in expert systems can be performed by patients either manually or by uploading the data stored in their glucose meter (El-Gayar, Timsina, Nawar, & Eid, 2013a). The automation of data entry is preferred as it minimizes transcription errors (Given et al., 2013), results in more data captured, simplifies the data entry process and increases patients' satisfaction (El-Gayar, Timsina, Nawar, & Eid, 2013b). One of the problems that expert systems in diabetes have to face is the management of incomplete glycaemia measurements. A measurement is considered incomplete if it lacks its association with a meal or with a moment of measurement. Newer glucose meters can include the functionality to allow registering these data manually, but even if they do, it is a time-consuming task and patients sometimes forget to introduce it. Without an accurate method to manage incomplete glycaemia data, expert systems cannot determine if patients' metabolic condition is altered due to a specific meal that should be adjusted or due to an extended fasting period.

The majority of studies available in literature about expert systems do not explicitly describe the method used to retrieve the associated meal and moment of measurement of glycaemia data or how they manage the lack of this information. Some expert systems allow patients to add a meal tag to glycaemia measurements after uploading data from the glucose meter (Cafazzo, Casselman, & Hamming, 2012; Lim, Kang, & Shin, 2011; Quinn, Clough, & Minor, 2008), but it has been observed that sometimes patients forget to label some of the measurements (Mackillop et al., 2014) so they cannot be automatically analyzed. To solve this problem, the expert system can preselect a meal tag for each measurement

downloaded and allow patients to modify it. Bromuri et al, preselect glycaemia meal tags based on previous measurements, so, if the last measurement was taken before dinner, an after dinner period is preselected (Bromuri, Puricel, & Schumann, 2016). However, this method might present problems when dealing with repeated or missing measurements, for example if the patient forgets to measure her glycaemia after dinner, she measures it the following day before breakfast and she, by mistake, accepts the preselected meal tag. Some commercial applications automatically classify the glycaemia measurements downloaded by patients according to patient's predefined mealtimes (Sanofi Diabetes, 2015), but this method might present an elevated rate of errors as we will see in the following sections. We propose an innovative method for glycaemia meal tag preselection using machine learning techniques.

This paper presents the methodology to design an automatic classifier to associate the appropriate meal and moment of measurement to each glycaemia data downloaded from a glucose meter, its integration within the *Sinedie* expert system for GDM and the classification results obtained in a pilot study at Hospital de Sabadell with 47 patients for 8 months.

2. Material and methods

This section describes the *Sinedie* expert system and how the automatic classifier is integrated with the BG levels uploading procedure. We explain the two different classification strategies studied to design the classifier: a simple algorithm based on the patient's mealtime schedules, measurements' time and BG level; and a more complex algorithm based on machine learning techniques. Finally, we describe the design of the clinical evaluation experiment.

2.1. *Sinedie* expert system for GDM

Sinedie is a telemedicine platform enhanced by an expert system to manage the treatment of GDM patients. It aims to improve health care processes by reducing the evaluation time per patient, avoiding unnecessary displacements and improving the access to specialized healthcare. The expert system available in *Sinedie* computes the patients' metabolic condition and generates advice, to both patients and physicians about treatment changes, including the need to start an insulin therapy. The BG classifier presented in this paper was integrated in the *Sinedie* system as a classification module, whose functionality is to assign an appropriate mealtime and a moment of measurement to each incomplete measurement uploaded to the system with the glucose meter. The glucose meter memory file provides information about date and time of each measurement, as well as its value (see Fig. 1). Additionally, patients can enter the corresponding moment of measurement (preprandial or postprandial) or the associated mealtime if they have a glucose meter that allows registering such information. The classifier allows structuring the BG levels obtained from the glucose meter file to be visualized in an e-logbook in *Sinedie* (Fig. 1). This completion procedure is executed in a preprocessing step prior to the automatic data analysis performed by the expert system and it is essential to detect anomalous conditions in patient's health. After each data download, patients verify if the automatic classification is accurate and otherwise correct it.

2.2. Classification problem analysis

A preliminary study (study 1) was carried out to determine which would be the optimal classifier to be implemented in *Sinedie*. As part of this study, we examined measurements' distribution along the day according to time and BG level to analyze the

GLUCOSE METER MEMORY FILE

```
<IMPORT>
<ACSPIX Type="2106" SN="UI00346822" Ver="3.00.02"/>
<DEVICE Name="Aviva" SN="36815015" Dt="2015-02-06"
Tm="10:49" BGUnit="mg/dL"/>
<RECENTREC Dt="2014-11-24" Tm="17:07"/>
<BGDATA>
<BG Val="118" Dt="2014-11-24" Tm="17:07" D="1"/>
<BG Val="126" Dt="2014-11-24" Tm="11:09" D="1"/>
<BG Val="99" Dt="2014-11-24" Tm="09:12" D="1"/>
<BG Val="116" Dt="2014-11-23" Tm="23:58" D="1"/>
<BG Val="135" Dt="2014-11-23" Tm="21:25" D="1"/>
<BG Val="132" Dt="2014-11-23" Tm="17:12" D="1"/>
<BG Val="133" Dt="2014-11-23" Tm="11:29" D="1"/>
<BG Val="94" Dt="2014-11-23" Tm="09:47" D="1"/>
<BG Val="116" Dt="2014-11-22" Tm="23:31" D="1"/>
<BG Val="121" Dt="2014-11-22" Tm="17:25" D="1"/>
<BG Val="144" Dt="2014-11-22" Tm="11:13" D="1"/>
<BG Val="144" Dt="2014-11-22" Tm="11:11" D="1"/>
<BG Val="83" Dt="2014-11-22" Tm="09:15" D="1"/>
```

Classification



E - LOGBOOK

Date	Prep. breakf.	Post. breakf.	Prep. lunch	Post. lunch	Prep. dinner	Post. dinner	Others
24/11/14	99	126		118			3
23/11/14	94	133		132	135	116	5
22/11/14	83			121		116	5
21/11/14	98	114		147		145	5
20/11/14	91	132		124		130	4
19/11/14	93	117		154		168	4
18/11/14	90	117		122		129	4
17/11/14	91	133		125		134	4
16/11/14	97	142		129		113	4
15/11/14	94	133		115		149	4
14/11/14	105	147		118		146	4
13/11/14	100	140		117		132	4
12/11/14	95					130	5

Fig. 1. Built patient e-logbook from the glucose meter memory file in the Sinedie system. The Sinedie e-logbook shows the 6 main intervals around main meals, when patients usually perform a BG measurement. Most of the patients also measure their glycaemia in other intervals, or perform several measurements in the same one. The column ("Others") shows the total number of measurements performed by the patient on a specific day, and shows all measurements' details when the cell is clicked.

feasibility of using a simple algorithm based on patient's mealtime schedules, measurements' time and BG level.

