

# Un análisis de soluciones basadas en Programación Genética para resolver problemas de regresión simbólica de datos de medidores continuos de glucosa

J. Ignacio Hidalgo, José Manuel Velasco, Juan Lanchares, Sergio Contador y Oscar Garnica  
Universidad Complutense de Madrid  
C/. Prof. José García Santesmases 9, 28040, Madrid, España  
Email: hidalgo@dacya.ucm.es

**Resumen**—Diabetes es el nombre de un grupo de enfermedades metabólicas asociadas con la incapacidad del cuerpo para regular los niveles de glucosa en sangre. Una solución ideal para la diabetes sería un páncreas artificial que fuera capaz de mantener un buen control de los niveles de azúcar en la sangre y permitir que el paciente tenga una vida normal, retrasando al mismo tiempo la aparición de complicaciones. Debido a esto, un páncreas artificial necesita realizar predicciones precisas de la evolución de la glucosa utilizando diferentes horizontes de tiempo. En este artículo analizamos diversas técnicas para la predicción de la glucosa. En concreto, comparamos la programación genética con selección estricta, una aproximación más tradicional de programación genética y *ARIMA* con ventana de 24 y 48 horas. El análisis de los resultados se realiza utilizando las métricas Clarke Error Grid y Parkes Error Grid.

## I. INTRODUCCIÓN

El páncreas es una glándula del sistema endocrino que produce insulina y glucagón. Estas dos hormonas mantienen los niveles de glucosa en sangre dentro de los rangos óptimos y regulan el metabolismo de los carbohidratos. Existe un grupo de enfermedades metabólicas asociadas con la incapacidad del cuerpo para regular el metabolismo que pueden producir niveles elevados de glucosa en sangre durante un periodo de tiempo prolongado. Estas enfermedades se llaman Diabetes [1]. En 2015, se estimó que más de 415 millones de personas en todo el mundo la padecían [2]. La Diabetes Mellitus Tipo 1 (DMT1), que suele manifestarse en edades tempranas, representa el 10 % de este número. La DMT1 es un trastorno autoinmune crónico por el que el sistema inmune ataca las células secretoras de insulina del páncreas. El resultado es que las células no asimilan el azúcar y, como consecuencia, hay un aumento en los niveles de glucosa en sangre que se conoce como hiperglucemia (niveles de glucosa en sangre superiores a 140 *mg/dl*). Si esta situación se extiende durante un largo periodo de tiempo, el

paciente puede desarrollar complicaciones graves a largo plazo, incluyendo enfermedades cardíacas, ceguera, insuficiencia renal o úlceras en los pies [3, 4].

La DMT1 solo puede tratarse con insulina sintética inyectada en el torrente sanguíneo. Sin embargo, esta no es una tarea fácil. Una dosis excesiva de insulina puede producir hipoglucemia (definida como un valor de glucosa en sangre inferior a 70 *mg/dl*). Si la hipoglucemia es muy grave (glucosa inferior a 50 *mg/dl*), puede llevar a la pérdida del conocimiento y a un coma diabético. En la Figura 1 podemos ver la evolución de los niveles de glucosa en sangre en un paciente diabético. Se observa como los niveles de glucosa se encuentran en la zona de hiperglucemia durante un periodo largo de tiempo y, los episodios de hiperglucemia grave o hipoglucemia son frecuentes.

Algunos pacientes con complicaciones graves han sido operados mediante un trasplante de páncreas, pero esta opción puede producir varios problemas nuevos. El paciente trasplantado debe tomar medicamentos inmunosupresores para evitar que su sistema inmune ataque su nuevo páncreas, y este tratamiento puede coadyuvar al desarrollo de infecciones bacterianas o virales y cáncer. Una solución ideal para la DMT1 sería un Páncreas Artificial (PA) [5] capaz de mantener el control de los niveles de azúcar en la sangre y permitir que el paciente tenga una vida normal, evitando al mismo tiempo, o al menos retrasando, la aparición de complicaciones. En la Figura 2 se muestra la estructura básica de un PA. Uno de los componentes es un sistema continuo de monitorización de la glucosa que consiste en un biosensor [6, 7] (generalmente un sensor que mide el nivel de glucosa intersticial [8]) y un transmisor inalámbrico. El otro componente es un dispositivo integrado con varias funcionalidades. Por un lado, recibe las muestras de glucosa y las almacena. Por otro lado, controla una bomba de insulina; un dispositivo que inyecta bolos de insulina en el torrente



sanguíneo. Para determinar la cantidad de insulina, el dispositivo tiene en cuenta varios factores, entre los que se encuentran el tipo y las características de la insulina. Principalmente, hay cuatro tipos de insulinas (insulina de acción rápida, de acción corta, de acción intermedia y de acción lenta) que se clasifican dependiendo del tiempo de retraso antes de que comiencen a actuar (inicio), el momento en que la insulina alcanza su máximo efecto (pico) y la demora final hasta que cesa su efecto (duración). La insulina de acción rápida, tiene un inicio entre 5-15 minutos, un pico entre 30-60 minutos y una duración entre 3-5 horas. La insulina de acción corta o regular, tiene un inicio de 30 minutos, un pico entre 2-3 horas y una duración entre 3-6 horas. La insulina de acción intermedia, tiene un inicio entre 2-4 horas, un pico entre 4-12 horas y una duración entre 12-18 horas. La insulina de acción lenta tiene un inicio entre 1-2 horas, un pico entre 6-8 horas y una duración entre 6-23 horas, dependiendo de la dosis suministrada. Debido a esto, surge la necesidad de realizar predicciones precisas de la evolución de la glucosa utilizando diferentes horizontes de tiempo.

Esta investigación está motivada por el desafío de predecir con precisión los niveles futuros de glucosa para que un sistema automático pueda decidir cuándo y cuánta insulina inyectar a fin de mantener el nivel de azúcar en la sangre en un rango saludable. Al mismo tiempo, es muy importante para un sistema de control de la glucosa en sangre, evitar las predicciones que puedan desencadenar tratamientos innecesarios o, incluso peor, tratamientos que son perjudiciales para el paciente. Aunque hay algunas aproximaciones clásicas, todavía hay mucho trabajo por hacer para las predicciones dentro de un horizonte de más de 60 minutos.

