

**XIII Congreso Español  
de Metaheurísticas,  
Algoritmos Evolutivos y  
Bioinspirados  
(XIII MAEB)**

**MAEB 9: SESIÓN ESPECIAL:  
APLICACIONES EN MEDICINA  
Y BIOTECNOLOGÍA**

**Organizadores:**

JOSÉ MANUEL COLMENAR Y JOSÉ IGNACIO HIDALGO







# Strain Design as Multiobjective Network Interdiction Problem: A Preliminary Approach\*

\*Note: The full contents of this paper have been published in the volume *Lecture Notes in Artificial Intelligence 11160* (LNAI 11160)

Marina Torres <i>Dept. of Computer Science and A.I.</i> <i>Universidad de Granada</i> Granada, Spain torresm@decsai.ugr.es	Shouyong Jiang <i>School of Computing</i> <i>Newcastle University</i> Newcastle, U.K. math4neu@gmail.com	David Pelta <i>Dept. of Computer Science and A.I.</i> <i>Universidad de Granada</i> Granada, Spain dpelta@decsai.ugr.es	Marcus Kaiser <i>School of Computing</i> <i>Newcastle University</i> Newcastle, U.K. marcus.kaiser@ncl.ac.uk
--	--	---	--

Natalio Krasnogor  
*School of Computing*  
*Newcastle University*  
Newcastle, U.K.  
natalio.krasnogor@ncl.ac.uk

**Abstract**—Computer-aided techniques have been widely applied to analyse the biological circuits of microorganisms and facilitate rational modification of metabolic networks for strain design in order to maximise the production of desired biochemicals for metabolic engineering. Most existing computational methods for strain design formulate the network redesign as a bilevel optimisation problem. While such methods have shown great promise for strain design, this paper employs the idea of network interdiction to fulfil the task. Strain design as a Multiobjective Network Interdiction Problem (MO-NIP) is proposed for which two objectives are optimised (biomass and bioengineering product) simultaneously in addition to the minimisation of the costs of genetic perturbations (design costs). An initial approach to solve the MO-NIP consists on a Nondominated Sorting Genetic Algorithm (NSGA-II). The shown examples demonstrate the usefulness of the proposed formulation for the MO-NIP and the feasibility of the NSGA-II as a problem solver.

**Index Terms**—strain design, network interdiction, metabolic networks, multiobjective bilevel optimisation

# 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.

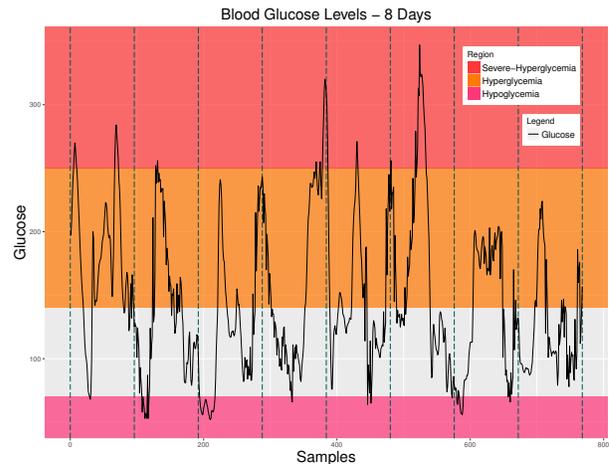


Figura 1: Valores de Glucosa Reales - 8 Días

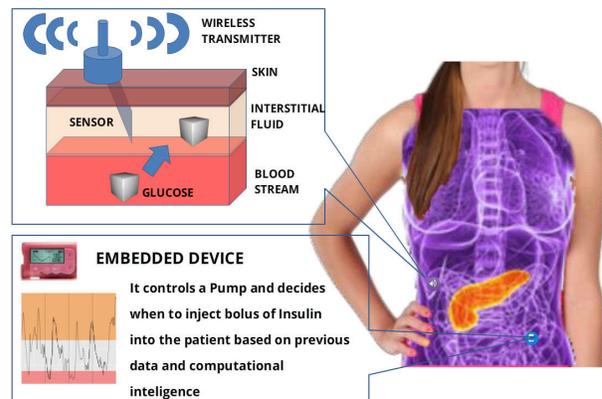


Figura 2: Páncreas Artificial

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<sub>com</sub>*, 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<sub>com</sub>* (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<sub>com</sub>*, 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

# Registrado de Imagen Médica mediante Algoritmos de Optimización basados en Arrecifes de Coral

E. Bermejo\*, M. Chica†, S. Damas\*, S. Salcedo-Sanz‡ and O. Cordon\*§

\* Instituto Andaluz en Data Science and Computational Intelligence (DaSCI), Universidad de Granada, Granada 18071.

Email: ebermejo@decsai.ugr.es, sdamas@ugr.es

† School of Electrical Engineering and Computing, The University of Newcastle, Callaghan 2380, Australia.

Email: manuel.chicaserrano@newcastle.edu.au

‡ Departamento de Teoría de la Señal y Comunicaciones, Universidad de Alcalá, Alcalá de Henares 28805.

Email: sancho.salcedo@uah.es

§ Centro de Investigación en Tecnologías de la Información, Universidad de Granada, Granada 18071

Email: oordon@decsai.ugr.es

**Resumen**—El Registrado de Imágenes (RI) conlleva la transformación de un conjunto de imágenes en un sistema de coordenadas común. Generalmente, la búsqueda de la correspondencia óptima entre imágenes es considerada un problema de optimización. Los enfoques basados en metaheurísticas permiten resolver el problema de RI superando las numerosas limitaciones de los métodos tradicionales. En esta contribución consideraremos un reciente método bio-inspirado: Optimización por Arrecifes de Coral (CRO). Este novedoso algoritmo simula los fenómenos naturales que subyacen a un arrecife de coral. Adaptaremos el algoritmo siguiendo dos enfoques de registrado diferentes: basado en características y basado en intensidades. Además, diseñaremos un exhaustivo estudio experimental considerando un problema de RI médica con resonancias magnéticas cerebrales. Los resultados muestran cómo CRO supera al estado del arte en RI tanto en robustez como en precisión y eficiencia siguiendo ambos enfoques.

## I. INTRODUCCIÓN

En el ámbito de la imagen médica existe un gran interés en relacionar la información procedente de diferentes imágenes. Esta tarea es relevante en un gran número de aplicaciones: diagnóstico, monitorización y seguimiento de enfermedades o cirugía asistida, entre otras. Generalmente, estas aplicaciones requieren la integración de información visual adquirida por diferentes dispositivos o condiciones. Así, el Registrado de Imágenes (RI) [1] es considerado como una tarea de preprocesamiento esencial en imagen médica, ya que permite alinear múltiples imágenes con contenido común.

Generalmente, los métodos de RI consideran una transformación espacial para alinear imágenes solapando regiones coincidentes. El problema se trata como un proceso de optimización iterativo que explora el espacio de transformaciones posibles. La calidad de una solución se define así por el grado de semejanza entre imágenes tras la transformación, que se mide mediante una métrica de similitud [2].

El proceso de RI permite emplear una imagen completa (enfoque de intensidades) o basarse en partes distintivas de la imagen (características). Los enfoques de características [1] agilizan el proceso de optimización reduciendo la complejidad del problema, al usar una porción de las imágenes. Sin embargo, estos métodos dependen en gran medida del proceso

de extracción de características, una etapa propensa a errores si la información extraída no es suficientemente representativa. Por otro lado, los enfoques de intensidades permiten procesar un volumen de datos mayor, a costa de incrementar los requisitos computacionales [2]. Así, la alineación se guía por la distribución de intensidades (niveles de gris) de la imagen, aumentando la precisión de estos métodos, pero también su sensibilidad al ruido o a las condiciones de iluminación.

Métodos tradicionales, como Iterative Closest Point (ICP) [3] tienden a quedar atrapados en mínimos locales debido a factores como el ruido, la discretización o la desalineación entre imágenes. Mientras tanto, aquellos basados en algoritmos evolutivos u otras metaheurísticas (MHs) permiten superar estas desventajas y ofrecen un buen rendimiento que los hace idóneos para problemas de RI médica.

Hasta la fecha se han propuesto numerosos estudios comparativos para analizar el rendimiento de diferentes MHs en problemas de RI empleando modelos craneales e imágenes médicas en 3D [4], [5]. En ambos estudios se incluyen métodos de RI relevantes: ICP, una versión adaptativa de descenso de gradiente estocástico (ASGD) y algoritmos genéticos (GAs). Entre las diversas MHs destaca una versión memética del método scatter search (SS) propuesta en [4] y extendida en [6], obteniendo resultados excelentes y consolidando su dominancia en el estado del arte sobre métodos convencionales.

Recientemente se ha propuesto una MH bio-inspirada denominada *Coral Reefs Optimization* (CRO) [7]. CRO es un algoritmo evolutivo basado en la simulación de los fenómenos naturales que ocurren durante la formación de un arrecife de coral. A lo largo de su vida, los corales atraviesan diversas fases, como la reproducción, el asentamiento de larvas o la lucha por un espacio en el arrecife donde sobrevivir. CRO emula dichos procesos favoreciendo un equilibrio entre diversidad y especificidad, ingredientes idóneos para abordar problemas complejos de optimización.

Dados los buenos resultados de CRO en diferentes problemáticas reales, consideramos que su diseño puede comportarse adecuadamente en problemas de RI complejos. Así, en este trabajo diseñamos un nuevo método de RI basado en



CRO donde tanto la función objetivo como la codificación han sido adaptados a un problema concreto de RI médica.

Para validar nuestra propuesta, hemos desarrollado un exhaustivo estudio experimental comparando CRO contra algunos métodos evolutivos del estado del arte en RI, considerando tanto enfoques de intensidades como de características. Esta comparativa es una extensión de nuestra propuesta donde adaptamos CRO al enfoque de intensidades [8]. Se han utilizado pares de imágenes monomodales extraídas de Brainweb, dataset disponible gracias a McGill University [9].