2.2.1. Glycaemia data description

The study 1 dataset (DS_G) includes retrospective data from 25 GDM patients from the Hospital de Sabadell in Barcelona, Spain. Fifteen of the patients were treated only with diet and ten of them also with insulin therapy. Patients with insulin therapy administered either fast insulin associated to one or more meal intakes, and/or slow insulin at night. Patients used the Accu-check Aviva or the Accu-check Nano glucose meter (Roche Diagnostics, 2013a, b) over an average period of 60.9 ± 33.5 days, and were requested to perform at least 4 measurements a day: before breakfast (breakfast preprandial) and after each main meal (breakfast postprandial, lunch postprandial and dinner postprandial). Patients also filled in a paper logbook with the four BG measurements requested. The DS_G dataset includes a total of 6025 BG measurements from patients' glucose meters (241.0 ± 134.8 per patient) and it is formed by two features obtained directly from the glucose meter memory file:

- "time": Time of the day when the measurement was taken in absolute minutes.
- "bg": Glycaemia value in mg/dl.

BG measurements were labeled with the meal information registered manually in the patients' paper logbooks. The concordance detected between meter files and logbooks was 93.50%, considering concordance as the measurements correctly registered in the logbook out of those registered at all (Given et al., 2013). The "underreported" measurements, the ones appearing in the memory file but not in the logbook, were 5.05% of the stored measurements and were labeled by the authors including the following tags: lunch preprandial, dinner preprandial, repeated, morning, afternoon and night. We will refer to these measurements in the rest of the document as "additional measurements". BG values differed between the meter file and the logbook in 1.78% of the measurements registered; in that case the meter file value was maintained. We also found that 1.98% of the logbook measurements registered did not appear in the meter file.

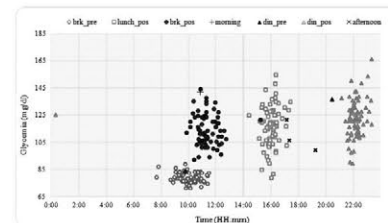


Fig. 2. Glycaemia measurements distribution of patient P5 along the day.

2.2.2. Measurements scatter plots

Twenty five scatter plots, one for each patient, were created in order to visually analyze the distribution of glycaemia measurements along the day. Each plot was generated according to measurement time and BG value. All scatter plots showed that measurements were clearly distributed in three groups corresponding to the three main meals: breakfast, lunch and dinner. It was also observed an overlap in breakfast preprandial and postprandial measurements time and, in some cases, BG values as well. An example of these assessments can be observed in Fig. 2, which shows the measurements' distribution of one patient treated only with diet.

Fig. 2 shows the time overlap in P5 breakfast measurements (white and black circles) from 9:43 to 11:20. It also presents four additional measurements: one at 10:52 in the morning (+ symbol) that falls in the breakfast interval and three in the afternoon (x symbols), two of them coincide with the postprandial lunch interval and the last one is a preprandial dinner (black triangle symbol). These additional measurements were observed in 18 out of the 25 patients, ranging from 1 to 134 measurements per patient. At the left side of the chart we can see an example of a midnight dinner which should be grouped with the rest of the meals at dinnertime instead of with the breakfast ones. It would be logical to think that the postprandial BG levels would be higher than the postprandial ones, due to carbohydrates intake as we observe in P5 values. But, as shown in Section 2.2.3, we found that this is not always true.

2.2.3. Statistical analysis

A further statistical analysis performed with the PSPP tool (PSPP, 2013) confirmed our visual appreciations. Four intervals

Table 1

Measurement time variability for each patient in the four mealtime intervals defined, distinguishing between weekdays and weekends (_w).

Patient	Bpre		Bpos		Lpos		Dpos	
	IQR	IQR _w	IQR	IQR _w	IQR	IQR _w	IQR	IQR _w
P1	277	143	94	49	46	45	83	53
P2	39	49	42	34	30	69	33	44
P3	7	144	32	85	13	39	32	41
P4	37	37	33	34	33	10	68	107
P5	60	60	60	62	32	91	27	41
P6	59	57	46	48	49	114	43	96
P7	188	84	135	49	65	73	62	114
P8	29	50	74	49	46	55	48	45
P9	78	67	87	97	43	26	20	43
P10	67	57	65	48	59	106	67	69
P11	86	92	88	87	79	62	47	37
P12	42	58	35	43	59	116	23	68
P13	19	29	19	23	39	47	26	33
P14	33	38	45	83	34	98	35	114
P15	80	80	76	110	49	66	66	46
P16	104	92	96	104	116	49	64	47
P17	85	61	92	82	40	48	30	42
P18	33	81	31	84	11	34	13	62
P19	19	74	27	63	27	13	44	47
P20	43	28	47	77	64	31	57	69
P21	38	75	62	45	47	40	37	60
P22	24	67	38	64	59	66	62	73
P23	60	34	54	47	48	58	70	39
P24	81	73	72	63	47	42	53	49
P25	63	68	60	73	43	63	66	36
Mean	66 ± 57	68 ± 29	60 ± 28	64 ± 23	47 ± 21	58 ± 29	47 ± 19	59 ± 25

Interquartile range expressed in minutes.

were established: preprandial breakfast (Bpre), postprandial breakfast (Bpos), postprandial lunch (Lpos) and postprandial dinner (Dpos) in order to study the variability in measurement time in each one of them. For each interval, we calculated the measurement mean time and interquartile range (IQR) for each patient, globally and differentiating between working days and weekends. We used the IQR instead of the standard deviation because it better shows the variability in the patients' usual measurement times, not considering the extreme values. Midnight measurements, from 0:00 to 3:20, were modified by adding 1440 min (24 h*60 min) to the original value, in order to elude cyclical distances issues. Table 1 shows the results of the statistical study about measurement time variability of each patient over the four meal intervals defined.

The observed average IQR exceeds 45 min in all the intervals. Focusing our attention on breakfast's intervals, the average IQR in both preprandial and postprandial subintervals is of 66 and 68 min respectively. This elevated IQR, over 60 min, was detected in 13 out of the 25 patients in "Bpos" interval and 12 out of 25 in "Bpre" interval. Furthermore, we detected a delay in the mean measurement time at weekends of 47 min at breakfast, 30 min at lunch and 16 min at dinner. These behavioral changes at weekends could be a criterion to take into account in our classification algorithm. However, in 3 patients no delay was observed.

In order to study the glycemic preprandial and postprandial ranges at breakfast time, we calculated the minimum and maximum BG level of the two breakfast intervals establishing for each patient its glycemic range: (Bpre_{gmax} - Bpre_{gmin}) and (Bpos_{gmax} - Bpos_{gmin}). We calculated the percentage of preprandial measurements whose BG value fell within the postprandial glycemic range and vice versa. The percentage of coincidence found for insulin treated patients (P3, P4, P6, P7, P8, P11, P16, P17, P18 and P19) was of 47.92% (656 measurements out of 1367) and of 41.71% (599 measurements out of 1436) for non-insulin treated patients, reaching more than 50% in 3 patients using insulin and in 5 non-

insulin treated patients. Fig. 3 shows the breakfast preprandial and postprandial measurements' boxplot where preprandial and postprandial BG levels overlap can be appreciated in the patients' breakfast measurements.

The variability observed in the statistical analysis indicates that a classification algorithm based only on the time of measurement according to patients' mealtime schedules could present problems when distinguishing between preprandial and postprandial measurements in the same mealtime interval, as well as additional measurements. In GDM this is especially problematic at breakfast interval, when it is necessary to discriminate between fasting and postprandial measurements.

Neither is enough to add the BG level information to the algorithm, since we observed over a 40% of coincidence in glycaemia ranges at breakfast time regardless insulin administration, which makes it really complex to discern whether the measurement is preprandial or postprandial. For these reasons, we consider that a machine learning algorithm is needed to perform a rigorous and more accurate classification of measurements to correctly complete the uploaded measurements with the associated meal and moment of measurement.

