

ORIGINAL

Identificación de patrones de glucemia en pacientes con diabetes tipo 1 mediante monitorización continua de glucosa y técnicas de clusterización

Sergio Contador Pachón^a, Marta Botella Serrano^b, Oscar Garnica Alcázar^c, José Manuel Velasco Cabo^c, Aranzanzu Aramendi Zurimendi^b, Remedios Rodríguez Martínez^b, Esther Maqueda Villaizán^d y José Ignacio Hidalgo Pérez^{c,*}

^a Universidad Rey Juan Carlos, Móstoles, Madrid, España

^b Servicio de Endocrinología y Nutrición, Hospital Universitario Príncipe de Asturias, Alcalá de Henares, Madrid, España

^c Departamento de Arquitectura de Computadores y Automática, Universidad Complutense de Madrid, Madrid, España

^d Servicio de Endocrinología y Nutrición, Hospital Virgen de la Salud, Toledo, España

Recibido el 28 de agosto de 2019; aceptado el 24 de diciembre de 2019



PALABRAS CLAVE

Diabetes mellitus;
Monitorización
continua de glucosa;
Clasificación;
Clusterización;
Detección automática
de interacción
chi-cuadrado

Resumen

Objetivo: Demostrar que mediante técnicas estadísticas es posible obtener un número reducido de perfiles de glucosa de 4 h con los que se puede identificar cualquier comportamiento de la glucosa en pacientes con diabetes mellitus tipo 1.

Material y métodos: Se ha realizado un estudio retrospectivo de 10 pacientes con diabetes mellitus tipo 1, con datos adquiridos mediante monitorización continua de glucosa. Se ha utilizado una técnica de minería de datos basada en árboles de decisión denominada CHAID (*CHi-square Automatic Interaction Detection*) para clasificar los perfiles de glucosa en grupos utilizando dos criterios de decisión. Por un lado, los diferentes días (lunes; martes; miércoles; jueves; viernes; sábado; domingo); por otro, diferentes franjas del día, dividiendo el día en seis tramos de 4 h cada uno (00:00-04:00 h; 04:00-08:00 h; 08:00-12:00 h; 12:00-16:00 h; 16:00-20:00 h; 20:00-24:00 h). Las agrupaciones se han realizado de acuerdo a los niveles de glucosa registrados, mediante diferencias estadísticamente significativas encontradas.

Resultados: Se han observado diferencias significativas ($p < 0,05$) y dependencias entre los perfiles de glucosa clasificados en función de las variables independientes día de la semana y franja horaria, siendo las relaciones encontradas distintas para cada paciente, demostrando la necesidad de hacer un estudio individualizado.

* Autor para correspondencia.

Correo electrónico: hidalgo@ucm.es (J.I. Hidalgo Pérez).

KEYWORDS

Diabetes mellitus;
Continuous glucose monitoring;
Classification;
Clustering;
Chi-square automatic detection interaction

Conclusiones: Los resultados obtenidos facilitarán la modelización matemática de la glucosa y podrán utilizarse para la creación de un clasificador individualizado para cada paciente que clasifique los perfiles de glucosa en función de las variables día de la semana y franja horaria. Utilizando este clasificador se podrán predecir los valores de glucosa del paciente conociendo en qué día de la semana se encuentra y en qué franja horaria, obteniendo modelos más precisos. También el profesional de la salud podrá mejorar los hábitos y las terapias de los pacientes.

© 2020 SEEN y SED. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

Identification of blood glucose patterns in patients with type 1 diabetes using continuous glucose monitoring and clustering techniques

Abstract

Objective: To show that statistical techniques allow for obtaining a reduced number of four-hour glucose profiles that can identify any glucose behavior in patients with type 1 diabetes mellitus.

Material and methods: A retrospective study of 10 patients with type 1 diabetes mellitus was conducted using data collected by continuous glucose monitoring. A data mining technique based on decision trees called CHAID (Chi-square Automatic Interaction Detection) was used to classify glucose profiles into groups using two decision criteria. These were: 1, the seven days of the week, and 2, different time slots, the day being divided into six sections of four hours each. Clustering was performed according to the glucose levels recorded using the statistically significant differences found.