## II. REGISTRADO DE IMÁGENES: PLANTEAMIENTO

Generalmente, en los problemas de RI médica se parte de dos imágenes, una de referencia o modelo ( $I_M$ ), y una imagen móvil o escena ( $I_E$ ) que se alinea con la geometría del modelo. El objetivo del proceso de registrado es encontrar una transformación geométrica  $T$  que haga el modelo  $I_M$  y la escena transformada  $T(I_E)$  sean lo más similares posible. El grado de semejanza entre imágenes se mide considerando una métrica de similitud. Así, el RI se formula como un problema de maximización en el espacio de transformaciones:

$$\operatorname{argmax}_{T \in \text{Transformaciones}} \text{Similitud}(I_M, T(I_E)).$$

El proceso iterativo de un método de RI consta de tres componentes principales: el modelo de transformación, la métrica de similitud, y el optimizador. En primer lugar, el optimizador estima una transformación geométrica para alinear las imágenes, que viene determinada por el *modelo de transformación*. La Figura 1 ilustra el efecto de aplicar diferentes tipos de transformación a una resonancia magnética (IRM) cerebral. La elección del modelo adecuado depende tanto de la aplicación concreta como de la naturaleza de las imágenes. En ciertos contextos un modelo más simple (transformación de traslación) puede ser adecuado, mientras que otras aplicaciones que suponen una estimación de movimiento requieren modelos deformables. De cualquier forma, el tipo de transformación debe ser seleccionado cuidadosamente, ya que de esta elección dependen tanto los requisitos computacionales en base al número de parámetros a estimar, como los efectos geométricos resultantes de la transformación.

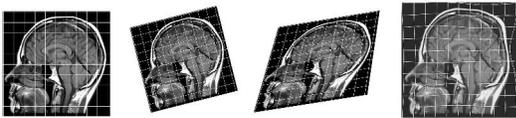


Figura 1: Aplicación de diferentes modelos de transformación: similitud, finidad y b-spline, sobre una escena inicial.

Tras ello, la *métrica de similitud* mide la calidad de la alineación entre la escena transformada y el modelo. La elección de esta métrica es un paso crucial en el diseño de cualquier método de RI y depende del enfoque considerado. Enfoques de características suelen considerar el error cuadrático medio (ECM) como métrica para medir la distancia entre

características correspondientes. Enfoques de intensidades se basan en la relación entre distribuciones de intensidad para evaluar la similitud. Existen multitud de métricas disponibles (correlación normalizada (CN) o información mutua (IM), entre otras) y su idoneidad viene determinada por el proceso de adquisición y la relación entre imágenes.

Por último, el *optimizador* refina la solución hasta alcanzar un criterio de parada (cuando se encuentra una solución adecuada o se realiza un número determinado de iteraciones). Así, la optimización de la transformación espacial supone un proceso de búsqueda iterativo sobre el espacio de transformaciones geométricas. De acuerdo con las características del espacio de búsqueda, podemos encontrar dos estrategias diferentes. El registrado puede considerarse como un problema de optimización continua cuando la búsqueda se realiza en base a los parámetros de la transformación. Alternativamente, la búsqueda se puede realizar en el espacio de correspondencias entre pares de características (enfoque de características) o entre áreas de la imagen (enfoque de intensidades).

### A. Algoritmos evolutivos relevantes en RI médica

#### RI basado en intensidades

*r-GA\**, Valsecchi et al.: r-GA\* se basa en un *algoritmo genético* [10]. El optimizador está diseñado en codificación real donde una solución almacena los parámetros de la transformación en un vector. Los operadores genéticos también son reales: cruce por mezcla alfa (BLX- $\alpha$ ) [11] y mutación aleatoria. r-GA\* es capaz de adaptar diferentes métricas de similitud y distintos modelos de transformación. Además, integra el uso de múltiples resoluciones en combinación con un mecanismo de reinicio y adaptación al espacio de búsqueda. Estos mecanismos permiten agilizar el proceso de optimización y mitigar una convergencia prematura.

*SS\**, Valsecchi et al.: SS\* [6] es una variante del diseño original de *scatter search* o búsqueda dispersa, donde el conjunto de referencia se divide en dos para contener las soluciones más diversas y las de mayor calidad. El algoritmo han sido diseñados específicamente para RI [4]. Así, SS\* considera el uso de memoria basada en frecuencia para el método de diversificación, la combinación de soluciones se basa en el cruce BLX- $\alpha$ , y el método de mejora en el operador PMX- $\alpha$  [12]. El método de actualización clasifica las soluciones según su calidad o diversidad. Finalmente, el método de control de duplicación previene la aparición de copias idénticas de una solución en la población.

#### RI basado en características

*HE-GA*, He y Narayana: Propuesto en [13], considera un diseño genético con codificación real combinado con *dividing rectangle*, un método de optimización global basado en ramificación y poda. El genético sigue un modelo generacional elitista con cruce aritmético y mutación uniforme, aplicado para estimar una solución preliminar. Esta solución es posteriormente refinada por búsqueda local. Incorpora un mecanismo de reinicio para evitar una convergencia prematura.

SS, Santamaría *et al.*: Este método basado en SS fue originalmente diseñado para problemas de RI de rango [4]. Los autores adoptan una métrica de similitud basada en el error cuadrático mediano (ECMed), robusta a imágenes con bajo solapamiento, integran una estructura de datos para acelerar el cálculo de la función objetivo, y una búsqueda local basada en cruce (XLS) como método de mejora. SS también incluye un mecanismo de reinicio para evitar mínimos locales.

### III. CORAL REEFS OPTIMIZATION ALGORITHM

CRO [7], [14] es un algoritmo evolutivo basado en los procesos naturales de un arrecife de coral. Denominamos  $\mathcal{R}$  a un arrecife representado por una cuadrícula  $R_1 \times R_2$ , donde cada posición  $(i, j)$  permite albergar a un coral,  $C_{i,j}$ , que representa una solución al problema de optimización considerado. El algoritmo inicializa varios corales en posiciones aleatorias de  $\mathcal{R}$ , dejando huecos libres para el asentamiento de futuras generaciones de corales. La ratio entre posiciones libres/ocupadas viene dado por un parámetro  $\rho_0$ , donde  $0 < \rho_0 < 1$ .

Existen diferentes mecanismos reproductivos en un arrecife de coral. En CRO se recrean aplicando diferentes operadores de forma secuencial durante la segunda fase del algoritmo:

#### 1. Reproducción sexual externa (*Broadcast Spawning*).

En cada iteración  $k$  del algoritmo se aplica:

- 1.a. Selección uniforme de una fracción aleatoria  $F_b$  de corales (denominados *broadcast spawners*) con respecto al total de corales existentes en el arrecife.
- 1.b. Formación de larvas. Se eligen dos broadcast spawners y se aplica un operador de cruce. Una vez seleccionados, éstos no vuelven a atravesar otra etapa reproductiva durante la iteración  $k$ . La selección puede ser uniforme, aleatoria o proporcional al fitness (selección por ruleta).

#### 2. Reproducción sexual interna (*Brooding*).

La reproducción de los corales hermafroditas se modela como una mutación sobre una fracción de los corales restantes  $1 - F_b$ . En estos casos, un porcentaje  $P_i$  del coral muta.

#### 3. Asentamiento de larvas.

Las nuevas larvas generadas seleccionan una posición aleatoria del arrecife para formarse, siempre que la posición  $(i, j)$  esté libre. En caso de que esté ocupada, la nueva larva se asentará si su valor de fitness mejora a la del coral existente. El máximo número de intentos para ocupar una posición en cada iteración  $k$  está limitado por el parámetro  $\eta = 3$ .

#### 4. Reproducción asexual.

Los corales pueden fragmentarse y reasentarse en el arrecife (*budding*). CRO modela este mecanismo ordenando los corales por su nivel de vida (fitness  $f(C_{i,j})$ ). Una pequeña fracción ( $F_a$ ) de los corales se duplica, muta con probabilidad  $P_a$ , e intenta asentarse en el arrecife como en el paso 3.

#### 5. Depredación.

Tras cada iteración  $k$ , se simula la muerte de un pequeño número de corales, liberando espacio en el arrecife para la siguiente generación. La depredación se aplica con una probabilidad muy pequeña ( $P_d$ ) a una fracción  $F_d$  de corales con peor vida.

## IV. DISEÑO EXPERIMENTAL

Diseñamos un estudio experimental considerando los algoritmos descritos en la Sección A ya que mejoran el comportamiento de alternativas clásicas [5], [6], [8], [10]. Así, pretendemos realizar un análisis exhaustivo del comportamiento de nuestra propuesta en comparación con los métodos evolutivos más relevantes en un problema de RI médica.

### A. Escenarios y dataset de imágenes

Las imágenes mono-modales utilizadas en esta experimentación han sido obtenidas de un repositorio público muy conocido, BrainWeb (McGill University [9]). Este dataset se compone de imágenes cerebrales simuladas a partir de IRMs sintéticas. Con el objetivo de considerar diferentes niveles de complejidad, las imágenes contienen ruido (hasta un 5%) y múltiples lesiones (esclerosis). Las imágenes tienen un tamaño común de  $60 \times 181 \times 217$  voxels.

A partir de la curvatura cerebral de cada imagen extraemos un conjunto de puntos mediante un detector de líneas de cresta [15]. En enfoques de características, los valores heurísticos de curvatura se emplean para guiar al optimizador durante la estimación de la transformación de registrado.

Para generar diferentes escenarios, cada imagen ha sido transformada usando una de las cuatro transformaciones de similitud (aplicando rotación, traslación y escalado uniforme) de la Tabla I. En total, el estudio consta de 16 instancias de RI creadas a partir de pares de imágenes con diferentes transformaciones. Los escenarios son:  $I_1$  vs.  $T_i(I_2)$ ,  $I_1$  vs.  $T_i(I_3)$ ,  $I_1$  vs.  $T_i(I_4)$ , e  $I_2$  vs.  $T_i(I_4)$ , para  $i = 1, 2, 3, 4$ .