2.3. Machine learning for classification

We evaluated different machine learning techniques applied to automatic classification. In order to evaluate whether the consideration of more parameters could improve classification accuracy, we obtained a larger set of features (DS18) from the DSG. We tried 6 different feature selection (FS) methods over the DS18 to select the most relevant features and to remove the redundant ones. Three datasets were obtained using FS (DSCFS, DSWB and DSWG) in addition to the initial DSG, the DS18, and a feature subset defined by an expert (DSE). We evaluated two learning algorithms in terms of accuracy, the

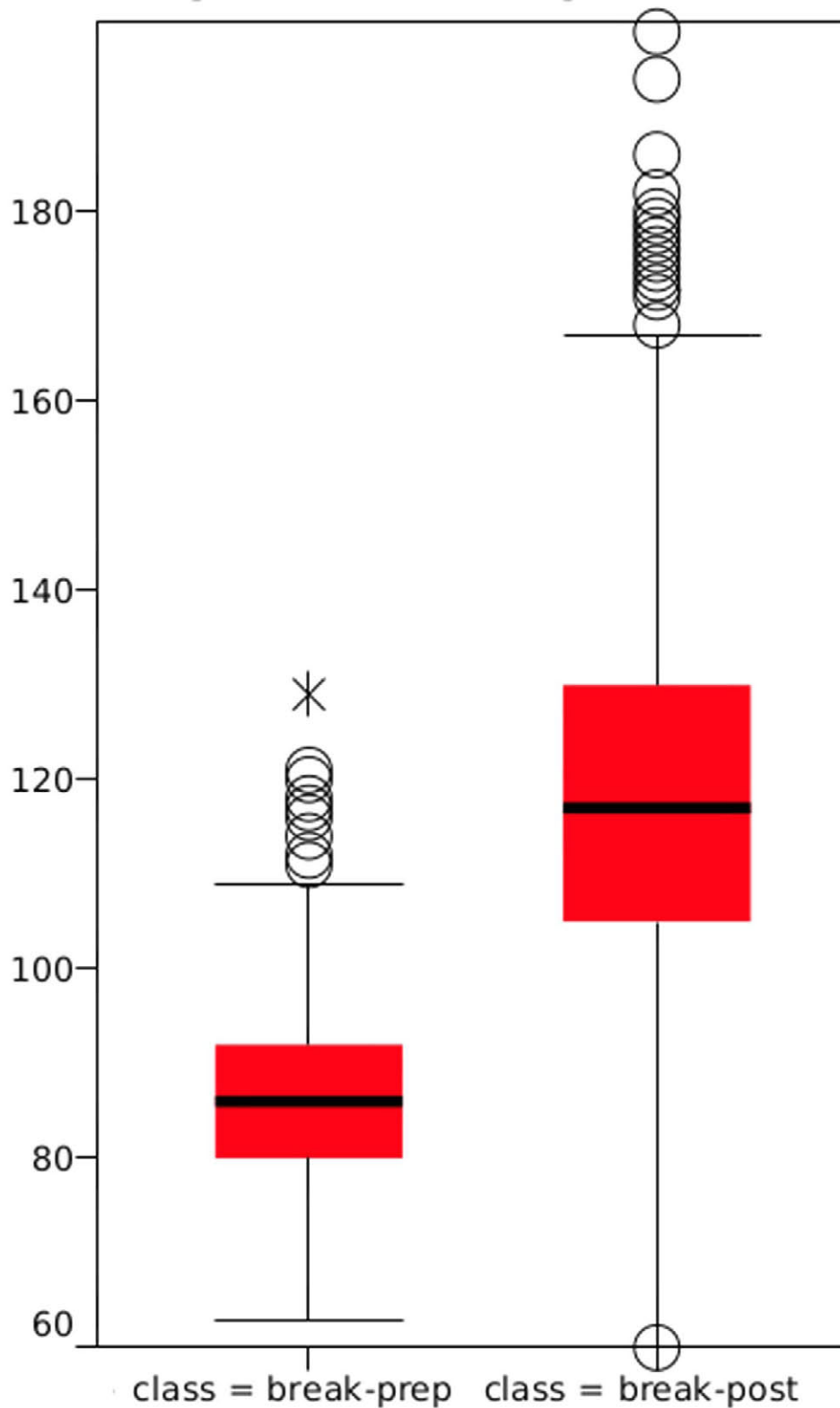


Fig. 3. Boxplot of breakfast preprandial and postprandial measurements of all patients. Blood glucose level in mg/dl.

C4.5 decision tree (Quinlan & J. Ross, 1993) and the Multilayer perceptron neural network (Haykin, 1994) with the 6 different subsets.

We used the Weka data mining tool (Witten, I. H. & Frank, E, 2005) due to its Java library that provides all the algorithms

implementation we needed. It was also chosen for its ease of integration in the *Sinedie* system, as it is also developed in Java. Weka default parameters were maintained in all the algorithms tested and a 10-fold cross-validation was applied for FS and learning algorithms evaluation.

2.3.1. Search of input and output features

We obtained a total of 18 input features from the DS_C and from the patients' clinical history. In a previous study performed for patients with type 1 diabetes (García-Sáez, Alonso, & Molero, 2009), an expert specified that for measurement's meal assignment, besides time and BG level, it was relevant to know when the previous measurement was taken and whether the patient was under insulin therapy or not. Following these recommendations, we obtained the DS_E subset with 'time', 'bg' and two additional features:

- "insulin": Boolean that indicates whether the patient has insulin therapy or not.
- "iprev": Time difference in minutes from previous measurement of the same day if any.

As the expert considered relevant the insulin treatment and time interval between measurements, we added 6 more features related to those parameters:

- "instype": Indicates the type of insulin administered and the moment of administration if any.
- "fib": Fast insulin dose at breakfast time.
- "fil": Fast insulin dose at lunchtime.
- "fid": Fast insulin dose at dinnertime.
- "sin": Slow insulin dose at night time.
- "ipos": Time difference in minutes until posterior BG measurement if any.

The consideration of time intervals between measurements led to the idea of evaluating groups of three measurements instead of evaluating isolated ones. So, two more features were calculated:

- "bgprev": Previous BG value if any.
- "bgpost": Posterior BG value if any.

To study if weekends, holidays or even the seasons of the year affect the patterns of the patients' measurements, and to identify which rules could be established related to this, we calculated the following 5 features from the date of measurement:

- "work": Boolean value that indicates if it is a working day or not.
- "dow": Day of the week.
- "doy": Day of the year.
- "day": Day of the month.
- "month": Month of the year.

As we verified in patients' scatter plots, measurements could be grouped in three clusters related to the three main meals, so that a last feature denominated "intake" was obtained. This feature indicates which cluster the measurement belongs to, and can take three different values: Breakfast, lunch or dinner. The feature is calculated clustering each patient's measurements according to its time with the Expectation Maximization algorithm (EM) (Dempster, Laird, & Rubin, 1977), specifying the number of clusters. Measurements after midnight were corrected, as explained in Section 2.2.3, to be grouped in the dinner cluster instead of the breakfast one. We also performed the clustering using the K-means (MacQueen, 1967) and X-Means (Pelleg & Moore, 2000) algorithms, but we obtained worse results than with the EM algorithm.