Results: Significant differences ($P < .05$) and dependencies were seen between the glucose profiles classified depending on the independent variables 'day of the week' and 'time slot'. The relationships found were different for each patient, showing the need for individualized studies.

Conclusions: The results obtained will facilitate mathematical modeling of glucose, and can be used to develop an individualized classifier for each patient that categorizes glucose profiles based on the day of the week and time slot variables. Using this classifier, it will be possible to predict the glucose levels of the patient knowing on which day of the week and in which time slot he/she is, leading to more precise models. Healthcare professionals will also be able to improve patient habits and therapies.

© 2020 SEEN y SED. Published by Elsevier España, S.L.U. All rights reserved.

Introducción

La colección de métodos conocidos como minería de datos ofrece soluciones técnicas y metodológicas para resolver problemas de análisis de datos médicos y creación de modelos de predicción. La minería de datos es el proceso de seleccionar, explorar y modelar grandes cantidades de datos para encontrar y descubrir patrones o relaciones desconocidas que brinden un resultado claro y útil¹. Es un campo de la ciencia que se ha desarrollado rápidamente en los últimos años y que ayuda a explicar los datos y a ganar conocimiento sobre ellos². Los árboles de decisión son una de las técnicas estadísticas de clasificación más potentes utilizadas en minería de datos³⁻⁸. Un árbol de decisión es una vía clara y concisa para poder examinar y tomar una decisión acerca de las posibles relaciones entre los datos, identificando grupos o segmentos de interés entre ellos.

La realización del presente trabajo tuvo como objetivo identificar, mediante la construcción de árboles de decisión, perfiles de glucosa clasificados en grupos obtenidos mediante las variables día de la semana (lunes; martes; miércoles; jueves; viernes; sábado; domingo) y franja horaria, definida como la división de los valores de glucosa en tramos de 4 h de duración (00:00-04:00 h; 04:00-08:00 h; 08:00-12:00 h; 12:00-16:00 h; 16:00-20:00 h; 20:00-24:00 h).

El resto del artículo está organizado de la siguiente manera. En la siguiente sección se describen los datos utilizados y la técnica empleada para la clasificación de la glucosa. El trabajo experimental y los resultados se muestran en la sección «Resultados». La discusión de los resultados y las conclusiones se exponen en la sección «Discusión».

Material y métodos

Pacientes

Se realiza un estudio retrospectivo de 10 pacientes con diabetes mellitus tipo 1. Se registraron mediciones cada 5 min utilizando sensores de monitorización continua de glucosa (MCG) Guardian Real Time y bombas de insulina Minimed, de la compañía Medtronic. Así mismo, se registraron las estimaciones de hidratos de carbono realizadas por los pacientes formados en el proceso de dieta por raciones. Las mediciones se realizaron en días no necesariamente consecutivos ni tampoco los mismos días para cada paciente. Se contemplan únicamente los tramos de 4 horas con al menos 46 valores. La tabla 1 muestra la caracterización de los pacientes con información del sexo, edad, peso, HbA1c medido en los 3 meses previos al estudio, y los períodos que llevan con el diagnóstico de la enfermedad y con el tratamiento con la bomba de insulina. Además, para cada paciente se muestran en esa misma tabla la glucosa media, la desviación estándar y los porcentajes de tiempo donde el paciente tiene niveles de glucosa por debajo de 70 mg/dl, por encima de 250 mg/dl y en el tiempo en rango [70,180] mg/dl.

Métodos estadísticos

Los árboles de decisión son una técnica de minería de datos que explora los datos para extraer información oculta en ellos. El objetivo de la construcción del árbol de decisión es crear un modelo para predecir el valor