Una solución que ha demostrado ser adecuada para predecir los niveles de glucosa es la utilización de algoritmos basados en la Evolución Gramatical (EG) [9]. Sin embargo, uno de los principales obstáculos que encontraron los investigadores para entrenar a los modelos de EG es la escasez de cantidades significativas de datos. Como en muchos otros campos de la medicina, la recopilación de datos de pacientes diabéticos reales es muy compleja. Los modelos de EG entrenados con un pequeño conjunto de datos, por lo general, sufren de sobreajuste y tienen un poder de predicción limitado. Para evitar esta situación, en [10] se ha propuesto aumentar los registros de datos de glucosa con datos sintéticos con buenos resultados. También se han realizado otras comparativas entre técnicas relacionadas con la EG, como la Programación Genética (PG) con selección estricta. Sin embargo, esta última tiene un alto tiempo de ejecución.

En este artículo investigamos otras opciones; en concreto queremos comparar la programación genética con selección estricta, una aproximación más tradicio-

nal de programación genética y el método *ARIMA* con ventana de 24 y 48 horas. Además, queremos ver si los resultados tienen una interpretación correcta comparándolos utilizando las métricas de Clarke Error Grid y Parkes Error Grid.

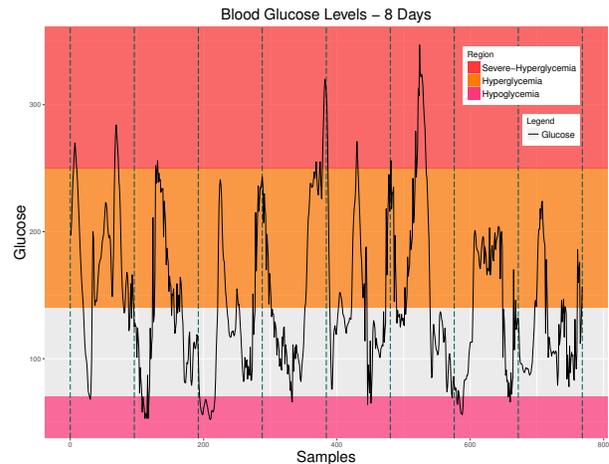


Figura 1: Valores de Glucosa Reales - 8 Días

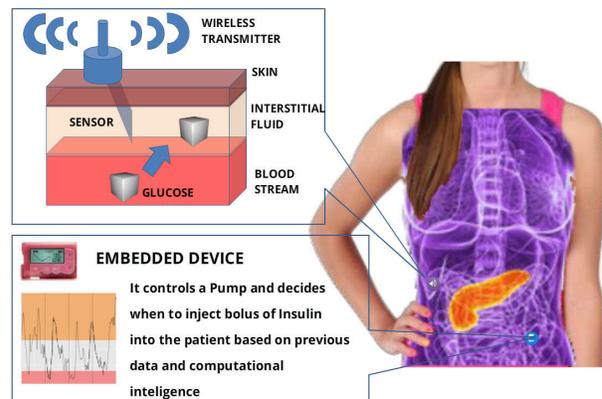


Figura 2: Páncreas Artificial

El resto del artículo está organizado de la siguiente manera. La sección II describe los antecedentes teóricos de las técnicas empleadas y explica en detalle nuestra metodología. El trabajo experimental y los resultados se muestran en la sección III.

## II. MÉTODOS

Nuestra propuesta es aplicar varias técnicas de aprendizaje con el fin de obtener modelos personalizados para la predicción del nivel de glucosa de cada paciente siguiendo un enfoque en el que los modelos se extraen de los datos. Levaremos a cabo una primera fase de entrenamiento sobre un conjunto de datos de pacientes que producirán un número de modelos que se evaluarán en una fase de prueba con datos de

los mismos pacientes. Consideramos una ventana de predicción de 2 horas. Los datos están disponibles en forma de series temporales con observaciones a intervalos de cinco minutos. Se adquieren las series de glucosa, ingestas de carbohidratos e insulinas. Hemos transformado este problema de predicción de series de tiempo en múltiples problemas de regresión independientes, donde el objetivo es encontrar modelos independientes para la predicción de la glucosa en sangre a 2 horas. De esta manera podemos aplicar cualquier técnica de regresión, incluida la regresión simbólica usando PG.

Un inconveniente de esta transformación es que los modelos no nos permiten hacer predicciones sobre ningún otro punto de la ventana de predicción. Sin embargo, esto no es un problema desde un punto de vista práctico, ya que para los pacientes diabéticos es suficiente con predecir los valores de cuatro puntos distintos en la ventana de predicción. La ventaja es que podemos utilizar los datos de los sistemas de Monitorización Continua de la Glucosa (MCG) directamente para hacer recomendaciones de insulina basadas en distintos horizontes temporales, utilizando las ecuaciones dadas por los modelos obtenidos.

Nuestro objetivo es obtener el siguiente modelo:

$$\hat{G}_{t+120} = f_{t+120}((G_{t+i}, i \in (-240 \dots 0)), (I_{t+i}, C_{t+i}, i \in (-240 \dots +120))) \quad (1)$$

donde  $t$  es el tiempo en minutos y estamos haciendo predicciones hasta 120 minutos.  $G$  es la serie temporal de valores medidos de concentración de glucosa en sangre,  $I$  es la serie temporal de insulina (bomba y bolo), y  $C$  es la cantidad de carbohidratos ingeridos según lo estimado por los pacientes. En cada instante temporal, los datos de  $t$  de hasta 4 horas antes están disponibles para la predicción. Para generar los modelos de predicción  $f_{t+120}$  usamos *ARIMA* y *PG* basada en árboles para la evolución de los modelos de regresión simbólica. Adicionalmente, consideramos dos predictores de referencia: *Last* y *Avg*. El primero considera la predicción como el último valor de glucosa conocido. El segundo considera la predicción como la glucosa media de los últimos valores conocidos en un intervalo de 2 horas.

### II-A. Preprocesado de datos

Los pacientes diabéticos pueden inyectarse insulina de dos formas: mediante múltiples dosis de insulina inyectada manualmente o mediante una bomba (infusor continuo) de insulina. Existen tres categorías de insulina:

- Insulina Basal
- Insulina Preprandial
- Insulina Correctora

En este artículo vamos a utilizar el modelo de absorción de Berger y Rodband [11] que describe la farmacocinética para distintos tipos de insulina.

A partir de las series temporales de glucosa, insulina y carbohidratos, definimos un conjunto de características que describimos a continuación.