Regarding the classifier output features, ten classes were defined: "breakfast-prep", "breakfast-post", "lunch-prep", "lunch-post", "dinner-prep", "dinner-post", corresponding to preprandial and postprandial main meals measurements, and "morning", "afternoon", "night" and "repeated" corresponding to the additional measurements that some patients make.

2.3.2. Feature selection

In order to discover the optimal feature subset that achieves the best accuracy with the minimum complexity we tested 6

different FS methods by the combination of 2 evaluators and 2 searching algorithms.

There are 2 main approaches in FS evaluators, wrappers and filters, so we tested one evaluator of each approach. The wrapper evaluates feature sets by using a learning scheme (Kohavi, 1996) and it is computationally intensive. Among the filter models we chose the Correlation based feature selection (CFS), that evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them (Hall, 1999).

As not all searching algorithms perform equally before different datasets, we tested two algorithms belonging to two of the three main categories available: "Sequential search" and "Random search" (Liu & Yu, 2005). We excluded the "Complete search" category as it might be computationally infeasible when working with a large number of features. From the first category we tested the Best First algorithm (Witten & Frank, 2005), which searches the space of features subsets by greedy hill climbing augmented with a backtracking facility. It is simple and fast. We tried the three searching strategies: forward, backward and bidirectional. The other algorithm tested was the Genetic search (Goldberg, D. E., 1989), which belongs to the second category that can avoid local minimal problems. These 6 FS methods were applied to the DS_{18} obtaining 6 features subsets.

To determine which features should form the resulting subsets from FS, we initially included the features selected the greatest number of times. Then, we followed a sequential forward strategy adding the rest of the features one by one, selecting the most popular each time and stopping when accuracy decreased.

2.3.3. Learning algorithms

The algorithms evaluated to build the classifier are the C4.5 decision tree and a neural network with the multilayer perceptron (MLP) architecture. Both algorithms have been widely and successfully used in diabetes clinical domain with multiple purposes like retinal disease diagnosis (Bourouis, Feham, Hossain, & Zhang, 2014), impaired glucose metabolism or type 2 diabetes mellitus identification (Hische, Luis-Dominguez, & Pfeiffer, 2010; Möhlig, Flöter, & Spranger, 2006; Shankaracharya, Mallick, & Shukla, 2012; Upadhyaya, Farahmand, & Baker-Demaray, 2013), or to identify significant factors influencing diabetes control (Huang, McCullagh, Black, & Harper, 2007). However, they have not been used for BG level classification in mealtime intervals. The C4.5 algorithm implementation used is the *J48* and for the MLP the *Multilayer Perceptron*, both defined in Weka.

2.4. Clinical evaluation

The classifier was integrated in the *Sinedie* system as part of a classification module, which is also responsible for acquiring the input features required by the classifier. The classifier implementation is based on the results obtained in study 1. We selected the classification algorithm and the set of input features that achieved the highest accuracy in the second part of the study (2.3. *Machine learning for classification*). The implemented classifier was trained with the same data used in study 1.

The clinical evaluation was performed at a Spanish hospital, the "Hospital de Sabadell" for 8 months. During this period, 42 patients with GDM have uploaded a total of 7113 measurements to the *Sinedie* system, which have been automatically classified by the classification module.

In order to evaluate the classification results that would have been obtained using an algorithm based in predefined mealtimes we implemented a second classifier based on usual Spanish mealtimes. At the end of the clinical study, we categorized the measurements that patients had uploaded into the system with

Table 2

Features selected by each method in 10-fold cross validation. In bold and underlined the features selected by us to be included in the features subsets.

Feature	Wrapper				CFS
	Best First		Genetic		
	J48	MLP	J48	MLP	
day	1	0	2	0	0
month	1	1	1	2	0
doy	1	0	0	1	0
dow	2	2	0	0	0
work	1	0	5	4	0
time	9	10	10	10	10
iprev	10	9	10	2	10
ipos	10	10	10	10	10
bgprev	1	0	4	7	50
bg	9	10	9	10	10
bgpos	0	10	0	10	10
insulin	1	6	4	10	0
instype	1	0	4	1	0
fib	1	4	4	4	0
fil	3	2	4	2	0
fid	1	4	2	8	0
sin	8	0	9	6	10
intake	9	10	10	10	10

the classifier based on fixed time intervals and compared the results obtained with the labels assigned by the classifier integrated in Sinedie, which were verified by the participants in the clinical study.

3. Results

In this section we present the results obtained in the machine learning techniques evaluation, considering FS and learning algorithms tests, the classification module structure specification and its performance in the *Sinedie* expert system where it was integrated.

3.1. Machine learning

3.1.1. Feature selection

Table 2 shows the number of times (from 0 to 10) that each feature was selected by each FS method during the 10-fold cross validation. The table is divided into three main columns (features name and the two evaluators tested, *Wrapper* and *CFS*). The column –corresponding to the *Wrapper* evaluator– is divided into two sub-columns, one for each searching algorithm used: *Best First* and *Genetic Search*. These two columns are subdivided into two sub-columns according to the learning algorithms tested (J48 and MLP). The *CFS* algorithm, as it evaluates attributes correlation, does not depend on any learning algorithm. *CFS* obtained the same results with the two searching algorithms so its results are shown in a single column, the last one. In the *Best First* algorithm, the result obtained with the “Forward” search strategy is shown as it had similar results than the “Bidirectional” strategy and slightly better results than the “Backwards” one in MLP.

All FS methods selected 7 features apart from the *wrapper-MLP-genetic* combination, which selects 9. There are 4 features selected by all the methods in 9 or 10 folds: “time”, “ipos”, “bg” and “intake”. There are two other features that were selected by all the methods except one: the “iprev” feature (not selected by the *wrapper-MLP-genetic*) and the slow insulin dose “sin” (not selected by the *wrapper-MLP-Best First*). The MLP combinations selected the boolean feature “insulin” in more folds than the “sin”, on the contrary of J48. The posterior BG measurement “bgpos” is selected by MLP in all the folds and *CFS* but it is never chosen by

the J48. The “work” feature seems not to have a high impact as it has been selected only in three methods in half of the 10 folds, but as it is shown in Table 3, it belongs to the combination that achieves the best accuracy.

3.1.2. Learning algorithms performance

Table 3 shows the accuracy results obtained by each candidate dataset obtained in the FS process, as well as with the DS_G , the DS_{18} and the DS_E with both learning algorithms (J48 and MLP).

The results show that the J48 learning algorithm achieves higher accuracy than MLP with all the datasets tested. Accuracy improves when increasing the attributes number from 2 (DS_G) to 18 (DS_{18}), and reaches the highest values when removing the redundant ones by feature selection (DS_{CFS} , DS_{W^*}). We achieved a 4% enhancement with the DS_E , adding two features to the initial DS_G . When adding the rest of features calculated, a total of 18 forming the DS_{18} , we obtained a slight improvement with the J48 algorithm (1.42%), and of 4.3% in MLP algorithm. This might be explained because the J48 achieved a 3.6% more accuracy than the MLP with the initial DS_G , remaining fewer margin for improvement. It is observed that with few features, 2–4, the J48 algorithm performs substantially better than the MLP but when we increase the features number up to 18 the performance is similar, over 94% with just a 0.67% of enhancement over the MLP. The MLP cross-validation execution time shows how the algorithm complexity also increases with the number of features. The J48 achieved the best accuracy result, 95.45% and it is also faster to be trained than the MLP (1 s vs. 2 min comparing the best accuracy results of each algorithm) so we decided to build the classification module based on the J48 learning algorithm discarding the MLP.