Tabla 1

Paciente	1	2	3	4	5	6	7	8	9	10
Sexo	mujer	mujer	hombre	mujer	mujer	mujer	mujer	mujer	hombre	mujer
Edad [años]	45	40	70	38	41	31	45	33	34	46
Peso [kg]	74.7	52.0	94.0	55.4	67.3	60.6	50.5	55.5	73.6	64.2
HbA1c [%]	7.2	7.1	7.4	6.9	6.9	8.5	7.6	6.9	7.3	6.9
Diagnóstico [años]	39	38	40	26	24	18	36	14	14	23
Terapia bomba [años]	14.0	9.0	15.0	13.0	8.0	12.0	12.0	1.5	1.5	14.0
\bar{G} [mg/dl]	157.67	145.44	143.46	150.47	139.17	142.54	176.33	135.34	146.82	166.13
Std [mg/dl]	62.25	64.02	45.45	56.70	67.93	60.08	68.35	46.11	59.82	86.12
$T \cdot G < 70$ [%]	4.48	8.33	2.14	4.12	14.17	10.08	3.65	4.85	7.54	9.33
$T \cdot G > 250$ [%]	7.31	6.18	2.26	5.14	7.10	4.44	14.28	1.84	5.20	16.16
$T \cdot G^{[70, 180]}$ [%]	37.57	41.83	49.18	41.29	41.89	41.21	27.07	54.06	39.49	35.69
Días	148	181	182	267	190	151	117	82	122	135
Valores de glucosa	37854	49634	53964	65654	43007	40500	27756	18062	28156	32817

Tabla 2

Paciente	1	2	3	4	5	6	7	8	9	10
Profundidad de árbol	2	2	3	2	2	3	2	2	2	2
Número de nodos	33	42	42	36	40	42	32	25	28	46
Número de nodos finales	23	30	28	23	27	32	24	17	21	31

de una variable dependiente/objetivo a partir de las variables independientes/predictoras consideradas. El árbol de decisión tiene 3 tipos de nodos, a saber, el nodo raíz, los nodos internos y los finales, cada uno representando a una clase caracterizada por los valores estadísticos de la variable objetivo, y las categorías de las variables predictoras que contiene cada nodo. Cada camino en la construcción del árbol de decisión está asociado a una regla de decisión establecida por el propio algoritmo. Así pues, de acuerdo con las reglas establecidas, el conjunto de datos es dividido de forma recursiva en subconjuntos independientes de datos más pequeños (algoritmo divisorio). Uno de los algoritmos más utilizados es el algoritmo de detección automática de interacción chi-cuadrado o CHAID (*CHi-square Automatic Interaction Detection*)⁹, empleado en nuestro estudio con el software de análisis predictivo IBM SPSS v.21¹⁰. Este algoritmo divide recursivamente los datos mediante una variable respuesta/objetivo usando múltiples divisiones entre las distintas variables de entrada/predictoras. Una división debe alcanzar un nivel umbral de significación entre los valores nominales de la variable objetivo y las ramas, o bien el nodo no se divide. La búsqueda termina cuando no se pueden juntar más ramas o no hay divisiones significativas. La última división se elige como la solución. Hay que tener en cuenta que la última división no tiene por qué ser la división más significativa examinada, es decir, podría existir otra agrupación más significativa diferente de la mostrada en la tabla, aunque esto es una propiedad intrínseca al algoritmo utilizado.

En este estudio se ha utilizado como variable objetivo los niveles de glucosa de cada paciente, como variables predictoras el día de la semana y la franja horaria, y como umbral de significación un nivel de confianza del 95% ($\alpha = 0,05$). Se ha utilizado el estadístico F de Snedecor como criterio de división y el ajuste Bonferroni para el número de valores categóricos de la variable de entrada, mitigando así el sesgo hacia entradas con muchos valores.

Resultados

Se realiza un estudio individualizado para cada paciente. Para la construcción de los árboles de decisión de cada uno de los pacientes se ha seleccionado una profundidad máxima de árbol de 3, un número mínimo de casos en el nodo padre de 100 y en el nodo hijo de 50. La profundidad final del árbol, el número de nodos y el número de nodos finales obte-

nidos para cada paciente se muestran en la **tabla 2**. El primer predictor utilizado en la construcción del árbol ha sido el día de la semana y el segundo predictor ha sido franja horaria.