Agregamos los valores de las series en intervalos de tiempo que se calculan utilizando la siguiente ecuación:

$$media(X, t, rango) = \frac{\sum_{i \in rango} (X_{i+1})}{|rango|} \quad (2)$$

El objetivo es reducir el número de características utilizadas para generar los modelos. Para cada instante de tiempo  $t$  definimos el conjunto de características  $F(t)$  que describen los valores históricos  $F_{his}$  de las series temporales  $G$ ,  $I$  y  $C$ , así como los valores futuros  $F_{fut}$  de  $I$  y  $C$ :

$$F(t) = F_{his}(G, t) \cup F_{his}(I, t) \cup F_{his}(C, t) \cup F_{fut}(I, t) \cup F_{fut}(C, t) \cup \{G, I, C\} \quad (3)$$

$$F_{his}(X, t) = \{media(X, t, [-15, 0]), media(X, t, [-30, -15]), media(X, t, [-45, -30]), media(X, t, [-60, -45]), media(X, t, [-90, -60]), media(X, t, [-120, -90])\} \quad (4)$$

$$F_{fut}(X, t) = \{media(X, t, [0, 15]), media(X, t, [15, 30]), media(X, t, [30, 45]), media(X, t, [45, 60]), media(X, t, [60, 75]), media(X, t, [75, 90]), media(X, t, [90, 105]), media(X, t, [105, 120])\} \quad (5)$$

### II-B. Programación Genética

Para cada paciente se crean cuatro modelos diferentes basados en programación genética (*PG*, *PG2*, *PG2<sub>com</sub>* y *PGOS*) para predecir el nivel de glucosa a 2 horas. Se utiliza validación cruzada de 10 iteraciones (*10-fold Cross-Validation*).

Cada modelo se generó usando el algoritmo de regresión simbólica basada en programación genética disponible en el software de código abierto HeuristicLab versión 3.3.15 [12]. Para el conjunto de terminales utilizamos las series temporales de glucosa, insulina, y carbohidratos para los modelos *PG*, *PG2* y *PGOS*; para *PG2<sub>com</sub>* utilizamos además las características precalculadas en el apartado *Preprocesado*



de datos. En todos los modelos utilizamos también constantes. Cada valor se acompaña con un peso de valor real que se inicializa aleatoriamente usando una distribución Normal  $N(\mu = 1, \sigma = 1)$ , y se muta aleatoriamente añadiendo un valor muestreado de  $N(\mu = 0, \sigma = 0.05)$ . Las constantes con valores reales son inicializadas al azar usando una distribución uniforme  $U(-20.0, 20.0)$ , y la mutación de constantes agrega un valor muestreado de  $N(\mu = 0, \sigma = 1.0)$ .

El conjunto de funciones utilizado es:  $\{+, -, *, /, \log(x), \exp(x)\}$ , donde se utilizan las variantes protegidas de las funciones división y logaritmo [13]. Los modelos en la población inicial se han generado usando *Probabistic Tree Creator* (PTC) [14] utilizando diferentes tamaños para la población, y para los árboles un límite para la profundidad máxima y un máximo número de nodos permitidos. Las mismas restricciones de profundidad y tamaño se han aplicado en las operaciones de cruce y mutación. Para el cruce, utilizamos un operador de cruce de subárbol. Para la mutación utilizamos una variedad de operadores que reemplaza un subconjunto completo del árbol con un árbol inicializado al azar, o mutan todos los nodos del árbol, o mutan solo un nodo del árbol seleccionado de forma aleatoria. El operador de mutación seleccionado al azar se ejecuta después de cada operación de cruce con una tasa de mutación diferente. Utilizamos un número máximo de generaciones y selección de aptitud proporcional de padres. No hemos sintonizado los parámetros de la PG específicamente para esta tarea de modelado, pero utilizamos configuraciones estándar robustas.

Para *PG* y *PGOS*, utilizamos un tamaño de población de 100 individuos y una profundidad máxima de árbol de 8. El número máximo de generaciones es 25 con una tasa de mutación del 25%. Para *PG2* y *PG2<sub>com</sub>*, utilizamos un tamaño de población de 300 individuos y una profundidad máxima de árbol de 11. El número máximo de generaciones es 100 con una tasa de mutación del 15%. En todos los casos utilizamos un máximo número de nodos de árbol permitidos de 100.

En el caso de *PGOS* utilizamos selección estricta de descendientes [15]. Cuando se usa selección estricta, los descendientes solo se aceptan si son mejores que ambos padres, en caso contrario son descartados. Esto tiene el efecto de que múltiples intentos repetidos de selección de padres, el cruce, la mutación opcional y la evaluación de la aptitud, son necesarios hasta que uno de los descendientes se agrega a la población de la próxima generación [16]. El umbral de presión de selección para detener el algoritmo se estableció en 100.

La función objetivo utilizada en todos los casos se calcula con el coeficiente de determinación  $R^2$

de Pearson entre los valores reales de glucosa en sangre y los valores obtenidos con el modelo [17]. Los modelos de predicción finalmente seleccionados se escalan linealmente para minimizar la suma de los errores al cuadrado entre los valores reales y los resultados del modelo.

### II-C. ARIMA

El modelo autorregresivo integrado de media móvil o *ARIMA* (Auto Regressive Integrated Moving Average) es un modelo estadístico que utiliza variaciones y regresiones de los datos, en particular series temporales, con el fin de encontrar los parámetros del modelo más adecuados para predecir los valores futuros. Se suele expresar como  $ARIMA(p, d, q)$ , donde los parámetros  $p$ ,  $d$  y  $q$ , indican el orden de las distintas componentes del modelo ( $p$  es la componente autorregresiva,  $d$  la integrada y  $q$  la media móvil). Si se considera el efecto de la estacionalidad de los datos, el modelo toma la forma  $ARIMA(p, q, d)(P, Q, D)_m$ , y está dado por la siguiente ecuación:

$$\Phi(B^m)\phi(B)(1 - B^m)^D(1 - B)^d y_t = c + \Theta(B^m)\theta(B)\varepsilon_t \quad (6)$$

donde  $\varepsilon_t$  es ruido blanco con media 0 y varianza  $\sigma^2$ ,  $B$  es el operador de retroceso, y  $\phi(z)$ ,  $\theta(z)$ ,  $\Phi(z)y\Theta(z)$ , son polinomios de orden  $p$ ,  $q$ ,  $P$  y  $Q$ , respectivamente.