Tabla I: Parámetros de la transformación: ángulo de rotación ( $\lambda$ ), eje de rotación ( $a_x, a_y, a_z$ ), vector de traslación ( $t_x, t_y, t_z$ ), y factor de escalado uniforme  $s$ .

	$\lambda$	$a_x$	$a_y$	$a_z$	$t_x$	$t_y$	$t_z$	$s$
$T_1$	115	-0.863	0.259	0.431	-26	15.5	-4.6	1
$T_2$	168	0.676	-0.290	0.676	6	5.5	-4.6	0.8
$T_3$	235	-0.303	-0.808	0.505	16	-5.5	-4.6	1
$T_4$	276.9	-0.872	0.436	-0.218	-12	5.5	-24.6	1.2

### B. Diseño Experimental y Configuración

Los métodos de la comparativa difieren en el proceso de optimización: los enfoques de características se guían por los puntos de curvatura y las métricas ECM o ECMed, mientras que los enfoques de intensidades se guían por los valores de intensidad y métricas como CN o IM. Los resultados de diferentes métricas no son comparables, por lo que es necesario definir un marco comparativo para evaluar objetivamente los resultados de la alineación. Así, consideramos una medida común para evaluar la calidad de todas las soluciones. Una vez el algoritmo alcanza una solución, se calcula el ECM entre las características anatómicas (líneas de cresta) de la escena transformada y las del modelo. Este valor será usado como comparativa de la calidad final para todos los métodos.

La transformación de similitud que relaciona las imágenes considera un rango de  $[-30, 30]$  para los parámetros de la



Tabla II: Parameter configuration for considered IR methods.

r-GA*		SS*		CRO*		HE-GA		SS		CRO	
Individuos	100	Tamaño población	12	Tamaño arrecife	80	Individuos	60	Tamaño población	30	Tamaño arrecife	70
Generaciones	75	Generaciones	18	Generaciones	65	Reinicios	5	Reinicios	5	Reinicios	15
Reinicios	5	Reinicios	3	Reinicios	8	Factor $\alpha$	0.3	XLS iter.	100	XLS iter.	100
Torneo	3	PMX iter.	12	PMX iter.	15	Prob. cruce	0.7	Factor $\alpha$	0.3	$\rho_0$	0.6
Factor $\alpha$	0.3	Factor $\alpha$	0.3	$\rho_0$	0.6	Prob. mutación	0.2	Conjunto Referencia	8	$F_{broad}$	0.9
Prob. cruce	0.5	conjunto referencia	4	$F_{broad}$	0.8					$P_d$	0.05
Prob. mutación	0.1			$P_d$	0.15					$k$	3

traslación y  $[0,75,1,25]$  para el factor de escalado. No se aplica restricción en el ángulo de rotación ( $[0,360]$ ), ni en sus ejes ( $[-1,1]$ ). De esta forma, la transformación se codifica mediante un vector real de siete parámetros en el enfoque de intensidades: versor de rotación  $(\theta_x, \theta_y, \theta_z)$ , vector de traslación  $(t_x, t_y, t_z)$  y escalado uniforme  $s$ . En el enfoque de características, la rotación se codifica con cuatro parámetros: tres para el eje de rotación  $(r_x, r_y, r_z)$  y uno para el ángulo  $\theta_r$ , resultando en una codificación de ocho parámetros.

Para evaluar CRO en comparación con el estado del arte en RI (Sección A), diseñamos un estudio experimental considerando métodos representativos de ambos enfoques (intensidades y características). Así, comparamos seis métodos diferentes: i) *Enfoque de intensidades*: r-GA\*, SS\* y CRO\*. ii) *Enfoque de características*: HE-GA, SS y CRO.

Dadas las diferencias en términos de requisitos computacionales, la configuración de parámetros óptima para cada algoritmo varía según el enfoque considerado (ver Tabla II). Estos parámetros han sido ajustados manualmente con un par de imágenes no incluido en la experimentación para evitar sesgos. Se han seguido las directrices marcadas por los autores en estudios previos [5], [7]. A continuación detallamos las características distintivas de los métodos:

*Diseño de métodos basados en intensidades*:: La implementación se ha realizado en Elastix [16], una herramienta open-source específicamente diseñada para RI médica de intensidades. El método basado en CRO se ha propuesto en [8] con resultados prometedores. El optimizador se guía por la métrica normalizada de información mutua (IMN):

$$NMI(I_A, I_B) = \frac{\sum_{a \in I_A} \sum_{b \in I_B} p_{AB}(a, b) \log(p_A(a)p_B(b))}{\sum_{a \in I_A} \sum_{b \in I_B} p_{AB}(a, b) \log p_{AB}(a, b)},$$

donde  $p_{AB}$  es la probabilidad conjunta y  $p_A, p_B$  son las probabilidades marginales discretas de las intensidades de las imágenes.

Además, el diseño de los enfoques de intensidades integra dos componentes: i) **Multi-resolución** para reducir el coste computacional del proceso, mediante una reducción de la resolución y un suavizado gaussiano para crear dos representaciones (pirámides) de la imagen. En la primera resolución, el optimizador considera una pirámide de baja resolución para aproximar la transformación. La segunda resolución consiste en una etapa de refinado para mejorar la calidad de la alineación. ii) **Mecanismo de reinicio** para asegurar que el algoritmo es capaz de recuperarse de un posible estancamiento y encuentre una buena solución. En la primera resolución, se

aplica un reinicio del algoritmo un número determinado de veces. La mejor solución de esta etapa pasa a la etapa de refinado. Dado el gran volumen de datos, el tiempo de ejecución total de cada algoritmo se ha limitado a 180 segundos.

*Diseño de métodos basados en características*:: Implementados en C++ y compilados con la herramienta GNU/g++, siguen una estructura similar al enfoque de intensidades. Puesto que ECM ha sido elegido para la comparativa final, consideramos una métrica de optimización diferente para evitar favorecer este diseño. Así, consideramos una variante de ECMed específicamente diseñada para IR:

$$F(f, I_E, I_M) = w_1 \cdot (1 / (1 + \sum_{i=1}^N ||(sR\vec{p}_i + \vec{t}_i) - \vec{p}_j^i||)) + w_2 \cdot (1 / (1 + |p_c^e - p^m|)),$$

donde  $f$  es la solución que codifica los parámetros de la transformación;  $p_i$  es un punto 3D de la escena y  $p_j$  su correspondencia más cercana en el modelo utilizando una estructura de indexación (*grid closest point*);  $w_1$  y  $w_2$  ( $w_1 + w_2 = 1$ ) son pesos para balancear los términos de la función;  $p_c^e$  es el radio de la esfera que engloba la escena transformada y  $p^m$  el radio de la esfera sobre el modelo. Nótese que el primer término de  $F$  corresponde con la función ECMed entre puntos vecinos.

Como componentes específicos de este enfoque, se aplica el reinicio de la población cuando el optimizador detecta estancamiento, por ejemplo, tras 15 iteraciones sin mejora. El conjunto de características (líneas de cresta) es bastante reducido, por lo que no se requiere una estrategia multi-resolución y el tiempo de parada se limita a 20 segundos.

### C. Análisis de Resultados

En la Tabla III se recogen los resultados de la alineación entre imágenes: el error entre pares de características (ECM) y el ranking parcial de cada método en los 16 escenarios. Nótese que el mínimo valor posible de ECM se explica por los niveles de ruido y la presencia de lesiones en la comparación entre imágenes. El valor óptimo de ECM se ha resaltado entre paréntesis en la Tabla III. Además, se ha realizado un análisis estadístico sobre los valores medios para buscar diferencias significativas entre el algoritmo con mejor ranking y el resto. Los resultados del ranking, test de Bonferroni-Dunn y test de Holm se recogen en la Tabla IV.

En general, los métodos basados en características mejoran a sus homólogos de intensidades en la mayoría instancias. El diseño del optimizador hace que este tipo de métodos pueda alcanzar una mayor precisión en menor tiempo. Así,

Tabla III: Mínimo (m), media ( $\mu$ ) y desviación estándar (sd) para los valores de ECM (mm) y ranking medio (R) para cada instancia de RI. Las columnas subrayadas corresponden con los métodos basados en intensidades.

Algorithm	$I_1$ vs $T_1(I_2)$ [31]				$I_1$ vs $T_2(I_2)$ [31]				$I_1$ vs $T_3(I_2)$ [31]				$I_1$ vs $T_4(I_2)$ [31]			
	m	$\mu$	sd	R												
r-GA*	37.32	134	>99	6	39.60	86.43	44	5	43.29	5506	>99	6	32.19	84.36	91	6
HE-GA	42.69	101	47	5	<b>31.85</b>	44.27	17	4	<b>31.96</b>	42.19	9	4	32.88	59.23	28	5
SS*	36.01	38.14	5.9	4	36.53	38.96	2.9	3	36.78	1755	>99	5	32.65	32.96	0.7	4
SS	32.06	32.84	2.2	2	32.11	119	>99	6	32.17	41.08	47	2	<b>32.00</b>	32.25	0.1	2
CRO*	36.35	36.79	0.3	3	36.26	36.78	0.3	2	40.72	41.41	0.5	3	32.67	32.79	0.1	3
CRO	32.11	<b>32.25</b>	0.1	1	32.10	<b>33.96</b>	5.3	1	32.14	<b>32.27</b>	0.1	1	32.14	<b>32.24</b>	0.0	1

Algorithm	$I_1$ vs $T_1(I_3)$ [42]				$I_1$ vs $T_2(I_3)$ [42]				$I_1$ vs $T_3(I_3)$ [42]				$I_1$ vs $T_4(I_3)$ [42]			
	m	$\mu$	sd	R												
r-GA*	50.61	142	>99	6	45.61	97.75	81	6	66.00	8470	>99	6	42.93	86.11	53	5
HE-GA	61.65	122	48	5	<b>42.91</b>	61.01	24	4	56.52	74.44	18	5	58.74	114	42	6
SS*	50.20	61.34	50	4	43.65	46.36	3.9	3	55.22	57.85	3.2	4	44.68	45.18	0.3	4
SS	<b>42.96</b>	58.69	59	3	43.08	83.82	>99	5	43.06	50.95	41	2	42.95	43.44	0.6	2
CRO*	51.23	51.77	0.3	2	43.54	43.73	0.1	1	55.78	56.53	0.4	3	44.61	45.17	0.3	3
CRO	43.01	<b>43.29</b>	0.2	1	42.94	<b>44.06</b>	2.9	2	<b>42.97</b>	<b>43.45</b>	0.2	1	<b>42.87</b>	<b>43.39</b>	0.3	1