En total tenemos siete categorías para la variable día de la semana y seis categorías para la variable franja horaria. Las categorías de las variables se representan con letras y números, como muestra la **tabla 3**. La **tabla 4** representa los grupos obtenidos para cada paciente. En un primer nivel aparecen los agrupamientos por día. Por ejemplo, podemos ver que en el paciente 1 se distinguen cuatro patrones glucémicos diferentes: uno para el martes, otro para el jueves, otro para el viernes y un último patrón para los lunes-miércoles-sábados-domingos, donde el menor valor de glucosa media es para los viernes ($163,53 \pm 58,06$ mg/dl) y el mayor es para los martes ($179,01 \pm 55,61$ mg/dl). En el segundo nivel aparecen los agrupamientos por franja horaria. El tamaño de los grupos en este nivel está separado por colores, y se puede apreciar que en algunos pacientes existen agrupaciones con una gran cantidad de franjas. En el caso del paciente 1, podemos decir que el viernes habitualmente tiene dos comportamientos diferentes, uno para la franja [20:00-04:00 h] y otro para el resto del día. El mismo análisis se puede realizar para el resto de los pacientes.

Tabla 4 Resultados obtenidos con el algoritmo CHAID para las agrupaciones obtenidas teniendo en cuenta los diferentes días de la semana y las franjas horarias. G es el promedio de glucosa, Std es la desviación estándar, y TG es el porcentaje de tiempo que los valores de glucosa pasan en cada rango (< 70 mg/dl, > 250 mg/dl y [70, 180] mg/dl o tiempo en rango). Los colores en la *franja horaria* representan el número de elementos que forman un grupo. El color azul se utiliza para representar agrupaciones de 1 elemento, el verde con 2, el amarillo con 3 y el rojo con 4

	\bar{G} [mg/dl]	Std [mg/dl]	$T \cdot G < 70$ [%]	$T \cdot G > 250$ [%]	$T \cdot G^{[70, 180]}$ [%]	Día de la semana	Franja horaria
paciente 1	170.62	61.36	2.22	10.45	57.56	L-X-S-D	0-1 2 3 4 5
	179.01	71.15	0.65	13.65	55.61	M	0-1 2-4 3 5
	167.22	61.34	1.47	9.81	61.28	J	0 1-2-3 4 5
	163.53	58.06	2.40	8.43	62.55	V	0-5 1-2-3-4
paciente 2	146.28	59.40	6.13	5.08	68.67	L	0 1-2-4 3 5
	152.32	58.96	4.42	6.31	67.77	M	0 1 2 3 4 5
	158.59	63.25	4.90	5.91	61.93	X-D	0 1 2-4 3 5
	151.08	57.70	5.56	5.92	66.63	J-S	0 1 2 3 5
	148.46	56.29	4.56	5.37	69.56	V	0-2 1 3-5 4
paciente 3	141.65	43.21	2.45	2.02	80.95	L-M	0-3 1 2-5 4
	148.25	44.30	2.03	2.56	76.96	X-D	0-1 2-3 4 5
	154.49	51.03	0.76	4.92	73.06	J	0-2-3 1 4-5
	139.64	42.90	1.92	0.96	81.83	V	0-2 1 3 4 5
	135.85	38.38	1.59	0.53	85.33	S	0 1 2-3-5 4
paciente 4	162.95	58.68	2.96	8.50	61.49	L-X	0 1 2 3 4-5
	152.25	53.76	2.88	4.31	71.18	M-V	0 2-3-4 1 5
	137.29	57.62	2.59	5.78	68.48	J-D	0 1-3 4 5
	154.55	53.01	2.11	5.41	69.49	S	0 2-3-4 1 5
paciente 5	141.40	66.10	10.71	7.19	62.65	L-V	0 1-3 2 4 5
	148.19	65.68	8.46	8.11	63.67	M-X	0 1-3-4 2 5
	140.77	57.92	8.69	3.91	68.96	J	0-2 1 3 4 5
	128.98	55.63	13.65	2.20	67.38	S	0 1-2-3-4 5
	132.90	64.48	11.19	6.45	69.40	D	0-1 2-4 3-5
paciente 6	156.91	61.08	5.81	7.43	61.41	L-M-S	0 1 2-4-5 3
	144.54	54.08	8.41	2.76	65.37	X	0 1-2 3 4 5
	140.60	54.43	7.17	5.74	71.34	V	0 1 2 3 4 5
	152.51	68.46	8.68	6.78	63.12	D	0 1-5 2 3 4
	147.10	64.95	10.42	6.67	63.43		0 1-5 2 3 4
paciente 7	173.27	70.28	3.28	14.77	55.49	L-S	0 1 2-4 3 5
	170.80	64.61	2.06	10.96	61.07	M	0-1-2 3 4-5
	179.45	67.80	3.08	15.08	50.18	X	0 1 2-4 3-5
	161.55	60.39	3.40	8.71	62.60	J-V	0-4 1-5 2-3
	173.84	63.19	3.07	11.85	53.03	D	0 1-3-4 2 5
paciente 8	138.60	42.17	2.83	1.03	81.14	L	0-1-3 2-4 5
	128.27	44.25	5.40	1.05	82.81	M-X-S	0-1-2 3 4-5
	131.77	48.90	7.13	1.95	77.48	J-V	0 1-4 2-5 3
	143.04	43.81	6.01	1.98	71.83	D	0 1 2-3 4-5
paciente 9	145.90	55.31	7.17	4.83	66.65	L	0-1-2-3 4 5
	150.42	61.20	4.25	5.74	66.93	M-X-D	0-1-2 3 4-5
	143.12	53.96	7.13	3.09	69.57	X	0 1-4-5 2 3
	155.23	62.26	3.25	7.84	66.36	V	0-3-5 1 2 4
paciente 10	172.27	76.22	4.58	14.29	48.78	L	0-2 1-3-5 4
	168.56	78.29	7.84	15.96	52.72	M-V	0 1 2 3 4 5
	163.74	76.13	10.45	14.86	51.35	X	0 1-3-5 2 4
	181.96	85.28	6.77	21.44	46.92	J-D	0 1-2 3 4 5
	190.48	90.48	4.89	23.50	45.24	S	0-1 2 3-5 4