Uno de los principales problemas en el uso de los modelos *ARIMA* es que el proceso de selección del orden de las componentes generalmente se considera subjetivo y difícil de aplicar. Por eso, hemos utilizado el modelo estacional *auto.arima* [18] con el software libre de análisis estadístico R versión 3.3.1 [19], que busca de forma automática las componentes del modelo seleccionando el orden apropiado.

### II-D. Clarke-Parkes Error Grid

La precisión de las predicciones se analizan con Clarke Error Grid y Parkes Error Grid. Clarke Error Grid [20] se desarrolló para calcular la precisión clínica de los sistemas de MCG en sangre para todo el rango de valores de glucosa, mediante las diferencias entre los valores medidos por el sistema de medición y los valores de referencia. Análogamente, se puede utilizar para calcular las diferencias entre los valores estimados en una predicción y los valores reales o de referencia. Los valores se representan en un gráfico con coordenadas cartesianas donde el eje X representa los valores reales y el eje Y los valores de la predicción, siendo  $Y = X$  la predicción ideal, donde el valor real coincide con el valor de la predicción. La particularidad de esta representación es que el gráfico se divide en cinco zonas dependiendo del grado de precisión

de las estimaciones de la glucosa. A continuación describimos cada una de las zonas:

- Zonas A y B: los valores en la zona A representan los valores de glucosa que se desvían de los valores de referencia menos del 20 %, o aquellos que están en el rango de hipoglucemia ( $<70$  *mg/dl*), cuando la referencia es también  $<70$  *mg/dl*. Los valores en la zona B se desvían de los valores de referencia en más del 20 %, aunque el tratamiento clínico será correcto con una alta probabilidad. Los valores que se encuentran dentro de las zonas A y B son clínicamente exactos y/o aceptables y, por lo tanto, el tratamiento será clínicamente correcto. Tratamos estas dos zonas de forma combinada en el análisis de los resultados de nuestro experimento.
- Zona C: los valores en esta zona pueden ser peligrosos en algunas situaciones. El objetivo es minimizar las predicciones incluidas en esta zona.
- Zonas D-E: los valores incluidos en estas zonas son potencialmente peligrosos, ya que la predicción está lejos de ser aceptable y el tratamiento indicado será diferente del correcto, y contrario en la zona E. De nuevo, el objetivo es minimizar las predicciones incluidas en esta zona.

Parkes Error Grid [21] se publicó en 2000 como una alternativa al uso de Clarke Error Grid. En él se redefinen las zonas basándose en las zonas de Clarke y en los límites establecidos por 100 médicos expertos en diabetes en una encuesta realizada en la conferencia *American Diabetes Association Meetings* de junio de 1994. Estas definiciones son más flexibles, y permiten incorporar el conocimiento de expertos para asignar las zonas basándose en su experiencia.

### III. RESULTADOS EXPERIMENTALES

Se realiza un estudio retrospectivo de seis pacientes con DMT1. Los datasets utilizados proceden de la Universidad de Ohio [22]. En concreto, hay dos hombres (563 y 570) y cuatro mujeres (559, 575, 588 y 591). Los datos son recolectados durante un periodo de ocho semanas mediante MCG intersticial registrada con el sistema Medtronic 530G (©Medtronic). Todos los pacientes siguen una terapia con bomba de insulina. El número de observables por paciente que tenemos se muestran en la Tabla I. Las características de cada paciente se muestran en la Tabla II. Para cada paciente se calcula 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 rango  $[70, 180]$  *mg/dl*.

Tabla I: Número de observables disponibles para cada paciente.

Paciente	Observables
559	10149
563	11853
570	10580
575	11697
588	12395
591	10601

Tabla II: Características de cada paciente para los niveles de glucosa.

Paciente	Media [ <i>mg/dl</i> ]	Std [ <i>mg/dl</i> ]	$T_G < 70$ [ % ]	$T_G > 250$ [ % ]	$T_G [70, 180]$ [ % ]
559	167.22	70.50	3.95	13.11	55.54
563	145.73	49.43	2.10	2.35	74.86
570	187.22	62.37	1.80	17.16	43.75
575	139.76	58.06	8.04	4.57	70.21
588	165.09	50.60	0.94	4.96	63.47
591	156.83	57.38	3.02	7.03	64.62

Tabla III: Tiempos medios de ejecución de los modelos junto con su desviación típica.

Modelo	Tiempo [ <i>min</i> ]
<i>PG</i>	$1.54 \pm 0.14$
<i>ARIMA</i> 24h	$15.18 \pm 2.24$
<i>PG</i> <sub>com</sub> <sup>2</sup>	$19.76 \pm 1.62$
<i>PG</i> <sup>2</sup>	$20.11 \pm 1.73$
<i>ARIMA</i> 48h	$24.40 \pm 4.42$
<i>PGOS</i>	$1453.43 \pm 140.60$

Tabla IV: Predicciones (en porcentaje) a 2h obtenidas para las distintas zonas con Clarke-Parkes Error Grid con *ARIMA* con entrenamiento de 24h y 48h.

Zona	Paciente					
	559	563	570	575	588	591
Clarke con Entrenamiento de 24h						
A	39.94	45.65	<b>58.52</b>	40.90	52.29	40.22
B	44.87	47.83	34.41	45.34	42.96	<b>49.16</b>
C	8.06	3.31	3.91	3.95	1.72	3.76
D	5.59	2.68	2.74	7.47	2.48	5.21
E	1.54	0.53	0.43	2.34	0.55	1.65
Parkes con Entrenamiento de 24h						
A	43.13	46.47	<b>61.99</b>	43.18	52.78	42.75
B	42.43	45.55	32.55	42.01	41.66	<b>46.91</b>
C	12.25	7.47	4.96	12.69	5.08	8.78
D	2.17	0.51	0.50	2.11	0.48	1.56
E	0.01	0.00	0.00	0.01	0.00	0.00
Clarke con Entrenamiento de 48h						
A	38.01	46.81	<b>61.30</b>	41.93	52.75	41.95
B	47.67	47.42	33.61	44.33	42.92	<b>48.43</b>
C	6.19	2.23	2.33	3.54	1.27	2.80
D	6.95	2.98	2.59	8.24	2.68	5.29
E	1.17	0.56	0.17	1.95	0.38	1.54
Parkes con Entrenamiento de 48h						
A	40.23	47.32	<b>63.93</b>	43.99	53.03	43.67
B	<b>45.67</b>	45.59	31.00	41.23	41.44	45.66
C	12.32	6.53	4.79	12.79	5.00	8.85
D	1.79	0.57	0.28	1.97	0.53	1.79
E	0.00	0.00	0.00	0.01	0.00	0.04



Las Tablas IV, V y VI (V y VI en ANEXO I) recogen los resultados para *ARIMA* y PG para las distintas zonas en términos de los errores Clarke y Parkes.