Algorithm	$I_1$ vs $T_1(I_4)$ [46]				$I_1$ vs $T_2(I_4)$ [46]				$I_1$ vs $T_3(I_4)$ [46]				$I_1$ vs $T_4(I_4)$ [46]			
	m	$\mu$	sd	R												
r-GA*	51.06	151	>99	6	58.92	124	98	4	60.76	9428.7	>99	6	<b>46.44</b>	82.74	45	5
HE-GA	60.07	119	51	5	84.3	1011	>99	6	51.18	78.92	25	5	51.35	92.56	30	6
SS*	52.53	53.67	1.3	3	46.48	48.12	1.5	2	54.32	60.10	5.2	3	46.65	47.58	0.5	3
SS	46.79	61.77	77	4	46.52	154	>99	5	46.87	64.26	68	4	46.67	<b>47.33</b>	0.5	1
CRO*	52.78	53.24	0.3	2	<b>46.02</b>	<b>46.37</b>	0.1	1	57.46	58.02	0.3	2	47.03	47.70	0.3	4
CRO	<b>46.66</b>	<b>48.85</b>	6.0	1	46.77	48.80	3.4	3	<b>46.80</b>	<b>47.57</b>	0.6	1	46.64	47.37	0.6	2

Algorithm	$I_2$ vs $T_1(I_4)$ [28]				$I_2$ vs $T_2(I_4)$ [28]				$I_2$ vs $T_3(I_4)$ [28]				$I_2$ vs $T_4(I_4)$ [28]			
	m	$\mu$	sd	R												
r-GA*	32.15	3293	>99	6	32.20	68.26	33	6	33.38	3656	>99	6	29.53	57.89	32	5
HE-GA	33.59	94.40	35	5	29.94	46.64	41	5	29.88	37.67	14	3	31.34	78.25	33	6
SS*	34.80	35.66	0.5	3	30.36	33.07	3.7	3	38.85	43.70	9.6	5	<b>28.44</b>	29.69	0.5	4
SS	28.54	44.21	79	4	<b>28.61</b>	35.27	35	4	<b>28.46</b>	<b>28.80</b>	0.1	1	28.57	<b>28.75</b>	0.1	1
CRO*	35.06	35.54	0.2	2	30.25	30.47	0.1	2	39.11	40.66	0.5	4	29.36	29.64	0.1	3
CRO	<b>28.32</b>	<b>28.79</b>	0.3	1	<b>28.61</b>	<b>29.01</b>	0.6	1	28.57	28.82	0.1	2	28.59	28.76	0.1	2

el enfoque de características permite a los métodos obtener mejores resultados mínimos, a expensas de incrementar la varianza. Por otro lado, los resultados de los métodos basados en intensidades (excepto r-GA\*) son más robustos, obteniendo una menor desviación que los métodos de características.

Ambos genéticos tienen un comportamiento inestable, pero alcanzan buenos resultados mínimos. r-GA\* alcanza la última posición del ranking (5.63) debido a obtener los mayores errores en la comparativa. A pesar de su pobre comportamiento, HE-GA es capaz de obtener el mejor resultado mínimo en tres de los dieciséis escenarios, con un ranking de 4.94.

SS\* y SS mejoran considerablemente a los genéticos, con un comportamiento consistente excepto en cuatro escenarios, donde no terminan de converger. Así, los métodos basados en SS alcanzan las posiciones medias del ranking (3.0 y 3.56, respectivamente). Aunque tienen un rendimiento similar en cuanto al ECM, SS mejora a SS\* al obtener mejores resultados mínimos a pesar de su mayor desviación estándar.

En cuanto a nuestras propuestas, CRO\* obtiene mejor ECM medio que ambos métodos basados en SS, pero es incapaz de alcanzar su precisión en valores mínimos. Por otro lado, la adaptación al enfoque de características (CRO) mejora tanto a CRO\* como al actual método del estado del arte, SS. En cuanto al ECM medio, CRO mejora a SS en doce escenarios. Así, CRO\* alcanza la segunda posición con 2.5, mientras que CRO es primero con un ranking de 1.38, gracias a su excelente balance entre exploración y explotación. Ambos enfoques basados en CRO son capaces de mejorar al actual estado del arte en RI, destacando nuestra propuesta basada en características por obtener el mejor rendimiento.

Al aplicar el test estadístico de Friedman obtenemos un resultado de  $\chi_F^2 = 92,0$ , con un p-valor correspondiente de  $< 10^{-16}$ . Dado que  $p$  es menor que el nivel de significancia considerado ( $\alpha = 0,01$ ), el test concluye que hay diferencias significativas entre métodos. Complementamos el análisis con el test de Holm (Tabla IV) comparando CRO (control) con el resto de métodos. El p-valor resultante para ambos tests revela que existen diferencias significativas entre CRO y tanto los métodos genéticos como aquellos basados en SS.

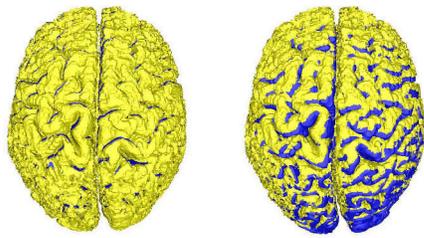
Tabla IV: Ranking de Friedman y resultados estadísticos con CRO como método de control para los test de Bonferroni y Holm, de acuerdo con el valor medio de ECM.

	Ranking	Bonferroni-Dunn p	Holm p
CRO	1.38	-	-
CRO*	2.50	0,44	0,09
SS	3.00	0,07	< 0,05
SS*	3.56	< 0,01	< 0,01
HE-GA	4.94	< 0,01	< 0,01
r-GA*	5.63	< $10^{-10}$	< $10^{-10}$

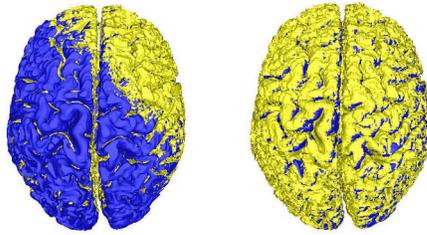
Por último, incluimos un análisis visual de los resultados obtenidos en uno de los escenarios más complejos. La Figura 2 muestra la mejor alineación entre imágenes obtenida por cada algoritmo. Destacamos los resultados de nuestras propuestas basadas en CRO, en especial el enfoque de características, ya que mejoran considerablemente la precisión de la alineación resultante.

## V. CONCLUSIONES

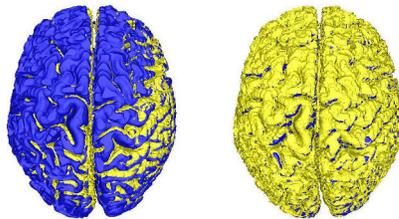
En este trabajo describimos el diseño e implementación de una novedosa técnica bio-inspirada, conocida como CRO, para



HE-GA (51.2) vs. r-GA\* (60.8)



SS (46.9) vs. SS\* (54.3)



CRO (46.8) vs. CRO\* (57.5)

Figura 2: Resultados del solapamiento entre el modelo (azul) y la escena (amarillo). La figura aporta una comparativa de los resultados entre el enfoque de características (izquierda) y de intensidades (derecha) para el escenario  $n^{\circ}11$  ( $I_1$  vs.  $T_3(I_4)$ ). Entre paréntesis se indica el mínimo ECM alcanzado.

resolver el problema de RI médica en 3D de forma robusta y eficiente. En particular, abordamos el problema considerando dos enfoques, usando la distribución de intensidades de las imágenes y un conjunto reducido de puntos característicos, cada uno con diferentes niveles de complejidad. CRO integra en su diseño una estrategia de búsqueda balanceada, que le permite obtener resultados excelentes en diferentes problemas reales. Así, adaptamos CRO al problema de RI siguiendo ambos enfoques y analizamos su comportamiento.

Para evaluar el rendimiento de nuestra propuesta comparamos ambas implementaciones de CRO con algunos de los métodos evolutivos más relevantes de la literatura en RI. Diseñamos un exhaustivo estudio experimental con 16 escenarios de RI mono-modal a partir de cuatro IMRs cerebrales sintéticos, con diferentes niveles de ruido y lesiones. Las imágenes provienen del repositorio BrainWeb (McGill University [9]).

CRO obtiene unos resultados excelentes considerando ambos enfoques de registrado, y mejora a métodos consolidados en RI demostrando gran robustez y eficiencia al abordar problemas de optimización complejos como es el RI médica.

#### AGRADECIMIENTOS

Este trabajo ha sido financiado parcialmente por los proyectos: TIN2015-67661-P, incluyendo al Fondo Europeo de Desarrollo Regional (ERDF), al Ministerio de Economía y TIN2014-54583-C2-2-R, de la Comisión Interministerial de Ciencia y Tecnología (CICYT).