Tabla 3

Día de la semana	Identificador	Franja horaria	Identificador
Lunes	L	00:00h-04:00h	0
Martes	M	04:00h-08:00h	1
Miércoles	X	08:00h-12:00h	2
Jueves	J	12:00h-16:00h	3
Viernes	V	16:00h-20:00h	4
Sábado	S	20:00h-24:00h	5
Domingo	D		

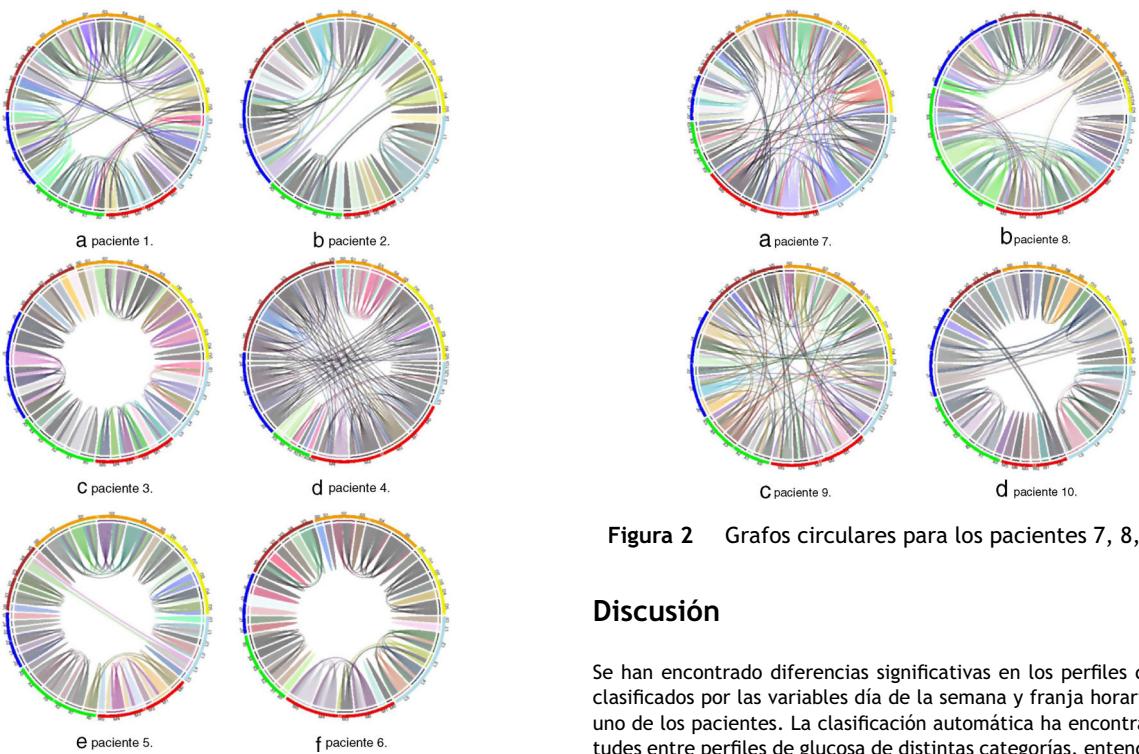


Figura 1 Grafos circulares para los pacientes 1, 2, 3, 4, 5 y 6.