The selected feature subset was the one with highest accuracy, the *Wrapper* evaluator and the Genetic search algorithm (DS_{WG}) using the J48 learning algorithm.

3.2. Classification module structure

We implemented the automatic BG measurements classifier as a C4.5 decision tree, with the J48 algorithm. The features chosen as inputs were the 7 selected by the FS method formed by the wrapper evaluator and the genetic search algorithm: “time”, “ipre”, “ipos”, “bg”, “intake”, “work” and “sin”.

The classification procedure is divided in two stages: 1) acquisition of input features and; 2) classification of the glycaemia measurements. In the first stage (“Acquisition of inputs features”), the set of measurements uploaded by the patient (M_{upld}), formed by 3 input features (date, time and BG level), is transformed to obtain a measurement set composed by 7 inputs features (M_7), which are required by the classifier in the second stage (“Classification”). The 7 features are obtained from: 1) M_{upld} ; 2) M_{prev} : dataset with the glycaemia measurements from the previous download if they exist; and 3) S : dataset with the slow insulin doses related to the measurements uploaded, in case the patient uses insulin treatment. M_{prev} and M_{upld} with midnight time rectification (“cotime”) are the inputs to the cluster that obtains the dataset (I) with the “intake” features related to each measurement. Clusters are recalculated in each measurement’s upload so the “intake” feature adapts to possible changes in patients’ schedules due to pregnancy progress or to other reasons. When patients upload their glycaemia data for the first time, if the number of measurements is lower than three, the “intake” attribute is obtained with a simple logic rule based on fixed mealtime intervals to solve errors detected in the clustering results with a small number of measurements.

The “Classification” stage involves the classification of the M_7 dataset, obtaining as a result the measurements’ relationship to main meals. The classifier output (C) is a set of pairs of values

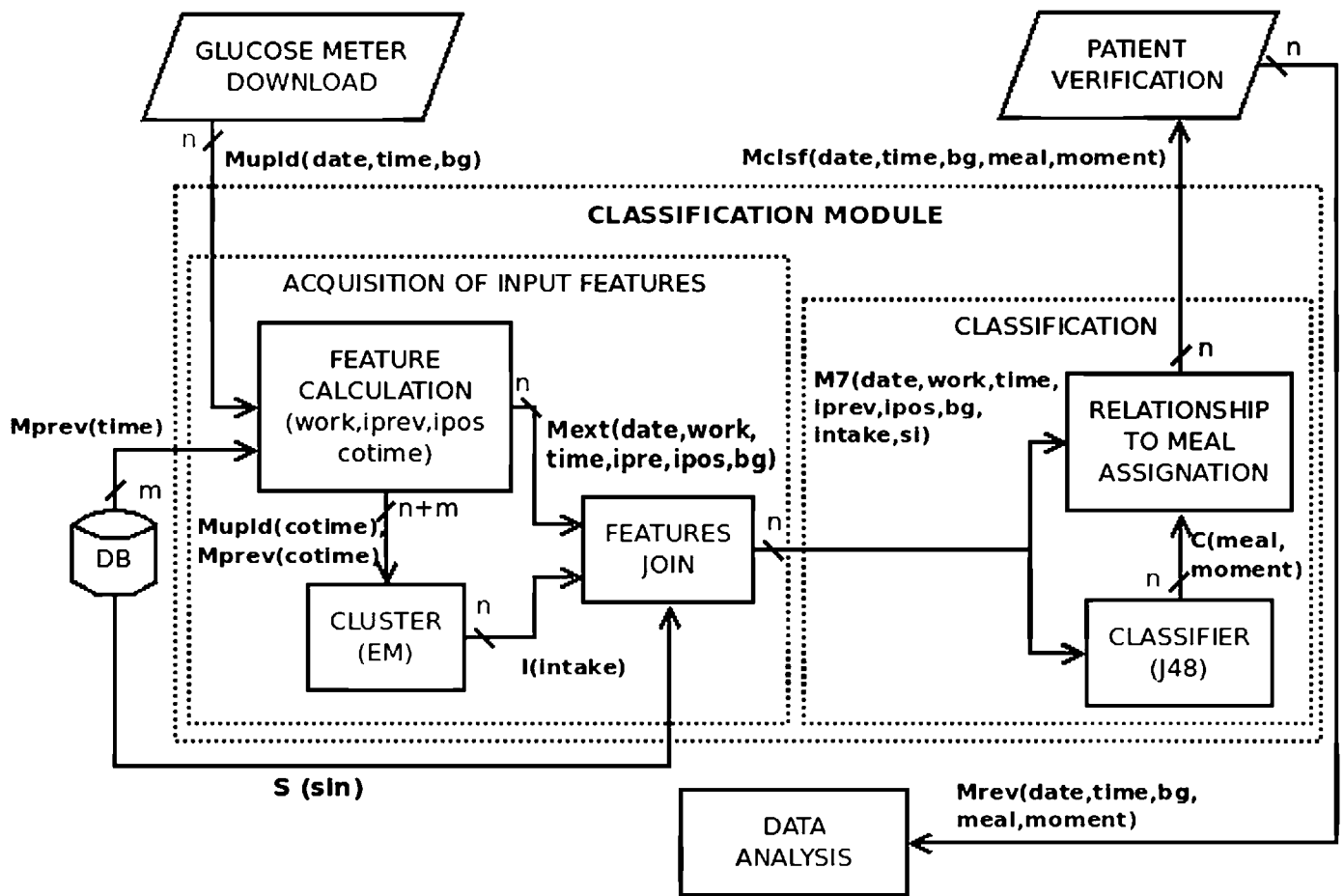


Fig. 4. Classification module structure diagram. The inputs and outputs of submodules are datasets whose features are indicated in brackets.

Table 3

Accuracy results of both learning algorithms with the 7 datasets tested with 10 fold cross-validation.

Set	J48			MLP		
	Accuracy	F-Measure	Execution Time	Accuracy	F-Measure	Execution Time
DS _G	89.20%	0.876	1s	85.61%	0.847	1 m13 s
DS _E	93.44%	0.925	1s	89.86%	0.895	1 m23 s
DS ₁₈	94.82%	0.947	1s	94.15%	0.938	8 m22 s
DS _{CFS}	95.19%	0.951	1s	94.58%	0.944	1 m59 s
DS _{WB}	95.20%	0.951	1s	95.10%	0.949	2 m4 s
DS _{WG}	95.45%	0.954	1s	95.00%	0.947	2 m26 s

DS_{CFS}, DS_{WB} y DS_{WG} correspond to features sets obtained by the CFS; and the wrapper evaluator with the Best First (WB) (Forwards) and the Genetic search algorithms (WG) respectively.