#### REFERENCIAS

- [1] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003.
- [2] A. A. Goshtasby, *2-D and 3-D Image Registration*. Wiley Interscience, 2005.
- [3] P. J. Besl and N. D. McKay, "A Method for Registration of 3-D Shapes," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.
- [4] J. Santamaría, O. Cordón, S. Damas, J. García-Torres, and A. Quirin, "Performance evaluation of memetic approaches in 3D reconstruction of forensic objects," *Soft Computing*, vol. 13, no. 8-9, pp. 883–904, 2009.
- [5] S. Damas, O. Cordón, and J. Santamaría, "Medical Image Registration Using Evolutionary Computation: An Experimental Survey," *IEEE Computational Intelligence Magazine*, vol. 6, no. 4, pp. 26–42, 2011.
- [6] A. Valsecchi, S. Damas, J. Santamaría, and L. Marrakchi-Kacem, "Intensity-based image registration using scatter search," *Artificial Intelligence in Medicine*, vol. 60, no. 3, pp. 151–163, 2014.
- [7] S. Salcedo-Sanz, J. Del Ser, I. Landa-Torres, S. Gil-López, and J. A. Portilla-Figueras, "The coral reefs optimization algorithm: A novel metaheuristic for efficiently solving optimization problems," *The Scientific World Journal*, vol. 2014, pp. 1–15, 2014.
- [8] E. Bermejo, M. Chica, S. Salcedo-Sanz, and O. Cordon, "Coral Reef Optimization for Intensity-based Medical Image Registration," in *IEEE Congress on Evolutionary Computation, CEC 2017 - Proceedings*. IEEE, 2017, pp. 533–540.
- [9] D. L. Collins, A. P. Zijdenbos, V. Kollkian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans, "Design and construction of a realistic digital brain phantom," *IEEE Trans. on Medical Imaging*, vol. 17, pp. 463–468, 1998.
- [10] A. Valsecchi, S. Damas, J. Santamaria, and L. Marrakchi-Kacem, "Genetic algorithms for Voxel-based medical image registration," in *2013 Fourth International Workshop on Computational Intelligence in Medical Imaging (CIMI)*, apr 2013, pp. 22–29.
- [11] M. Takahashi and H. Kita, "A crossover operator using independent component analysis for real-coded genetic algorithms," *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546)*, vol. 1, pp. 643–649, 2001.
- [12] M. Lozano, F. Herrera, N. Krasnogor, and D. Molina, "Real-coded memetic algorithms with crossover hill-climbing," *Evolutionary computation*, vol. 12, no. 3, pp. 273–302, 2004.
- [13] R. He and P. A. Narayana, "Global optimization of mutual information: application to three-dimensional retrospective registration of magnetic resonance images," *Computerized Medical Imaging and Graphics*, vol. 26, no. 4, pp. 277–292, 2002.
- [14] Salcedo-Sanz, "A review on the coral reefs optimization algorithm: new development lines and current applications," *Progress in Artificial Intelligence*, vol. 6, no. 1, pp. 1–15, 2017.
- [15] O. Monga, S. Benayoun, and O. Faugeras, "From partial derivatives of 3-D density images to ridge lines," in *Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1808. Champaign, Illinois, USA: IEEE, 1992, pp. 354–359.
- [16] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, "Elastix: A toolbox for intensity-based medical image registration," *IEEE Trans. on Medical Imaging*, vol. 29, no. 1, pp. 196–205, 2010.

# Registrado de imágenes 3D-2D para identificación forense mediante comparación de radiografías

Oscar Gómez\*, Oscar Ibáñez\*, Andrea Valsecchi\*, Oscar Cordon\* y Tzipi Kahana†

\*Instituto Andaluz Interuniversitario DaSCI (Data Science and Computational Intelligence), Universidad de Granada, España

† Facultad de Medicina, Universidad Hebrea de Jerusalén, Jerusalem, Israel

Emails: ogomez@decsai.ugr.es, oscar.ibanez@decsai.ugr.es, valsecchi.andrea@gmail.com, ocordova@decsai.ugr.es, kahana.tzipi@gmail.com

**Abstract**—La comparación de radiografías es una técnica de identificación forense que conlleva la comparación manual de radiografías ante-mortem (AM) y post-mortem (PM). Sin embargo, éste es un proceso lento y propenso a errores por lo que su utilidad se ve reducida. El objetivo de este trabajo es proponer y validar un nuevo paradigma asistido por ordenador para comparación de radiografías. En concreto, se centra en reproducir de manera automática los parámetros de adquisición de la radiografía AM en una proyección del modelo 3D PM mediante una metodología basada en el registrado de imágenes 3D-2D. Este proceso de registrado será guiado por la silueta de la estructura anatómica utilizada para identificar en la radiografía AM y un modelo 3D PM de la misma. El problema de optimización subyacente es complejo ya que no se puede asumir una inicialización cercana y tampoco se puede depender del valor de intensidad de los píxeles (enfoque tradicional en imagen médica). Para abordarlo se ha utilizado una meta-heurística, differential evolution, y un método clásico que ha sido adaptado para este problema concreto. Se han realizado múltiples experimentos para estudiar la precisión y robustez de la propuesta utilizando imágenes sintéticas de clavículas, rótula y senos frontales obteniendo resultados prometedores.

**Index Terms**—Identificación forense, comparación de radiografías, registrado de imágenes 3D-2D, computación evolutiva.

## I. INTRODUCCIÓN

La identificación forense es de suma importancia en nuestra sociedad [1]. Los métodos principales de identificación son la odontología forense, el análisis de huellas dactilares y el análisis de ADN. Sin embargo, la aplicación de estos métodos no es posible cuando no se dispone de la suficiente información ante-mortem (AM) o post-mortem (PM) debido a la falta de datos por estado de conservación del cadáver y/o la ausencia de muestras AM contra las que comparar. En este aspecto, las técnicas de identificación basadas en el esqueleto (SFI, por sus siglas en inglés, *skeleton-based forensic identification*) pueden suponer la última posibilidad de identificación. Esto es debido a que el esqueleto generalmente sobrevive a procesos de descomposición natural y no natural, como por ejemplo es habitual en escenarios de desastres masivos.

Las técnicas de SFI son usadas para identificación y/o filtrado dependiendo del poder identificativo de cada hueso o cavidad [2]. Dentro de las técnicas de SFI, las más importantes son la superposición craneofacial [3], la reconstrucción facial [4] y la comparación de radiografías (CR). En concreto, CR

[1] es la única técnica SFI considerada primaria por la Interpol (junto con la comparación de implantes quirúrgicos) [5]. En la literatura, numerosas estructuras anatómicas como huesos y cavidades se han mostrado útiles para realizar una identificación positiva [6]. Éstas están localizadas principalmente en la zona del cráneo (p. ej. dientes [7] y los senos frontales [8] y del pecho (p. ej. clavículas [9] y vértebras [10]). Aunque también hay otras fuera de estas regiones tales como los huesos de la mano [11] y la rótula [12], entre muchas otras.

Sin embargo, a pesar de ser el método más empleado y útil en SFI, su aplicación todavía se basa en una comparación manual de los datos AM-PM a través de un proceso de inspección visual que depende por completo de la habilidad y experiencia del experto. Como consecuencia, su utilidad se reduce debido a que es un proceso tedioso y propenso a errores debidos al cansancio del experto.

Guiados por el análisis de la literatura forense, se han distinguido tres etapas consecutivas en proceso automático de identificación mediante CR:

- 1) Obtención de los datos AM de todos los candidatos y el procesado del material PM. Incluye la mejora y segmentación de las imágenes AM y PM.
- 2) Producir una radiografía PM que simule la transformación proyectiva de cada una de las radiografías AM del candidato.
- 3) Toma de decisiones en base a la superposición obtenida, junto con otros factores tales como la calidad de las imágenes o el poder discriminatorio del hueso utilizado.

Pese a la necesidad de métodos automáticos, esta técnica ha sido objeto de pocos trabajos de investigación. En concreto, se han utilizado técnicas de morfometría geométrica, tales como el análisis de la elíptica de Fourier, para la CR de senos frontales [13], bóvedas craneales [14], clavículas [15], y rótulas [12]. El enfoque de los dos primeros trabajos consiste en la segmentación manual del hueso o cavidad en las radiografías AM y PM y finalmente compararlas utilizando el análisis de la elíptica de Fourier. Mientras tanto, los dos últimos se basan en la adquisición de modelos 3D de los huesos PM utilizando un escáner de luz estructurada, pero la decisión final aún implica la comparación de la segmentación de la radiografía AM contra un conjunto predefinido de imágenes obtenidas a través de rotaciones pre-establecidas



del modelo 3D PM. No hay, sin embargo, ningún método automático que busque la mejor proyección posible del modelo 3D.

El proceso de superposición para CR es complejo debido principalmente a dos razones. La primera, la radiografía AM fue tomada en unas condiciones desconocidas y por tanto no se puede asumir una inicialización cercana. La segunda, las intensidades en las imágenes no son fiables o incluso hay escenarios en los que no se dispone de ellas (véase la Sección 2). Esto lleva a que las técnicas habituales de registro de imágenes (RI) 3D-2D no sean adecuadas para CR y se deben investigar técnicas diferentes para resolverlo satisfactoriamente, como la utilización de algoritmos evolutivos [16].

El objetivo de este trabajo es proponer y validar un método asistido por ordenador para realizar la superposición de un modelo 3D de cualquier hueso o cavidad PM y una radiografía AM segmentada. El método sigue una metodología de RI 3D-2D basada en la silueta del hueso o cavidad. Para validar se han utilizado dos tipos de huesos (clavículas y rótulas) y uno de cavidad (senos frontales).

Este artículo se organiza de la siguiente manera. En la sección II se hace una revisión del estado del arte de RI. La sección III describe nuestra propuesta. La sección IV presenta los experimentos realizados junto con sus resultados. Por último en la sección V se presentan las conclusiones de este trabajo.

## II. ANTECEDENTES Y JUSTIFICACIÓN

RI [17], [18] es el proceso de poner dos imágenes en un mismo sistema de coordenadas, donde una imagen se mantiene fija y la otra es transformada. En el escenario 3D-2D, la imagen transformada es siempre la 3D y la 2D se mantiene fija. Los métodos de RI 3D-2D son clasificados según quién el proceso de optimización utilizando intensidades o características de las imágenes.