Podemos observar que en algunos pacientes en los que el control es muy bueno, las técnicas funcionan de una manera similar. No obstante, el mejor modelo generado que más porcentajes de valores tiene en las zonas A y B es  $PG2_{com}$ , donde el porcentaje medio en las zonas ha sido  $91.95 \pm 4.27$  [%]. El modelo generado con PG que más porcentajes de valores tiene en las zonas A y B es  $PG2_{com}$  (aparece en 8 de los 12 casos). El segundo modelo es *PGOS* (aparece en 4 de los 12 casos). Comparando *PG* con *PGOS*, el modelo mejora cuando se utiliza selección de descendencia. Comparando  $PG2$  con  $PG2_{com}$ , el modelo mejora cuando se utilizan más variables (en este caso las características) para entrenarlo. El modelo generado con *ARIMA* que más porcentajes de valores tiene en las zonas A y B es *ARIMA* con entrenamiento de 48h (aparece en 11 de los 12 casos). El modelo *ARIMA* mejora cuando se utiliza un periodo de tiempo de entrenamiento más largo.

El modelo *PGOS* no obtiene mejores resultados en la predicción a 2 horas. Por lo tanto, el tiempo empleado en la obtención del modelo (Tabla III) puede ser fundamental para la calidad de las predicciones y para la seguridad del paciente. También constatamos que la métrica Parkes es más precisa que Clarke y por consiguiente, más conveniente.

#### AGRADECIMIENTOS

Este trabajo ha sido financiado por el Ministerio de Economía y Competitividad de España a través de los proyectos con referencias TIN2014-54806-R y TIN2015-65460-C2.

#### REFERENCIAS

- [1] Leonid Poretsky Ed. *Principles of Diabetes Mellitus - 2017*. Springer International Publishing, 2017.
- [2] James Stephen Krinsley and Richard L. Jones. Cost analysis of intensive glycemic control in critically ill adult patients. *Chest*, 129(3):644 – 650, 2006.
- [3] American Diabetes Association. Intensive diabetes treatment and cardiovascular outcomes in type 1 diabetes: The dcct/edic study 30-year follow-up. *Diabetes Care*, 2016.
- [4] Ya-Xiong Tao (Eds.). *Glucose Homeostasis and the Pathogenesis of Diabetes Mellitus*. Progress in Molecular Biology and Translational Science 121. Academic Press, 1 edition, 2014.
- [5] Giuseppe De Nicolao, Lalo Magni, Chiara Dalla Man, and Claudio Cobelli. Modeling and control of diabetes: Towards the artificial pancreas. *{IFAC} Proceedings Volumes*, 44(1):7092 – 7101, 2011. 18th {IFAC} World Congress.
- [6] Emma P. Córcoles and Martyn G. Boutelle. *Biosensors and Invasive Monitoring in Clinical Applications*. SpringerBriefs in Applied Sciences and Technology, 2013.
- [7] Eun-Hyung Yoo and Soo-Youn Lee. Glucose biosensors: An overview of use in clinical practice. *Sensors*, 10(5):4558–4576, 2010.
- [8] Alexander Hildenbrand Hansen, Anne Katrine Duun-Henriksen, Rune Juhl, Signe Schmidt, Kirsten Nørgaard, John Bagterp Jørgensen, and Henrik Madsen. Predicting plasma glucose from interstitial glucose observations using bayesian methods. *Journal of Diabetes Science and Technology*, 8(2):321–330, 2014.
- [9] J. Ignacio Hidalgo, Esther Maqueda, José L. Risco-Martín, Alfredo Cuesta-Infante, J. Manuel Colmenar, and Javier Nobel. glucodel: A monitoring and modeling system for chronic diseases applied to diabetes. *Journal of Biomedical Informatics*, 48:183 – 192, 2014.
- [10] Jose Manuel Velasco, Oscar Garnica, Juan Lanchares, Marta Botella, and J. Ignacio Hidalgo. Combining data augmentation, edas and grammatical evolution for blood glucose forecasting. *Memetic Computing*, Jun 2018.
- [11] Markus Berger and David Rodbard. Computer simulation of plasma insulin and glucose dynamics after subcutaneous insulin injection. *Diabetes care*, 12(10):725–736, 1989.
- [12] S. Wagner and M. Affenzeller. Heuristiclab: A generic and extensible optimization environment. pages 538–541, 2005.
- [13] John R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, USA, 1992.
- [14] Sean Luke. Two fast tree-creation algorithms for genetic programming. *IEEE Trans. Evolutionary Computation*, 4:274–283, 2000.
- [15] M. Affenzeller and S. Wagner. Offspring selection: A new self-adaptive selection scheme for genetic algorithms. pages 218–221, 2005.
- [16] Michael Affenzeller, Stephan Winkler, Stefan Wagner, and Andreas Beham. Genetic algorithms and genetic programming: Modern concepts and practical applications. 01 2009.
- [17] Maarten Keijzer. Improving symbolic regression with interval arithmetic and linear scaling. pages 70–82, 2003.
- [18] Rob Hyndman and Yeasmin Khandakar. Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software, Articles*, 27(3):1–22, 2008.
- [19] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.
- [20] W.L. Clarke, D. Cox, L.A. Gonder Frederick, W. Carter, and S.L. Pohl. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. *Diabetes Care*, 10(5):622–628, September 1987.
- [21] Pardo S Ginsberg BH. Parkes JL, Slatin SL. A new consensus error grid to evaluate the clinical significance of inaccuracies in the measurement of blood glucose. *Diabetes Care*, 23(8):1143–1148., 2000.
- [22] Cindy Marling and Razvan Bunescu. The ohio1dm dataset for blood glucose level prediction - draft.

## ANEXO I

Tabla V: Predicciones (en porcentaje) obtenidas para las distintas zonas con Clarke Error Grid para las medias y desviaciones típicas de los valores obtenidos con validación cruzada de cada paciente para los distintos modelos con horizonte de tiempo a 2 horas.