(meal-moment) related to each measurement uploaded. The four outputs related to the additional measurements “morning”, “afternoon”, “night” and “repeated” have moment “unknown”. Only incomplete measurements are automatically classified. Whenever a patient introduces either the meal or the moment in her glucose meter, this data is maintained and the missing value is completed with the automatic classifier output. Classified measurements (M_{clsf}) are presented to the patient through the *Sinedie* graphical user interface for verification, offering the possibility of reclassifying them (M_{rev}) before the automatic data analysis is performed by the expert system. The classification module structure is shown in Fig. 4.

3.3. Clinical evaluation

The classification module integrated in the *Sinedie* expert system achieved a 98.79% of accuracy when being used by patients.

Only 1.21% (86) of the measurements was reclassified by patients, out of the 7113 measurements transmitted by them and automatically classified by the system. Among the reclassified ones, 53.49% (46) were automatically classified as “repeated”, 15.12% (13) as “preprandial-breakfast”, 12.79% (11) as “postprandial lunch”, 10.47% (9) as “postprandial breakfast”, 3.49% (3) as “preprandial dinner”, 3.49% (3) as “postprandial dinner” and 1.16%(1) as “preprandial lunch”. Table 4 shows the reclassification changes made by the patients.

Most classification errors (46 out of 86) are made in the identification of repeated measurements. However, in half of these cases (23 out of 46) patients only reclassify the moment of measurement from “unknown” to “postprandial” or “preprandial” maintaining the meal determined by the classifier (“Repeated”). We consider the automatic classification of these 23 measurements correct, as the classifier does not specify the moment of measurement in repeated measurements.

Table 4
Classifier confusion matrix.

Aclsf ^a / Prclsf ^b	Bpre	Bpos	Morn.	Lpre	Lpos	Aftrn	Dpre	Dpos	Night	Runk	Rpre	Rpos
Bpre	1743	5	0	0	0	0	0	6	0	0	2	0
Bpos	1	1724	1	0	2	0	0	1	0	0	0	4
Morn.	0	0	0	0	0	0	0	0	0	0	0	0
Lpre	0	0	0	1	1	0	0	0	0	0	0	0
Lpos	0	2	4	2	1718	0	0	0	0	0	0	3
Aftrn	0	0	0	0	0	0	0	0	0	0	0	0
Dpre	0	0	0	0	1	2	0	0	0	0	0	0
Dpos	0	0	0	0	1	0	1	1706	1	0	0	0
Night	0	0	0	0	0	0	0	0	0	0	0	0
Runk	3	10	0	0	4	0	0	6	0	107	1	22

The rows show the tags assigned by the automatic classifier and the columns the tags assigned by the patients. Bpre: Breakfast preprandial; Bpos: Breakfast postprandial; Morn: Morning; Lpre: Lunch preprandial; Lpos: Lunch postprandial; Aftrn: Afternoon; Dpre: Dinner preprandial; Dpos: Dinner postprandial; Runk: Repeated unknown; Rpre: Repeated preprandial; and Rpos: Repeated postprandial.

^a Aclsf: Automatic classification

^b Prclsf: Patient reclassification.

Other noteworthy errors are the 6 measurements reclassified as dinner postprandial. These measurements belong to the same patient and were taken between 2a.m and 5a.m.

The accuracy achieved by the classifier based in Spanish pre-defined mealtimes, was 78.00%, a 21.04% lower than the classifier based in machine learning techniques.

4. Discussion

Machine learning techniques have confirmed our initial hypothesis about the need of using more variables than the 3 initially proposed (mealtime schedules, BG level and measurement time) to obtain better accuracy results that optimize the automatic classification process. Increasing the number of input features improved classification accuracy in both learning algorithms tested, but it is by the application of feature selection methods how we achieved the highest performance. The common 6 features selected by all the FS methods are: measurement time (time), time interval between measurements (iprev, ipost), BG level, insulin treatment (sin or insulin) and mealtime interval (intake). The selection of time intervals between the measurements confirms the importance of evaluating groups of three measurements instead of isolated ones. The best result in MLP selects the boolean attribute “insulin” instead of the integer “sin” selected in CFS or in C4.5 best results. The slow insulin is administered to lower fasting glycaemia, so its selection could help to differentiate between preprandial and postprandial measurements at breakfast time interval. Apart from the 6 common features mentioned before, while the CFS selects the “bgpos” as relevant, the Wrapper for the C4.5 does not. Instead, the combination that obtained the best result, *wrapper-C4.5 genetic*, selects the “work” attribute, which gives information about whether the measurement was taken during the weekend or not. A delay in the average measurement time during weekends was observed in all the patients but three, so this also validates the idea that this information should be considered.

The classification module presented in this paper was successfully integrated in the *Sinedie* expert system, where it was validated in a clinical environment during 8 months at one hospital. In this period, 42 patients sent a total of 7113 measurements which were completed by the automatic classifier developed. Only 1.21% of the measurements were reclassified by patients, so we consider that 98.79% were correctly classified by the module. The classifier presented has a good generalization ability, as the accuracy achieved in the clinical evaluation with new glycaemia data is 3% better than the one achieved in study 1 with the cross-validation of the training dataset.

Even if all patients had a glucose meter with the functionality to manually add the meal information to each measurement, the automatic classifier integrated in *Sinedie* would still be an essential part of the expert system to manage the incomplete measurements that patients may forget to label. In the clinical study, the *Sinedie* expert system was able to detect all situations that required a therapy adjustment due to altered glycaemia values, and all recommendations generated about diet modifications affected the appropriate meal intake.

The published studies on expert systems for diabetes that describe the method used for glycaemia measurements' classification do not report the accuracy achieved with the method used. For that reason, it is not possible to compare our results with the results obtained by other classification methods. Considering our results, the classifier defined in this article presents the strength of obtaining higher accuracy than the algorithms based on patients' mealtimes' schedules or on previous measurements. A greater accuracy in glycaemia classification results in lower patient intervention when registering monitoring data into expert systems, which speeds up the process and makes it less tedious. The potential limitation of the tested classifier would be determined by the training set used and the generalization ability in the presence of a different population than the one treated at the Hospital de Sabadell. In case of reduced accuracy with different populations the classifier could be trained with a different training set.

Possible future research lines include: 1) evaluation of the generalization ability of the classifier; 2) enhancement of the identification of repeated measurements and those belonging to optional moments of measurement (preprandial, morning, afternoon and night); 3) comparison of the results obtained with the classifier proposed and the ones that would be obtained with mealtime schedules specified by patients; 4) extension of the classifier functionality to include other types of diabetes like type 1 and type 2; 5) Integration of the classifier in a mobile application that connects with the *Sinedie* expert system.

To evaluate the generalization ability of the classifier it would be necessary to test it with patients from different regions from Spain or different countries. In case the classifier does not generalize correctly, it would require to be trained with a different data set from patients treated at the hospital where the classifier will be used.

Training the classifier with a more homogeneous dataset could improve the identification of repeated and optional measurements. In the dataset from study 1, which was used to train the GDM classifier, the measurements labeled as “repeated” only represent the 3.5%, of the whole data set. The training of the classifier can be performed in parallel without requiring a system interruption.

Then, it would be necessary to replace the file where the trained classifier is stored with the new retrained one, and reboot the system for the changes to take effect. In any case, to minimize the reclassification of this type of measurements is complicated, as the criteria about which measurement, either the first or the last one, should be considered as "Repeated", is ambiguous. We also observed that, sometimes, when facing several repeated measurements, some patients tend to label the ones with higher BG values as "Repeated".