La [tabla 5](#) incluye información de control glucémico sobre los resultados de la [tabla 4](#). Para favorecer la interpretación de resultados se incluye únicamente la información del paciente 10. Los datos del resto de los pacientes se incluyen como [material adicional](#).

Las relaciones encontradas entre las variables independientes y los valores de glucosa se representan también como gráficos circulares. Los gráficos se han creado utilizando la librería *circlize*¹¹, con el software libre de análisis estadístico R versión 3.5.2¹².

Las [figuras 1 y 2](#) representan los gráficos circulares de los resultados de la [tabla 4](#). Para cada paciente se crea un gráfico circular dividido en 7 segmentos, uno por cada día, y cada segmento en otros 6, correspondientes a cada franja ([tabla 3](#)) y las líneas representan las relaciones encontradas entre los días y las franjas.

Tabla 5 Resultados obtenidos con el algoritmo CHAID para las agrupaciones obtenidas teniendo en cuenta los diferentes días de la semana y las franjas horarias. G es el promedio de glucosa, Std es la desviación estándar, y TG es el porcentaje de tiempo que los valores de glucosa pasan en cada rango ($< 70 \text{ mg/dl}$, $> 250 \text{ mg/dl}$ y $[70, 180] \text{ mg/dl}$ o tiempo en rango). Los colores en la franja horaria representan el número de elementos que forman un grupo. El color azul se utiliza para representar agrupaciones de 1 elemento, el verde con 2, el amarillo con 3 y el rojo con 4.

Día de la semana	Franja horaria	\bar{G} [mg/dl]	Std [mg/dl]	$T.G_{<70}$ [%]	$T.G_{>250}$ [%]	$T.G_{[70,180]}$ [%]	\bar{T} [U/4h]	\bar{C} [gr]
paciente 10								
L	(0-2)	159.21	71.65	5.20	11.22	61.01	2.63	13.75
L	(1-3-5)	179.09	75.79	4.49	13.69	46.73	3.66	29.67
L	(4)	202.40	77.01	3.79	20.98	34.04	5.47	36.50
M-V	(0)	142.51	76.45	15.00	8.48	62.62	2.32	6.50
M-V	(1)	163.57	69.95	6.32	13.74	58.54	3.21	41.50
M-V	(2)	148.95	64.78	9.55	8.35	58.12	4.06	36.25
M-V	(3)	156.74	75.08	11.90	11.58	52.61	2.81	21.00
M-V	(4)	194.02	79.76	1.91	24.54	45.86	5.62	44.00
M-V	(5)	202.58	81.48	2.50	28.11	40.94	4.74	36.75
X	(0)	154.82	80.87	11.64	12.70	55.42	2.64	7.50
X	(1-3-5)	166.01	80.11	11.83	16.8	47.14	3.53	30.50
X	(2)	145.31	60.95	10.98	7.23	67.47	3.73	39.00
X	(4)	190.19	63.54	3.99	19.82	41.78	4.25	15.50
J-D	(0)	174.50	89.03	5.97	19.30	52.19	2.84	9.25
J-D	(1-2)	161.44	75.46	7.48	13.23	57.97	3.03	22.50
J-D	(3)	166.24	78.74	9.78	13.98	51.29	3.35	25.50
J-D	(4)	226.94	86.61	3.90	42.04	24.95	6.04	38.75
J-D	(5)	194.15	83.52	6.06	21.18	41.01	4.76	31.50
S	(0-1)	169.50	82.26	3.46	27.98	34.58	2.81	15.25
S	(2)	135.70	69.00	11.76	7.26	62.58	3.87	38.60
S	(3-5)	173.16	83.28	4.83	16.49	57.30	3.77	28.75
S	(4)	247.22	96.84	1.70	45.93	22.36	5.48	31.50

Figura 2 Grafos circulares para los pacientes 7, 8, 9 y 10.