Modelo	t+120 train				t+120 test			
	A+B	C	D	E	A+B	C	D	E
<b>Paciente 559</b>								
<i>Avg</i>	84.4 ± 0.6	1.1 ± 0.2	0.1 ± 0.4	1.5 ± 0.2	81.9 ± 7.2	5.2 ± 9.2	11.1 ± 6.9	1.8 ± 2.2
<i>Last</i>	84.8 ± 0.8	7.3 ± 0.4	6.6 ± 0.3	1.3 ± 0.1	84.8 ± 6.9	7.3 ± 3.7	6.6 ± 2.5	1.3 ± 1.0
<i>PG</i>	88.1 ± 1.3	1.2 ± 0.4	10.2 ± 2.9	0.5 ± 0.2	88.3 ± 5.6	1.2 ± 0.9	10.1 ± 3.5	0.5 ± 0.5
<i>PG2</i>	89.3 ± 0.6	1.3 ± 0.1	8.8 ± 0.5	0.6 ± 0.1	89.2 ± 4.4	1.4 ± 0.8	8.8 ± 2.8	0.5 ± 0.5
<i>PG2<sub>com</sub></i>	<b>90.2 ± 0.7</b>	1.3 ± 0.2	8.1 ± 0.4	0.4 ± 0.0	<b>90.0 ± 4.2</b>	1.5 ± 0.9	8.1 ± 2.8	0.4 ± 0.4
<i>PGOS</i>	89.6 ± 0.5	1.4 ± 0.2	8.3 ± 0.4	0.6 ± 0.1	89.5 ± 4.9	1.6 ± 0.7	8.3 ± 2.7	0.6 ± 0.6
<b>Paciente 563</b>								
<i>Avg</i>	94.4 ± 0.9	0.0 ± 0.0	5.2 ± 0.3	0.3 ± 0.1	94.3 ± 7.9	0.4 ± 0.9	4.8 ± 2.6	0.5 ± 1.1
<i>Last</i>	94.4 ± 0.3	2.1 ± 0.1	2.8 ± 0.2	0.6 ± 0.1	94.4 ± 2.6	2.1 ± 1.0	2.8 ± 2.1	0.6 ± 0.6
<i>PG</i>	94.9 ± 2.5	0.2 ± 0.1	4.7 ± 0.7	0.2 ± 0.1	94.8 ± 6.0	0.2 ± 0.3	4.7 ± 2.9	0.3 ± 0.4
<i>PG2</i>	94.8 ± 2.9	0.1 ± 0.1	5.0 ± 0.8	0.1 ± 0.1	94.2 ± 7.1	0.1 ± 0.3	5.5 ± 4.0	0.1 ± 0.3
<i>PG2<sub>com</sub></i>	<b>95.5 ± 0.7</b>	0.3 ± 0.1	4.1 ± 0.3	0.2 ± 0.1	<b>95.2 ± 6.7</b>	0.3 ± 0.3	4.4 ± 2.8	0.2 ± 0.3
<i>PGOS</i>	95.4 ± 10.8	0.1 ± 0.0	4.3 ± 1.1	0.1 ± 0.1	94.8 ± 13.2	0.2 ± 0.3	4.8 ± 3.0	0.2 ± 0.4
<b>Paciente 570</b>								
<i>Avg</i>	90.9 ± 1.4	1.6 ± 0.2	6.8 ± 0.7	0.7 ± 0.1	89.7 ± 10.0	2.9 ± 4.3	7.1 ± 7.5	0.3 ± 0.7
<i>Last</i>	94.6 ± 0.9	2.2 ± 0.2	2.9 ± 0.3	0.3 ± 0.0	94.6 ± 8.3	2.2 ± 1.5	2.9 ± 2.5	0.3 ± 0.4
<i>PG</i>	95.8 ± 2.4	1.0 ± 0.2	2.9 ± 0.7	0.3 ± 0.4	95.7 ± 10.7	0.9 ± 0.9	3.1 ± 3.1	0.2 ± 0.4
<i>PG2</i>	95.8 ± 1.1	0.9 ± 0.1	3.0 ± 0.3	0.2 ± 0.0	95.9 ± 10.4	1.0 ± 0.9	3.0 ± 3.0	0.2 ± 0.4
<i>PG2<sub>com</sub></i>	<b>96.2 ± 2.3</b>	0.8 ± 0.1	2.6 ± 0.6	0.4 ± 0.4	<b>96.0 ± 10.5</b>	0.9 ± 0.7	2.8 ± 2.7	0.3 ± 0.4
<i>PGOS</i>	95.9 ± 1.0	0.9 ± 0.1	3.0 ± 0.3	0.2 ± 0.0	95.7 ± 9.9	0.9 ± 0.9	3.2 ± 3.2	0.2 ± 0.4
<b>Paciente 575</b>								
<i>Avg</i>	85.6 ± 0.6	0.1 ± 0.1	13.9 ± 0.5	0.4 ± 0.2	84.9 ± 4.2	0.8 ± 2.2	13.9 ± 4.7	0.4 ± 1.0
<i>Last</i>	86.0 ± 0.6	3.4 ± 0.3	7.5 ± 0.4	3.0 ± 0.2	86.0 ± 5.6	3.4 ± 2.3	7.5 ± 3.5	3.0 ± 2.0
<i>PG</i>	86.9 ± 0.5	0.3 ± 0.1	12.4 ± 0.5	0.3 ± 0.1	86.9 ± 7.0	0.3 ± 0.5	12.5 ± 4.1	0.3 ± 0.4
<i>PG2</i>	87.2 ± 0.7	0.3 ± 0.1	12.2 ± 0.6	0.3 ± 0.1	87.0 ± 7.0	0.3 ± 0.5	12.3 ± 4.1	0.4 ± 0.4
<i>PG2<sub>com</sub></i>	<b>87.5 ± 0.9</b>	0.3 ± 0.1	11.9 ± 0.6	0.3 ± 0.1	<b>87.5 ± 8.9</b>	0.4 ± 0.5	11.7 ± 3.8	0.4 ± 0.5
<i>PGOS</i>	<b>87.5 ± 0.7</b>	0.1 ± 0.1	12.0 ± 0.6	0.4 ± 0.1	87.3 ± 6.5	0.1 ± 0.2	12.2 ± 3.9	0.4 ± 0.3
<b>Paciente 588</b>								
<i>Avg</i>	94.4 ± 0.7	0.2 ± 0.0	5.4 ± 0.5	0.0 ± 0.0	94.4 ± 6.6	0.3 ± 0.5	5.2 ± 4.3	0.0 ± 0.1
<i>Last</i>	95.8 ± 0.6	1.4 ± 0.1	2.3 ± 0.2	0.5 ± 0.1	95.8 ± 5.9	1.4 ± 1.0	2.3 ± 1.6	0.5 ± 0.8
<i>PG</i>	95.6 ± 2.9	0.2 ± 0.1	4.0 ± 1.5	0.1 ± 0.1	95.8 ± 8.0	0.2 ± 0.3	3.9 ± 2.4	0.1 ± 0.3
<i>PG2</i>	96.1 ± 0.8	0.2 ± 0.1	3.5 ± 0.3	0.1 ± 0.0	<b>96.1 ± 7.4</b>	0.2 ± 0.3	3.5 ± 2.1	0.1 ± 0.3
<i>PG2<sub>com</sub></i>	<b>96.3 ± 0.9</b>	0.2 ± 0.1	3.4 ± 0.3	0.1 ± 0.0	<b>96.1 ± 7.6</b>	0.3 ± 0.3	3.5 ± 2.2	0.1 ± 0.2
<i>PGOS</i>	95.3 ± 3.6	0.2 ± 0.1	4.4 ± 1.9	0.1 ± 0.1	95.4 ± 8.6	0.3 ± 0.4	4.3 ± 4.0	0.0 ± 0.1
<b>Paciente 591</b>								
<i>Avg</i>	88.6 ± 1.1	0.2 ± 0.1	10.6 ± 0.7	0.6 ± 0.1	88.0 ± 6.6	0.3 ± 0.4	11.3 ± 6.1	0.5 ± 0.7
<i>Last</i>	90.4 ± 0.2	2.8 ± 0.2	5.4 ± 0.2	1.5 ± 0.1	90.4 ± 2.5	2.8 ± 2.0	5.4 ± 1.7	1.5 ± 0.9
<i>PG</i>	90.7 ± 0.7	0.4 ± 0.1	8.6 ± 0.9	0.3 ± 0.1	90.3 ± 6.1	0.4 ± 0.4	9.0 ± 3.8	0.3 ± 0.5
<i>PG2</i>	91.1 ± 0.5	0.4 ± 0.1	8.1 ± 0.3	0.4 ± 0.1	90.9 ± 5.9	0.5 ± 0.5	8.2 ± 2.9	0.4 ± 0.5
<i>PG2<sub>com</sub></i>	91.2 ± 1.3	0.5 ± 0.1	7.9 ± 1.0	0.3 ± 0.1	<b>91.4 ± 5.0</b>	0.5 ± 0.6	7.8 ± 2.7	0.4 ± 0.5
<i>PGOS</i>	<b>91.5 ± 0.6</b>	0.5 ± 0.1	7.6 ± 0.3	0.5 ± 0.1	<b>91.4 ± 5.9</b>	0.4 ± 0.4	7.7 ± 2.9	0.5 ± 0.5