En imagen médica, los métodos basados en intensidades son los más extendidos debido a que no requieren segmentaciones, las cuales introducen subjetividad y errores. Estos han sido utilizados satisfactoriamente en problemas de registro de Tomografías Computarizadas, TACs, (3D) y radiografías (2D) [19]. Sin embargo en el problema que nos ocupa (CR), los niveles de intensidad han podido cambiar desde el momento de la adquisición de la radiografía y el TAC. Esto se debe a que la densidad de los huesos puede cambiar por muchos motivos provocando un cambio en las intensidades capturadas (p. ej. edad, osteoporosis, tiempo transcurrido desde el fallecimiento, etc). Otro motivo importante es el hecho de que muchos laboratorios forenses hacen uso en la actualidad de escáneres de luz estructurada [20] debido a su gran disponibilidad y bajo coste, mientras que solo unos pocos de ellos tienen TACs. Estas limitaciones descartan los métodos basados en intensidades, dejando únicamente los basados en características. Los antropólogos forenses consideran que la morfología ósea (silueta) es un aspecto fiable para realizar la identificación basada en CR. Por otro lado, no es posible utilizar otras características como puntos homólogos por la imposibilidad de

establecer una correspondencia entre los puntos 3D y 2D. Los datos 3D son el hueso o cavidad PM en forma de un modelo de superficie PM 3D, ya sea que se obtiene al segmentar una imagen volumétrica como una CT o directamente desde un escáner 3D. Los datos 2D son la silueta del hueso en una radiografía AM. Los métodos de RI 3D-2D IR para dominios médicos [21], [22] pueden asumir que trabajan en un entorno controlado. En consecuencia pueden limitar la transformación a solo 6 grados de libertad (trasladaciones y rotaciones) sin tener en cuenta los parámetros relacionados con las distorsiones de perspectiva. Además utilizan inicializaciones cercanas al *Ground Truth* (p. ej. en [22] consideran un error máximo de 20 mm y de rotación de 20 °, de 16 mm en [21], etc). Sin embargo, estas asunciones no son adecuadas para CR. Por supuesto, hay algunas excepciones como Feldman et al. [23] que propuso un método IR 3D-2D basado en la silueta que no depende de las inicializaciones pero es solo aplicable en el caso calibrado (considerando solo 6 DoF). No existe ningún enfoque de RI 3D-2D basado en la silueta que considere 7 grados de libertad sin depender de una inicialización cercana.

## III. REGISTRADO DE IMÁGENES PARA COMPARACIÓN DE RADIOGRAFÍAS

No hay un estándar universal para diseñar un método de RI. No obstante, los métodos de RI suelen tener los siguientes 5 componentes (en la Figura 1 se puede ver un esquema de la propuesta de RI 3D-2D para CR): (1) el modelo 3D (modelo 3D del hueso PM) y una imagen fija (radiografía AM); (2) la transformación proyectiva responsable de generar una proyección 2D del modelo 3D; (3) el conocimiento experto que delimita la transformación proyectiva; (4) una métrica de similitud que mide las similitudes entre la proyección y la imagen fija; y (5) un optimizador que busca los parámetros de la transformación proyectiva según la métrica de similitud.

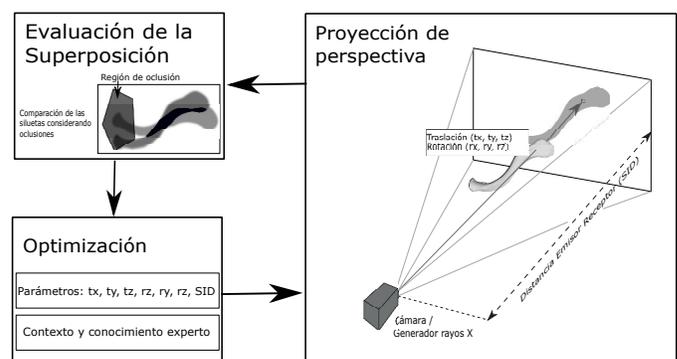


Fig. 1: Esquema de la propuesta de RI 3D-2D para CR.

### A. Datos AM y PM

Las imágenes a registrar son las siguientes:

- La imagen 2D: la radiografía AM.
- El modelo 3D: un modelo 3D de la superficie del hueso o cavidad. Obtenido mediante el uso de un escáner

de luz estructurada del "hueso limpio" o mediante la segmentación del hueso en un TAC.

En esta propuesta, el proceso de RI se guía por la silueta del hueso o cavidad por lo que es necesario su segmentación en ambas imágenes. En la radiografía AM, la silueta puede estar ocluida por otras estructuras. Por ello aparte de la silueta es necesario segmentar una región de oclusión donde el experto tiene dudas (véase la Figura 2). En el caso 3D hay dos escenarios posibles en función de si se parte de un modelo 3D de superficie obtenido con un escáner de luz estructurada o de un TAC. En el primer escenario no es necesario ningún preprocesamiento. Por otro lado, en el TAC los huesos se segmentan fácilmente mediante el uso de las unidades Hounsfield <sup>1</sup>.

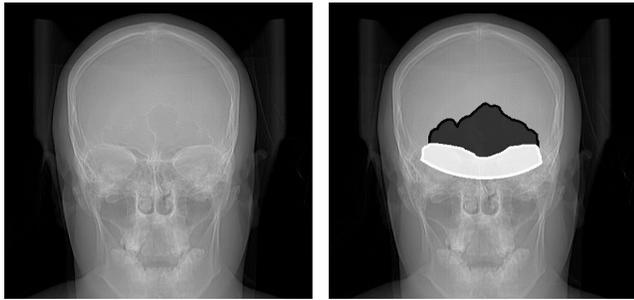


Fig. 2: Radiografía del cráneo: sin segmentar (izquierda), con la segmentación de los senos frontales en negro y la región de oclusión en blanco.

### B. Transformación proyectiva

Una transformación proyectiva es una transformación entre el mundo 3D y una imagen 2D. Las funciones proyectivas son de dos tipos: proyecciones de perspectiva y ortográficas [24].

La transformación que origina una radiografía la modelamos mediante una transformación de perspectiva simple utilizando solo 7 parámetros: 6 parámetros extrínsecos (3 traslaciones y 3 rotaciones de la cámara) y 1 intrínseco (la distancia entre el emisor y el receptor, SID, por sus siglas en inglés, *source image distance*) [25] (véase la Figura 1). La mayoría de trabajos en imagen médica consideran solo 6 parámetros dado que asumen que en un escenario controlado el parámetro SID es conocido [21], [22]. Pero en CR, cómo ya se ha comentado en las secciones anteriores, es necesario considerar los 7 parámetros.

Por otro lado, la transformación ortográfica es un caso particular de la transformación de perspectiva donde la cámara está situada en el "infinito" y por tanto no modela distorsiones de perspectiva. La mayoría de las radiografías tienen distorsiones de perspectiva (a excepción de aquellas con una SID elevada como las imágenes cefalométricas con una SID de 4 metros). No obstante, el estudio del modelo ortográfico merece la pena debido a que matemáticamente es más sencillo y podría ser suficiente para identificar. Esta transformación tiene solo 6 parámetros: 2 traslaciones que representan la posición

<sup>1</sup>Es una escala cuantitativa utilizada en los estudios de TACs para describir los diferentes niveles de radiodensidad de los tejidos humanos.

del centro de la silueta en la imagen 2D, 3 rotaciones y 1 escala que representa el porcentaje de píxeles de la imágenes ocupados por el *bounding box* (el cuadrado delimitador) de la silueta en la imagen.

### C. Parámetros y su delimitación con conocimiento experto

Los parámetros están solo delimitados por el protocolo de adquisición de radiografías [26]. Las radiografías son tomadas con el cuerpo en una posición conocida (posterioranterior o lateral) y la rotación por tanto puede ser delimitada para ambas transformaciones con un margen de error (p. ej.  $\pm 10^\circ$  o  $\pm 20^\circ$  en ángulos de Euler). En la transformación de perspectiva, el protocolo de adquisición también sirve para delimitar la traslación, la SID y también para conocer las dimensiones del receptor y su resolución (píxeles por mm). La traslación en los ejes x e y es delimitada por el ancho y alto del receptor, mientras que la SID del protocolo sirve para delimitar la SID con un rango de error y también la traslación en el eje z ya que el cuerpo se suele colocar pegado al receptor, mientras que en la transformación ortográfica la traslación es delimitada por los límites de la imagen en coordenadas normalizadas (de -1 a 1) y la escala por el porcentaje de píxeles que se espera que ocupe la silueta en la radiografía (desde el 5% al 80%), no requiriendo conocimiento experto. Por último, en el ortográfico no es necesario conocer las dimensiones del receptor ni su resolución.

### D. Métrica de similitud

Para medir las similitudes entre dos siluetas la métrica más utilizada en la literatura es la métrica *DICE* [27]. En este trabajo, la métrica *DICE* ha sido adaptada creando una nueva métrica, *Masked DICE* (véase la Eq. 1), mediante la inclusión de la región de oclusión. La región de oclusión define un área que es ignorada por la métrica permitiendo así solapamientos parciales y obviando regiones dubitativas (incertidumbre en la segmentación). En los casos en los que no haya región de oclusión el valor de la métrica *Masked DICE* será igual al de *DICE*. En el caso particular donde la proyección del hueso esté fuera del campo de visión, el valor de la métrica es fijado a 1.5.

$$\text{Masked DICE} = \frac{2 \cdot |(I_A \setminus M) \cap (I_B \setminus M)|}{|I_A \setminus M| + |I_B \setminus M|} \quad (1)$$

donde  $I_A$  son los píxeles de la silueta del objeto A (el hueso segmentado AM),  $I_B$  son los píxeles del objeto B (la proyección de hueso PM), y  $M$  es la región de oclusión.

### E. Optimizador

El problema de optimización es complejo dado que no se puede hacer uso de las intensidades o de una inicialización cercana, como ya se comentó en las secciones anteriores. Además se asume una alta multimodalidad del espacio de búsqueda. Para abordar el problema se ha estudiado el uso de dos optimizadores diferentes. El primero está basado en un método de optimización numérica llamado BOBYQA [28], que ya ha sido aplicado a problemas de RI. El segundo está basado en *Differential Evolution* (DE) [29], metaheurística que



ha mostrado un gran comportamiento en los problemas de optimización global [30].