Discusión

Se han encontrado diferencias significativas en los perfiles de glucosa clasificados por las variables día de la semana y franja horaria en cada uno de los pacientes. La clasificación automática ha encontrado similitudes entre perfiles de glucosa de distintas categorías, entendiendo por categoría todos los perfiles correspondientes a una misma franja horaria de un mismo día de la semana (p.ej., los lunes de 00:00 a 04:00 h). Un perfil de glucosa está formado por los valores de glucosa medidos en esa franja horaria.

Las agrupaciones obtenidas con la variable día de la semana son heterogéneas, es decir, están formadas por una, dos, tres y hasta cuatro categorías. Lo mismo sucede para las agrupaciones con respecto a la variable franja horaria. El tamaño de las agrupaciones que más se ha repetido para la variable día de la semana es dos categorías (en todos los pacientes) y la que menos tres categorías (pacientes 6 y 7) y cuatro (pacientes 1 y 9). Por lo que respecta a la variable franja horaria el resultado es similar, aunque se han encontrado un mayor número de similitudes entre perfiles, es decir, más agrupaciones que para la variable día de la semana, y para un mayor número de los pacientes analizados.

Del análisis de las [tablas 4 y 5](#) de manera conjunta se pueden obtener conclusiones de aplicabilidad clínica. En el caso del paciente 10 se puede observar que los sábados están más hiperglucémico que otros días ([tabla 4](#)), y con la información de la [tabla 5](#) para el sábado se pueden tomar

medidas para reducir este comportamiento, como modificar su pauta de insulina o su dieta.

Si analizamos los gráficos circulares de las **figuras 1 y 2** se observa que para los pacientes 3 y 6 los centros de los gráficos están libres de líneas. Esto es debido a que las agrupaciones en estos pacientes están formadas por pocos elementos, habiendo diferencias entre los perfiles de glucosa obtenidos, tanto para los días como para las franjas. En los pacientes 4 y 9 sucede lo contrario. Las agrupaciones están formadas por varios elementos y los centros de los gráficos están ocupados con muchas líneas. En este caso, los perfiles de glucosa obtenidos son parecidos, existiendo un mayor control glucémico en estos pacientes (mayor tiempo en rango).

No obstante, hay gráficos (pacientes 3, 5 y 6) con pocas conexiones debido a que el algoritmo detecta pequeñas diferencias entre los valores de glucosa para los diferentes días que previamente agrupó en un solo conjunto y por ello no aparecen líneas en los gráficos. Es decir, el algoritmo realiza una clasificación con mayor profundidad del árbol (**tabla 2**). Por ejemplo, en el paciente 3 hay un primer nivel que agrupa los lunes y martes, y después agrupa en un segundo nivel las franjas 0-3, 2-5, 1 y 4. Si el gráfico se realizará con respecto a este segundo nivel, se observaría un mayor número de conexiones entre categorías. Por lo tanto, es necesario establecer previamente a la aplicación del algoritmo el número de niveles.

Conclusiones

Las conclusiones obtenidas de este trabajo son:

- Se han observado diferencias significativas y dependencias entre los perfiles de glucosa clasificados en función de las variables día de la semana y franja horaria.
- Las agrupaciones encontradas han sido diferentes para cada paciente, demostrando la necesidad de hacer un estudio individualizado.
- Las **tablas 4 y 5** permiten buscar diferencias significativas para corregir y mejorar hábitos o terapias en los pacientes, y obtener modelos más precisos mediante técnicas de aprendizaje automático e inteligencia artificial.
- Los resultados obtenidos indican que las técnicas aplicadas pueden facilitar la modelización matemática de la glucosa, y podrán utilizarse para la creación de un clasificador individualizado para cada paciente que clasifique los perfiles de glucosa en función de las variables día de la semana y franja horaria. Utilizando este clasificador, se podrán predecir los valores de glucosa del paciente conociendo en qué día se encuentra y en qué franja horaria, obteniendo modelos más precisos.