Tabla VI: Predicciones (en porcentaje) obtenidas para las distintas zonas con Parkes Error Grid para las medias y desviaciones típicas de los valores obtenidos con validación cruzada de cada paciente para los distintos modelos con horizonte de tiempo a 2 horas.

Modelo	t+120 train				t+120 test			
	A+B	C	D	E	A+B	C	D	E
<b>Paciente 559</b>								
<i>Avg</i>	81.5 ± 0.7	15.8 ± 0.5	2.7 ± 0.3	0.0 ± 0.0	80.8 ± 7.3	14.9 ± 4.0	4.3 ± 5.4	0.0 ± 0.0
<i>Last</i>	84.8 ± 0.7	13.1 ± 0.4	2.1 ± 0.1	0.0 ± 0.0	84.8 ± 6.5	13.1 ± 3.3	2.1 ± 1.0	0.0 ± 0.0
<i>PG</i>	86.1 ± 1.5	12.2 ± 1.6	1.7 ± 0.2	0.0 ± 0.0	85.8 ± 6.4	12.4 ± 2.3	1.7 ± 1.7	0.0 ± 0.0
<i>PG2</i>	86.6 ± 0.7	11.7 ± 0.4	1.7 ± 0.2	0.0 ± 0.0	86.5 ± 4.8	11.8 ± 2.4	1.8 ± 1.7	0.0 ± 0.0
<i>PG2<sub>com</sub></i>	<b>87.9 ± 0.6</b>	10.8 ± 0.4	1.4 ± 0.1	0.0 ± 0.0	<b>87.8 ± 4.4</b>	10.7 ± 2.1	1.5 ± 1.4	0.0 ± 0.0
<i>PGOS</i>	86.7 ± 0.6	11.6 ± 0.4	1.7 ± 0.2	0.0 ± 0.0	86.3 ± 4.9	11.8 ± 2.3	1.8 ± 1.6	0.0 ± 0.0
<b>Paciente 563</b>								
<i>Avg</i>	90.7 ± 0.9	8.8 ± 0.6	0.5 ± 0.1	0.0 ± 0.0	89.0 ± 5.9	10.1 ± 6.6	0.9 ± 1.4	0.0 ± 0.0
<i>Last</i>	92.0 ± 0.3	7.1 ± 0.3	0.8 ± 0.1	0.0 ± 0.0	92.0 ± 2.4	7.1 ± 2.5	0.8 ± 0.7	0.0 ± 0.0
<i>PG</i>	92.4 ± 2.6	7.1 ± 0.7	0.5 ± 0.1	0.0 ± 0.0	91.8 ± 6.1	7.6 ± 4.2	0.5 ± 0.6	0.0 ± 0.0
<i>PG2</i>	92.4 ± 3.2	7.2 ± 1.1	0.4 ± 0.1	0.0 ± 0.0	92.1 ± 8.1	7.4 ± 4.4	0.5 ± 0.8	0.0 ± 0.0
<i>PG2<sub>com</sub></i>	93.6 ± 0.6	6.0 ± 0.4	0.4 ± 0.1	0.0 ± 0.0	93.2 ± 7.2	6.3 ± 3.9	0.4 ± 0.6	0.0 ± 0.0
<i>PGOS</i>	<b>93.7 ± 11.6</b>	5.9 ± 2.0	0.3 ± 0.1	0.0 ± 0.0	<b>93.3 ± 13.7</b>	6.3 ± 4.5	0.4 ± 0.5	0.0 ± 0.0
<b>Paciente 570</b>								
<i>Avg</i>	89.8 ± 1.7	8.7 ± 0.7	1.5 ± 0.2	0.0 ± 0.0	91.1 ± 15.0	8.0 ± 4.6	1.0 ± 1.1	0.0 ± 0.0
<i>Last</i>	<b>94.9 ± 0.9</b>	4.6 ± 0.3	0.5 ± 0.1	0.0 ± 0.0	<b>94.9 ± 7.8</b>	4.6 ± 2.4	0.5 ± 0.6	0.0 ± 0.0
<i>PG</i>	93.9 ± 2.3	5.4 ± 0.6	0.6 ± 0.4	0.0 ± 0.0	94.1 ± 10.3	5.4 ± 3.4	0.5 ± 0.6	0.0 ± 0.0
<i>PG2</i>	94.3 ± 1.1	5.3 ± 0.4	0.5 ± 0.1	0.0 ± 0.0	94.0 ± 9.4	5.5 ± 3.8	0.5 ± 0.6	0.0 ± 0.0