### 1) EG-BOBYQA

Varios métodos de optimización numérica (tanto de búsqueda lineal tales como Nelder-Mead, BFGS y LBFGS, cómo de región de confianza tales como Levenberg-Marquardt y BOBYQA) han sido probados para afrontar este problema de RI utilizando la librería DLIB [31]. El que mejores resultados obtuvo en los experimentos preliminares fue BOBYQA.

Sin embargo, pese a ser el mejor, sus resultados en la experimentación preliminar fueron insuficientes debido a que mostró una gran dependencia de la inicialización, especialmente con los parámetros relacionados con la rotación. Para afrontar este problema, se ha propuesto EG-BOBYQA (*estimation grid-BOBYQA*). Este método consiste en realizar diferentes ejecuciones de BOBYQA con diferentes inicializaciones mediante el uso de un *grid* en los tres parámetros de la rotación (limitando el número de evaluaciones de BOBYQA para que sea computacionalmente abordable). Además para cada configuración del *grid*, los parámetros de la traslación y la escala son estimados en la transformación ortográfica. Hay que buscar un compromiso entre la granularidad del *grid* y el tiempo de ejecución, ya que por ejemplo un *grid* con pasos de un grado en los tres ejes de rotación es computacionalmente inabordable incluso para el menor de los rangos de rotación considerados (de  $20^\circ$  en cada eje). Este problema se resuelve utilizando varios *grid* de forma iterativa reduciendo el rango de rotación y el tamaño del paso hasta alcanzar un paso menor a un grado. El paso inicial ha sido establecido de manera experimental en un cuarto del rango de rotación y en cada iteración es dividido entre cuatro, mientras que el rango de rotación es reducido de la siguiente forma [mejor rotación -  $\frac{paso}{2}$ , mejor rotación +  $\frac{paso}{2}$ ]. Finalmente la mejor solución del último *grid* es mejorada de nuevo utilizando BOBYQA sin establecer un número límite de evaluaciones.

Este optimizador es solo aplicable a la transformación ortográfica dado que la estimación de la escala y la traslación no puede realizarse en la transformación perspectiva dado que dependen tanto de la traslación en eje z como de la SID.

### 2) Differential Evolution

DE es una metaheurística propuesta por Storn y Price [29]. Se ha aplicado con éxito a problemas de optimización multimodales debido a su gran robustez, rápida convergencia y al reducido número de parámetros a fijar [32]. Además también se ha utilizado en problemas de RI 3D-3D [33]. Los parámetros a fijar de DE son los siguientes:  $F$  (controla la tasa de mutación),  $P_c$  (controla la tasa de recombinación),  $n$  (tamaño de la población) y  $m$  (número de generaciones).

## IV. EXPERIMENTOS

Los experimentos realizados están dedicados al estudio de la precisión y robustez de los dos optimizadores en problemas sintéticos de CR con dos tipos de huesos, clavículas y rótulas, y un tipo de cavidad, senos frontales, en casos positivos, es decir, el caso AM y el PM pertenecen a la misma persona.

Todos los experimentos se han realizado en un servidor con 12 nodos, cada uno con un Intel Core i7 4930k 3.4 GHz con el sistema operativo Ubuntu 16.04.

### A. Generación del dataset

Para medir la capacidad de RI 3D-2D de las propuestas se han utilizado casos sintéticos, cada uno compuesto por un modelo de superficie 3D y una proyección de perspectiva (2D) aleatoria (dentro del rango de los parámetros). Esto permite evaluar objetivamente la calidad de una superimposición incluso en escenarios con oclusiones debido a que la proyección sin oclusiones también es conocida.

En este experimento se han utilizado datos sintéticos de clavículas, rótulas y senos frontales. En concreto, se han escaneado 10 clavículas y 10 rótulas (procedentes del laboratorio de antropología física de la Universidad de Granada) con un escáner Artec *Spider<sup>TM</sup>* 3D, y 10 senos frontales han sido segmentados manualmente de TACs (procedentes del Hospital de Castilla la Mancha) usando 3D Slicer 4.5.0-1 (véase la sección III.A). Para cada uno de los 30 modelos 3D se han generado 5 proyecciones de perspectiva aleatorias dentro de los rangos de la Tabla I. Además, el tamaño del receptor para senos frontales y rótulas es de 240 mm  $\times$  300 mm y la SID es de 1000 mm, y en clavículas de 430 mm  $\times$  350 mm y 1800 mm (resultado en imágenes de 750  $\times$  900 píxeles y de 1290  $\times$  1050, respectivamente). Para cada una de las 150 imágenes simuladas, dos imágenes adicionales han sido generadas añadiendo oclusiones en la silueta del 15% y 30% con sus respectivas zonas de incertidumbre. En total se han generado 450 casos positivos de CR con sus respectivos GTs.

TABLE I: Rango de los parámetros para la transformación de perspectiva delimitados según los protocolos de adquisición [26]

Parámetro	Hueso/Cavidad		
	Senos frontales	Rótulas	Clavículas
$t_x$ (mm)	[-125, 125]	[-125, 125]	[-210, 210]
$t_y$ (mm)	[-150, 150]		[-175, 175]
$t_z$ (mm)	[700, 1100]		[700, 1900]
$r_x, r_y, y r_z$ (grados)	[-10°, 10°], [-20°, 20°]		
SID (mm)	[9000, 1100]		[1600, 1900]

### B. Diseño experimental

Los parámetros de los optimizadores se han fijado en base a una experimentación preliminar, siendo sus mejores configuraciones las siguientes. En BOBYQA la región de confianza inicial se fija en 0.05. En DE 100 individuos, 50000 evaluaciones, una probabilidad de cruce  $P_c$  de 0.5, y un F de 0.5. Con DE, se han realizado 16 ejecuciones independientes de cada experimento, dado que DE se basa en un proceso estocástico, para poder estudiar tanto su precisión como robustez. La inicialización de cada ejecución es aleatoria en todo el rango de los parámetros. Para ambos, se ha establecido un criterio de parada cuando el error de solapamiento es inferior al 0.01% de los píxeles.

En total se han llevado a cabo 2700 experimentos que se desglosan de la siguiente manera:

- 1) Transformación ortográfica: 2 optimizadores (EG-BOBYQA y DE), 2 rangos de rotación ( $\pm 10^\circ$  y  $\pm 20^\circ$ ),

con los 450 casos simulados (véase que todos los casos fueron generados con la transformación de perspectiva). Un total de 1800 experimentos.

- 2) Transformación perspectiva: 1 optimizador (DE), 2 rangos de rotación ( $\pm 10^\circ$  y  $\pm 20^\circ$ ), con los 450 casos simulados. Un total de 1800 experimentos. Con esta transformación no se ha podido probar EG-BOBYQA por los motivos vistos en la sección III.E.

### C. Resultados

Los resultados obtenidos se muestran en la Tabla II para las métricas *Masked DICE* y mRPD (error medio de la distancia de retroproyección, para más información véase [34]). Las métricas *Masked DICE* y mRPD están correlacionadas según el test de correlación de Pearson [35] (obteniendo un coeficiente de 0.780), lo cual indica que *Masked DICE* es una buena métrica para guiar el proceso de optimización en CR. También se ve que la transformación de perspectiva obtiene siempre mejores resultados que la ortográfica, lo cual se confirma con el test de Wilcoxon [36] y el del signo [37] obteniendo unos valores p de  $2.2 \cdot 10^{-16}$  y  $8.6 \cdot 10^{-203}$ , respectivamente. Esta diferencia puede explicarse ya que solo la transformación de perspectiva puede reproducir las distorsiones de perspectiva presentes en los casos. Por último, DE obtiene siempre mejores resultados tanto en media como en desviación estándar (sd) que EG-BOBYQA, lo cual se confirma también con el test de Wilcoxon y el de signo obteniendo unos valores p de  $2.2 \cdot 10^{-16}$  y de  $9.2 \cdot 10^{-110}$ .

TABLE II: Resumen de los resultados bajo las métricas DICE y mRPD según hueso o cavidad, transformación y optimizador.

Hueso	Optimizador	Transf.	Masked DICE		mRPD (mm)	
			media	sd	media	sd
Clavícula	DE	Orto.	0.015	0.011	12.128	14.442
		Persp.	<b>0.001</b>	<b>0.003</b>	<b>0.055</b>	<b>0.088</b>
	EG-BOBYQA	Orto.	0.044	0.025	12.714	14.373
		Persp.	0.014	0.016	11.981	16.270
Rótula	DE	Orto.	0.035	0.020	12.148	16.892
		Persp.	<b>0.005</b>	<b>0.008</b>	<b>0.761</b>	<b>1.544</b>
	EG-BOBYQA	Orto.	0.014	0.055	8.471	3.571
		Persp.	<b>0.001</b>	<b>0.003</b>	<b>0.028</b>	<b>0.067</b>
Senos frontales	DE	Orto.	0.029	0.034	8.285	3.399
		Persp.	<b>0.001</b>	<b>0.003</b>	<b>0.028</b>	<b>0.067</b>
	EG-BOBYQA	Orto.	0.029	0.034	8.285	3.399
		Persp.	<b>0.001</b>	<b>0.003</b>	<b>0.028</b>	<b>0.067</b>

La robustez de ambos optimizadores es mejorable dado que tienen algunas ejecuciones con errores muy altos (un 40%<sup>2</sup> con *Masked DICE* y 60 mm con mRPD). Sin embargo, el error mínimo del DE (es decir, el mejor resultado de las 16 ejecuciones) es siempre inferior a 0.5% con *Masked DICE* y de 0.1 mm con mRPD con las tres estructuras anatómicas para la transformación de perspectiva. Mientras que para el ortográfico es de 3% y 1 mm, respectivamente. Además ni la oclusión ni el rango de rotación han mostrado una influencia significativa en la precisión, afectando ligeramente a la robustez. Por último, también puede verse que existe una gran diferencia según la estructura anatómica, siendo los senos frontales los que obtienen mejores resultados seguidos de las clavículas, y por último de las rótulas.