Financiación

Este trabajo ha sido parcialmente financiado por:

Ayudas de la Fundación Eugenio Rodríguez Pascual 2019: Desarrollo de sistemas adaptativos y bioinspirados para el control glucémico con infusores subcutáneos continuos de insulina y monitores continuos de glucosa.

El proyecto B2017/BMD3773 (GenObIA-CM) y Y2018/NMT-4668 (Micro-Stres - MAP-CM), financiado por la Comunidad de Madrid y cofinanciado con los Fondos Estructurales de la UE.

El proyecto Y2018/NMT-4668 (Micro-Stres - MAP-CM), financiado por la Comunidad de Madrid y cofinanciado con los Fondos Estructurales de la UE.

La subvención del Ministerio de Economía y Competitividad española RTI2018-095180-B-I00.

Autoría

Marta Botella, José Ignacio Hidalgo y Oscar Garnica han diseñado el estudio y han participado en la redacción y la revisión del artículo.

Sergio Contador Pachón ha programado el procesado de los datos y las técnicas de clusterización y ha participado en la redacción del artículo.

Jose Manuel Velasco ha realizado las gráficas de clusterización y ha participado en la redacción del artículo.

Esther Maqueda ha participado en la revisión del artículo.

Aranzazu Aramendi y Remedios Martínez han realizado la toma de datos y la formación de los pacientes.

Conflictos de intereses

Los autores declaran no tener ningún conflicto de intereses.

Agradecimientos

Agradecimientos: Almudena Sánchez Sánchez participó en las discusiones y en el origen del trabajo.

Anexo. Material adicional

Se puede consultar material adicional a este artículo en su versión electrónica disponible en [doi:10.1016/j.endinu.2019.12.011](https://doi.org/10.1016/j.endinu.2019.12.011).

Bibliografía

1. Bellazzi R, Zupan B. Predictive data mining in clinical medicine: Current issues and guidelines. *Int J Med Inform.* 2008;77:81–97.
2. Witten IH, Eibe F, Hall MA. *Data Mining: Practical Machine Learning Tools and Techniques.* Morgan Kaufmann. 2013.
3. Delen D, Walker G, Kadam A. Predicting breast cancer survivability: A comparison of three data mining methods. *Artif Intell Med.* 2005;34:113–27, <http://dx.doi.org/10.1016/j.artmed.2004.07.002>.
4. Chen HY, Chuang CH, Yang YJ, Wu TP. Exploring the risk factors of preterm birth using data mining. *Expert Syst Appl.* 2011;38:5384–7, <http://dx.doi.org/10.1016/j.eswa.2010.10.017>.
5. Chang CD, Wang CC, Jiang BC. Using data mining techniques for multi-diseases prediction modeling of hypertension and hyperlipidemia by common risk factors. *Expert Syst Appl.* 2011;38:5507–13, <http://dx.doi.org/10.1016/j.eswa.2010.10.086>.
6. Meng XH, Huang YX, Rao DP, Zhang Q, Liu Q. Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. *Kaohsiung J Med Sci.* 2013;29:93–9, <http://dx.doi.org/10.1016/j.kjms.2012.08.016>.
7. Breault JL, Goodall CR, Fos PJ. Data mining a diabetic data warehouse. *Artif Intell Med.* 2002;26:37–54, [http://dx.doi.org/10.1016/S0933-3657\(02\)00051-9](http://dx.doi.org/10.1016/S0933-3657(02)00051-9).
8. Ramezankhani A, Pournik O, Shahrabi J, Khalili D, Azizi F, Hadaegh F. Applying decision tree for identification of a low risk population for type 2 diabetes. *Tehran lipid and glucose study. Diabetes Res Clin Pract.* 2014;105:391–8, <http://dx.doi.org/10.1016/j.diabres.2014.07.003>.
9. Kass GV. An exploratory technique for investigating large quantities of categorical data. *J R Stat Soc Ser C Appl Stat.* 1980;29:119–27, <http://dx.doi.org/10.2307/2986296>.
10. Field A. *Discovering Statistics Using IBM SPSS statistics.* Sage; 2013.
11. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize implements and enhances circular visualization in R. *Bioinformatics.* 2014;30:2811–2.
12. The R Development Core Team. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing; 2013. <http://www.R-project.org/>