<i>PG2<sub>com</sub></i>	94.6 ± 1.6	4.9 ± 0.7	0.6 ± 0.3	0.0 ± 0.0	94.4 ± 9.1	5.1 ± 3.9	0.5 ± 0.6	0.0 ± 0.0
<i>PGOS</i>	94.3 ± 1.1	5.1 ± 0.4	0.5 ± 0.1	0.0 ± 0.0	93.9 ± 9.0	5.6 ± 3.8	0.5 ± 0.6	0.0 ± 0.0
<b>Paciente 575</b>								
<i>Avg</i>	83.9 ± 0.6	13.4 ± 0.5	2.7 ± 0.2	0.0 ± 0.0	82.5 ± 5.8	15.1 ± 3.4	2.4 ± 1.3	0.0 ± 0.0
<i>Last</i>	83.2 ± 0.5	14.4 ± 0.7	2.3 ± 0.1	0.0 ± 0.0	83.2 ± 4.7	14.4 ± 6.1	2.3 ± 1.1	0.0 ± 0.0
<i>PG</i>	84.6 ± 0.4	13.4 ± 0.6	1.9 ± 0.2	0.0 ± 0.0	84.4 ± 6.1	13.7 ± 6.0	1.9 ± 0.8	0.0 ± 0.0
<i>PG2</i>	84.7 ± 0.6	13.5 ± 0.6	1.8 ± 0.2	0.0 ± 0.0	84.5 ± 6.2	13.6 ± 6.2	1.9 ± 0.9	0.0 ± 0.0
<i>PG2<sub>com</sub></i>	<b>86.7 ± 1.2</b>	11.9 ± 0.4	1.4 ± 0.1	0.0 ± 0.0	<b>86.4 ± 7.7</b>	12.2 ± 6.0	1.5 ± 1.0	0.0 ± 0.0
<i>PGOS</i>	84.7 ± 0.6	13.5 ± 0.7	1.8 ± 0.1	0.0 ± 0.0	84.4 ± 5.9	13.8 ± 6.1	1.8 ± 0.9	0.0 ± 0.0
<b>Paciente 588</b>								
<i>Avg</i>	94.6 ± 0.8	5.1 ± 0.4	0.3 ± 0.1	0.0 ± 0.0	95.0 ± 7.0	4.7 ± 3.3	0.4 ± 0.6	0.0 ± 0.0
<i>Last</i>	94.3 ± 0.7	5.1 ± 0.3	0.6 ± 0.1	0.0 ± 0.0	94.3 ± 6.9	5.1 ± 2.6	0.6 ± 0.8	0.0 ± 0.0
<i>PG</i>	94.9 ± 3.8	4.7 ± 0.5	0.4 ± 0.1	0.0 ± 0.0	94.7 ± 7.4	4.8 ± 3.2	0.4 ± 0.7	0.0 ± 0.0
<i>PG2</i>	95.1 ± 0.8	4.5 ± 0.3	0.4 ± 0.1	0.0 ± 0.0	94.9 ± 7.0	4.6 ± 3.2	0.5 ± 0.8	0.0 ± 0.0
<i>PG2<sub>com</sub></i>	<b>95.3 ± 0.8</b>	4.4 ± 0.3	0.4 ± 0.1	0.0 ± 0.0	<b>95.0 ± 7.1</b>	4.6 ± 3.4	0.4 ± 0.7	0.0 ± 0.0
<i>PGOS</i>	94.9 ± 5.0	4.5 ± 0.4	0.5 ± 0.1	0.0 ± 0.0	94.3 ± 8.0	5.2 ± 3.9	0.5 ± 0.9	0.0 ± 0.0
<b>Paciente 591</b>								
<i>Avg</i>	87.0 ± 1.0	11.2 ± 0.3	1.8 ± 0.2	0.0 ± 0.0	87.3 ± 6.4	10.8 ± 3.8	1.8 ± 2.2	0.0 ± 0.0
<i>Last</i>	89.8 ± 0.3	8.6 ± 0.3	1.6 ± 0.1	0.0 ± 0.0	89.8 ± 2.8	8.6 ± 2.9	1.6 ± 1.0	0.0 ± 0.0
<i>PG</i>	90.2 ± 1.0	8.2 ± 0.6	1.6 ± 0.2	0.0 ± 0.0	90.2 ± 5.8	8.2 ± 3.4	1.7 ± 1.8	0.0 ± 0.0
<i>PG2</i>	90.6 ± 0.5	7.7 ± 0.3	1.6 ± 0.2	0.0 ± 0.0	90.4 ± 6.0	8.0 ± 3.0	1.7 ± 1.8	0.0 ± 0.0
<i>PG2<sub>com</sub></i>	90.2 ± 1.6	8.4 ± 0.7	1.4 ± 0.2	0.0 ± 0.0	89.9 ± 5.2	8.6 ± 3.5	1.5 ± 1.5	0.0 ± 0.0
<i>PGOS</i>	<b>90.8 ± 0.6</b>	7.7 ± 0.3	1.5 ± 0.2	0.0 ± 0.0	<b>90.5 ± 6.4</b>	7.7 ± 2.7	1.7 ± 1.7	0.0 ± 0.0