In order to confirm the theory proposed in this article, about the accuracy enhancement that machine learning algorithms present in comparison to the accuracy achieved by algorithms based on patient's mealtime schedules, it would be interesting to compare the performance of both types of algorithms in the Sinedie system. Patients could specify their mealtime's schedules when they are registered in the system, with the possibility of modifying them in the system whenever they want. The evaluation could be done comparing the accuracy results obtained with the machine learning classifier (already integrated in Sinedie), and those that would be obtained using the mealtime schedules specified by each patient.

A similar methodology as the one presented in this paper could be followed to build a classifier for other types of diabetes. As glycaemia ranges and recommended moments of measurement are different depending on the type of diabetes, it would be necessary to implement a new classifier, in order to classify other types of diabetes measurements, performing a new feature selection with retrospective data of patients with the corresponding type of diabetes and create a new training dataset. Based on the learning curve of the presented GDM classifier, we estimate that the size of the dataset needed to perform a new study for a different type of diabetes should be over 5000 instances. However, from 2500 instances our classifier accuracy is stabilized at around 95% with the training dataset from study 1, achieving the maximum accuracy value with around 5000 instances.

The last future research direction identified is the implementation and integration of the classifier presented in the article into a mobile phone application. Mobile phones are smaller than laptops and therefore easier to carry when patients are out of home. Patients would download their monitoring data to their mobile phones and then the application would transmit the classified data to the Sinedie expert system for evaluation. Finally, patients would be able to verify the results of the data classification in the same mobile application.

5. Conclusions

Machine learning tools have proved to be effective to design an automatic classifier for glycaemia measurements, required to provide automatic recommendations based on an expert system for gestational diabetes. All methods tested obtained accuracy results above 80%, but the highest performance was achieved with the C4.5 decision tree learning algorithm with 7 inputs selected by a wrapper evaluator and the Genetic search algorithm. The selected classifier not only performed well with the training dataset with cross-validation, achieving a 95.45% of accuracy, but also in the Sinedie expert system, when being used to classify new glycaemia measurements uploaded by patients in clinical practice (98.79%).

The scarce methods available in literature for glycaemia classification based on previous measurements or predefined mealtimes present a high rate of errors that prevent expert systems from correctly determining patients' condition or generating recommendations about therapy adjustments. If the expert system detects altered glycaemia values but it is not able to determine whether they are caused by the breakfast or the lunch, it will not be able

to recommend the reduction of carbohydrates in the appropriate intake or the administration of insulin to metabolize them.

Automatic classification is essential in Sinedie expert system to analyze incomplete glycaemia measurements registered, to minimize patients' intervention, to automatically detect the patients' metabolic condition and to generate recommendations about therapy changes in specific mealtimes. In addition, the information added by the classification module in the Sinedie expert system allows structuring measurements in relationship to meals so that the medical team can visualize and evaluate them as fast as they usually do with paper logbooks.

Acknowledgment

This work has been funded by the Spanish grant Sinedie (PIIO/O1125), co-funded by FEDER. We would like to thank the patients of Hospital de Sabadell for their collaboration and support in this research.

References

- Boney, C. M., Verma, A., Tucker, R., & Vohr, B. R. (2005). Metabolic syndrome in childhood: association with birth weight, maternal obesity, and gestational diabetes mellitus. *Pediatrics*, *115*(3), e290–e296. doi:10.1542/peds.2004-1808.
- Bourouis, A., Feham, M., Hossain, M. A., & Zhang, L. (2014). An intelligent mobile based decision support system for retinal disease diagnosis. *Decision Support Systems*, *59*, 341–350. doi:10.1016/j.dss.2014.01.005.
- Bromuri, S., Puricel, S., Schumann, R., Krampf, J., Ruiz, J., & Shumacher, M. (2016). An expert Personal Health System to monitor patients affected by Gestational Diabetes Mellitus: A feasibility study. *Journal of Ambient Intelligence and Smart Environments*, *8*(2), 219–237.
- Cafazzo, J. A., Casselman, M., Hamming, N., Katzman, D. K., & Palmert, M. R. (2012). Design of an mHealth app for the management of adolescent type 1 diabetes: A pilot study. *Journal of Medical Internet Research*, *14*(3), e70. doi:10.2196/jmir.2058.
- Carral, E., Ayala, M., del, C., Fernández, J. J., González, C., Piñero, A., & García, G. (2015). Web-based telemedicine system is useful for monitoring glucose control in pregnant women with diabetes. *Diabetes Technology & Therapeutics*, *17*(5), 349–354. doi:10.1089/dia.2014.0223.
- Dalfrà, M. G., Nicolucci, A., & Lapolla, A. on behalf of the TISG. (2009). The effect of telemedicine on outcome and quality of life in pregnant women with diabetes. *Journal of Telemedicine and Telecare*, *15*(5), 238–242. doi:10.1258/jtt.2009.08.1213.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(1), 1–38.
- Duran, A., Saenz, S., Torrejon, M. J., Bordiú, E., del Valle, L., & Galindo, M. (2014). Introduction of IADPSG criteria for the screening and diagnosis of gestational diabetes mellitus results in improved pregnancy outcomes at a lower cost in a large cohort of pregnant women: The St. carlos gestational diabetes study. *Diabetes Care*, *37*(9), 2442–2450. doi:10.2337/dc14-0179.
- El-Gayar, O., Timsina, P., Nawar, N., & Eid, W. (2013a). A systematic review of IT for diabetes self-management: Are we there yet? *International Journal of Medical Informatics*, *82*(8), 637–652. doi:10.1016/j.ijmedinf.2013.05.006.
- El-Gayar, O., Timsina, P., Nawar, N., & Eid, W. (2013b). Mobile applications for diabetes self-management status and potential. *Journal of Diabetes Science and Technology*, *7*(1), 247–262.
- Ferrara, A., Hedderson, M. M., Ching, J., Kim, C., Peng, T., & Crites, Y. M. (2012). Referral to telephonic nurse management improves outcomes in women with gestational diabetes. *American Journal of Obstetrics and Gynecology*, *206*(6), 491.e1–491.e5. doi:10.1016/j.ajog.2012.04.019.
- Franks, P. W., Looker, H. C., Kobes, S., Touger, L., Tataranni, A., & Hanson, R. I. (2006). Gestational Glucose Tolerance and Risk of Type 2 Diabetes in Young Pima Indian Offspring. *Diabetes*, *55*, 460–465.
- García-Sáez, G., Alonso, J.M., & Molero, J., Rigla, M., Martínez-Sarriegui, I., de Leiva, A. et al. (2009). Mealtime blood glucose classifier based on fuzzy logic for the diabetel telemedicine system. In *Lectures Notes in Artificial Intelligence* (Vol. 5651, pp. 295–304). Verona, Italy: Springer. http://doi.org/10.1007/978-3-642-02976-9_42.
- Given, J. E., O'Kane, M. J., Bunting, B. P., & Coates, V. E. (2013). Comparing patient-generated blood glucose diary records with meter memory in diabetes a systematic review.pdf. *Diabetic Medicine*, *30*, 901–913. doi:10.1111/dme.12130.
- Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley.
- Hall, M.A. (1999). Correlation-based feature selection for machine learning. The University of Waikato. Retrieved from <https://www.lri.fr/~pierres/donn%E9es/save/these/articles/lpr-queue/hall99correlationbased.pdf>
- Hartling, L., Dryden, D. M., Guthrie, A., Muise, M., Vandermeer, B., & Donovan, L. (2013). Benefits and harms of treating gestational diabetes mellitus: A systematic review and meta-analysis for the US preventive services task force and the national institutes of health office of medical applications of research. *Annals of Internal Medicine*, *159*(2), 123–129.

- Haykin, S.S. (1994). *Neural networks: A comprehensive foundation*. New York: Macmillan College Publishing Company Inc.
- Hernando, M. E., Gómez, E. J., Corcoy, R., & del Pozo, F. (2000). Evaluation of DIABNET, a decision support system for therapy planning in gestational diabetes. *Computer Methods and Programs in Biomedicine*, 62(3), 235–248.
- Hische, M., Luis-Dominguez, O., Pfeiffer, A. F. H., Schwarz, P. E., Selbig, J., & Spranger, J. (2010). Decision trees as a simple-to-use and reliable tool to identify individuals with impaired glucose metabolism or type 2 diabetes mellitus. *European Journal of Endocrinology*, 163(4), 565–571. doi:10.1530/EJE-10-0649.
- Huang, Y., McCullagh, P., Black, N., & Harper, R. (2007). Feature selection and classification model construction on type 2 diabetic patients' data. *Artificial Intelligence in Medicine*, 41(3), 251–262. doi:10.1016/j.artmed.2007.07.002.
- International Association of Diabetes and Pregnancy Study Groups Consensus Panel. (2010). International Association of Diabetes and Pregnancy Study Groups Recommendations on the Diagnosis and Classification of Hyperglycemia in Pregnancy. *Diabetes Care*, 33(3), 676–682. doi:10.2337/dc09-1848.
- International Diabetes Federation. (2015). *IDF diabetes atlas seventh edition* (7th ed). Brussels, Belgium: International Diabetes Federation, Retrieved from www.diabetesatlas.org.
- Klonoff, D. C., & True, M. W. (2009). The missing element of telemedicine for diabetes: Decision support software. *Journal of Diabetes Science and Technology*, 3(5), 996–1001.
- Kohavi, R., & John, G. H. (1996). Wrappers for feature subset selection. *Artificial Intelligence*, 97, 273–324.
- Lim, S., Kang, S. M., Shin, H., Lee, H. J., Won Yoon, J., & Yu, S. H. (2011). Improved glycemic control without hypoglycemia in elderly diabetic patients using the ubiquitous healthcare service, a new medical information system. *Diabetes Care*, 34(2), 308–313. doi:10.2337/dc10-1447.
- Liu, H., & Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering*, 17(4), 491–502.
- Mackillop, L., Loerup, L., Bartlett, K., Farmer, A., Gibson, O. J., Hirst, J. E., et al. (2014). Development of a Real-Time Smartphone Solution for the Management of Women With or at High Risk of Gestational Diabetes. *Journal of Diabetes Science and Technology*, 8(6), 1105–1114. doi:10.1177/1932296814542271.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1, 281–297.
- Metzger, B. E., Lynn, M. D., Lowe, P., & Dyer, A. R. Northwestern University Feinberg School of Medicine, Trimble, E. R. The HAPO Study Cooperative Research Group. (2008). Hyperglycemia and Adverse Pregnancy Outcomes. *The New England Journal of Medicine*, 358, 1991–2002. doi:10.1056/NEJMoa0707943.
- Mohan, V., & Pradeepa, R. (2014). Telemedicine in Diabetes Care: In rural India, a new prevention project seeks to fill in the screening gap. *IEEE Pulse*, 5(3), 22–25. doi:10.1109/MPUL.2014.2309575.
- Möhlig, M., Flöter, A., Spranger, J., Weickert, M. O., Schill, T., & Schlösser, W. (2006). Predicting impaired glucose metabolism in women with polycystic ovary syndrome by decision tree modelling. *Diabetologia*, 49(11), 2572–2579. doi:10.1007/s00125-006-0395-0.
- Pelleg, D., & Moore, A. (2000). X-means. Extending K-means with efficient estimation of the number of clusters. In *Proceedings of the Seventeenth International Conference on Machine Learning* (pp. 727–734). San Francisco: Morgan Kaufmann. Retrieved from <http://www.cs.cmu.edu/~dpelleg/download/xmeans.pdf>.
- Pérez Ferre, N., Galindo, M., Fernández, M. D., Velasco, V., de la Cruz, M. J., & Martín, P. (2010). A Telemedicine system based on Internet and short message service as a new approach in the follow-up of patients with gestational diabetes. *Diabetes Research and Clinical Practice*, 87(2), e15–e17. doi:10.1016/j.diabres.2009.12.002.
- Polonsky, W. H., Jelsovsky, Z., Panzera, S., Parkin, C. G., & Wagner, R. S. (2009). Primary care physicians identify and act upon glycemic abnormalities found in structured, Episodic blood glucose monitoring data from non-insulin-treated type 2 Diabetes. *Diabetes Technology & Therapeutics*, 11(5), 283–291. (Version 0.7.9). Retrieved from PSPP. <https://www.gnu.org/software/pspp/>.
- Quinlan, J. Ross (1993). *C4.5: programs for machine learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- Quinn, C. C., Clough, S. S., Minor, J. M., Lender, D., Okafor, M. C., & Gruber-Baldini, A. (2008). WellDoc mobile diabetes management randomized controlled trial: change in clinical and behavioral outcomes and patient and physician satisfaction. *Diabetes Technology & Therapeutics*, 10(3), 160–168. doi:10.1089/dia.2008.0283.
- Retrieved from Diagnostics, Roche. *Accu-Check Aviva user's manual* <https://www.accu-check.ca/multimedia/documents/aviva2-user-manual.pdf>.
- Retrieved from Diagnostics, Roche. *Accu-Check Nano user's manual* <https://www.accu-check.com/us/glucose-meters/nano-how-to-use.html>.
- Diabetes, Sanofi (2015). *iBGStar® Diabetes Manager Application - MyStar (Version 2.2.1)*. Sanofi Diabetes. Retrieved from <http://www.mystarsanofi.com/web/products/software/iBGstar-application>.
- Shankaracharya, Odedra, D., Mallick, M., Shukla, P., Samanta, S., & Vid-yarthi, A. S. (2012). Java-based diabetes type 2 prediction tool for better diagnosis. *Diabetes Technology & Therapeutics*, 14(3), 251–256. doi:10.1089/dia.2011.0202.
- Upadhyaya, S., Farahmand, K., & Baker-Demaray, T. (2013). Comparison of NN and LR classifiers in the context of screening native American elders with diabetes. *Expert Systems with Applications*, 40(15), 5830–5838. doi:10.1016/j.eswa.2013.05.012.
- Witten, I. H., & Frank, E. (2005). *Data Mining, practical machine learning tools and techniques* (2nd ed). San Francisco: Morgan Kaufmann Publishers.
- Wojcicki, J. M., Ladyzynski, P., Krzymien, J., Jozwicka, E., Blacjowic, J., & Janczewska, E. (2001). What we can really expect from telemedicine in intensive diabetes treatment: Results from 3-year study on type 1 pregnant diabetic women. *Diabetes Technology & Therapeutics*, 3(4), 581–589.