<sup>2</sup>La métrica DICE multiplicada por 100 es equivalente al porcentaje de píxeles no coincidentes entre dos siluetas.

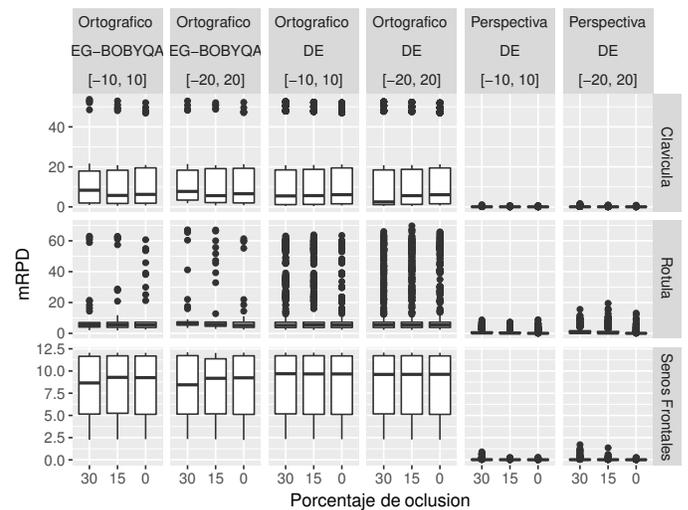


Fig. 3: Boxplots del error mínimo obtenido según el hueso o cavidad, transformación y optimizador bajo la métrica mRPD.

El principal punto débil de esta propuesta es el alto tiempo de ejecución necesario para obtener los resultados (30 minutos de media), haciendo difícil volver a ejecutar DE cuando éste obtiene una mala superimposición. Esto se debe al alto número de evaluaciones que necesita realizar el optimizador.

En resumen, DE obtiene buenas superposiciones con ambas transformaciones, los dos rangos de rotación y los diferentes niveles de oclusión, pero solo siendo robusto con los senos frontales debido a la singularidad de su silueta.

## V. CONCLUSIONES

La contribución de este trabajo ha sido abordar el problema del RI 3D-2D para CR basándose únicamente en la silueta del hueso o cavidad. Se han considerado dos optimizadores (EG-BOBYQA y DE) y dos transformaciones proyectivas (ortográfica y perspectiva). Se han obtenido resultados prometedores con un error de solapamiento inferior al 2% de los píxeles entre la silueta AM y la proyección del modelo 3D PM encontrada por el optimizador. A nivel general, DE ha ganado a EG-BOBYQA en todos los sentidos tanto en precisión como en robustez. Además la transformación de perspectiva también ha sido mejor que la ortográfica mostrando la importancia de poder reproducir las distorsiones de perspectiva. Estos resultados varían significativamente según el hueso o la cavidad. Los peores resultados han sido obtenidos con las rótulas presumiblemente por su baja singularidad con un error promedio de 0.761 mm y 1.5 % de píxeles mal superpuestos de acuerdo con las métricas mRPD y *Masked DICE*, respectivamente. Con clavículas por otro lado se han logrado errores promedio de 0.055 mm y 0.2 %, respectivamente. Mientras que los mejores resultados han sido obtenidos con los senos frontales con un error promedio de 0,028 mm y 0,2 % de píxeles mal superpuestos, respectivamente, debido a su singularidad. Estos resultados concuerdan con el uso que las rótulas y clavículas tienen en identificación limitado principalmente al filtrado de



casos [9], [12], mientras que los senos frontales han sido utilizados directamente para identificación [8].

En conclusión, se ha logrado automatizar el proceso de RI necesario para una aplicación adecuada de la técnica CR con reproducibilidad, objetividad y alta precisión. El principal inconveniente de esta propuesta es el tiempo de cálculo requerido para obtener las superposiciones (30 minutos en promedio por superposición). El primer trabajo futuro es la realización de un escenario  $n$  contra  $n$  con casos positivos y negativos que permitirá comprobar si los errores obtenidos son suficientes para identificación. Además de realizar un estudio comparativo de diferentes metaheurísticas de rápida convergencia [38] para reducir el tiempo de ejecución de la propuesta.

#### AGRADECIMIENTOS

Esta investigación ha sido subvencionada por el Ministerio de Economía y Competitividad de España bajo el proyecto NEWSOCO [TIN2015-67661-P], y por el departamento andaluz de Innovación, Ciencia y Empresa [TIC2011-7745], que incluye fondos europeo para el desarrollo de la región. El trabajo del Sr. Gómez ha sido financiado por el MECD mediante una beca FPU [FPU14/02380]. El trabajo del Dr. Ibáñez ha sido financiado por el MINECO mediante una beca Juan de la Cierva-Incorporación [JCI-2014-22433].

#### BIBLIOGRAFÍA

- [1] M. J. Thali, B. Brogdon, M. D. Viner, *Forensic radiology*, CRC Press, 2002.
- [2] M. Page, J. Taylor, M. Blenkin, Uniqueness in the forensic identification sciences—fact or fiction?, *Forensic Science International* 206 (1) (2011) 12–18.
- [3] M. I. Huete, O. Ibáñez, C. Wilkinson, T. Kahana, Past, present, and future of craniofacial superimposition: Literature and international surveys, *Legal Medicine* 17 (4) (2015) 267 – 278.
- [4] C. Wilkinson, *Forensic facial reconstruction*, Cambridge University Press, 2004.
- [5] Interpol, *Disaster victim identification guide*, Available at: <https://www.interpol.int/Media/Files/INTERPOL-Expertise/DVI/DVI-Guide-20082>.
- [6] T. Kahana, J. Hiss, Identification of human remains: forensic radiology, *Journal of Clinical Forensic Medicine* 4 (1) (1997) 7 – 15.
- [7] I. Pretty, A look at forensic dentistry - part 1: The role of teeth in the determination of human identity, *British Dental Journal* 190 (7) (2001) 359–366.
- [8] G. Quatrehomme, P. Fronty, M. Sapanet, G. Grévin, P. Baillet, A. Ollier, Identification by frontal sinus pattern in forensic anthropology, *Forensic Science International* 83 (2) (1996) 147–153.
- [9] C. Stephan, A. Winburn, A. Christensen, A. Tyrrell, Skeletal identification by radiographic comparison: Blind tests of a morphoscopic method using antemortem chest radiographs, *Journal of Forensic Sciences* 56 (2) (2011) 320–332.
- [10] T. Kahana, L. Goldin, J. Hiss, Personal identification based on radiographic vertebral features, *The American Journal of Forensic Medicine and Pathology* 23 (1) (2002) 36–41.
- [11] M. G. Koot, N. J. Sauer, T. W. Fenton, Radiographic human identification using bones of the hand: A validation study, *Journal of Forensic Sciences* 50 (2) (2005) 263–268.
- [12] E. Niespodziewanski, C. N. Stephan, P. Guyomarc'h, T. W. Fenton, Human identification via lateral patella radiographs: A validation study, *Journal of Forensic Sciences* 61 (1) (2016) 134–140.
- [13] A. M. Christensen, Testing the reliability of frontal sinuses in positive identification, *Journal of Forensic Sciences* 50 (1) (2005) 18–22.
- [14] A. B. Maxwell, A. H. Ross, A radiographic study on the utility of cranial vault outlines for positive identifications, *Journal of Forensic Sciences* 59 (2) (2014) 314–318.
- [15] C. N. Stephan, B. Amidan, H. Trease, P. Guyomarc'h, T. Pulsipher, J. E. Byrd, Morphometric comparison of clavicle outlines from 3d bone scans and 2d chest radiographs: a shortlisting tool to assist radiographic identification of human skeletons, *Journal of Forensic Sciences* 59 (2) (2014) 306–313.
- [16] S. Damas, O. Cordón, J. Santamaría, Medical image registration using evolutionary computation: An experimental survey, *IEEE Computational Intelligence Magazine* 6 (4) (2011) 26–42.
- [17] P. Markelj, D. Tomaževič, B. Likar, F. Pernuš, A review of 3D/2D registration methods for image-guided interventions, *Medical Image Analysis* 16 (3) (2012) 642–661.
- [18] F. P. Oliveira, J. M. R. Tavares, Medical image registration: a review, *Computer Methods in Biomechanics and Biomedical Engineering* 17 (2) (2014) 73–93.
- [19] P. Markelj, D. Tomazevic, F. Pernus, B. Likar, Robust gradient-based 3-D/2-D registration of CT and MR to x-ray images, *IEEE Transactions on Medical Imaging* 27 (12) (2008) 1704–1714.
- [20] S. Damas, O. Cordón, O. Ibáñez, J. Santamaría, I. Alemán, M. Botella, F. Navarro, Forensic identification by computer-aided craniofacial superimposition: a survey, *ACM Computing Surveys* 43 (4) (2011) 27.
- [21] D. B. Russakoff, T. Rohlfing, K. Mori, D. Rueckert, A. Ho, J. R. Adler, C. R. Maurer, Fast generation of digitally reconstructed radiographs using attenuation fields with application to 2D-3D image registration, *IEEE Transactions on Medical Imaging* 24 (11) (2005) 1441–1454.
- [22] J. Jomier, E. Bullitt, M. Van Horn, C. Pathak, S. R. Aylward, 3D/2D model-to-image registration applied to TIPS surgery, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2006, pp. 662–669.
- [23] J. Feldmar, N. Ayache, F. Betting, 3D-2D projective registration of free-form curves and surfaces, in: *Computer Vision, 1995. Proceedings., Fifth International Conference on, IEEE, 1995*, pp. 549–556.
- [24] R. Hartley, A. Zisserman, *Multiple view geometry in computer vision*, Cambridge University Press, 2003.
- [25] D. Mery, *X-ray Testing*, Springer International Publishing, Cham, 2015, pp. 1–33.
- [26] K. L. Bontrager, J. Lampignano, *Textbook of radiographic positioning and related anatomy*, Elsevier Health Sciences, 2013.
- [27] T. Sørensen, A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons, *Kongelige Danske Videnskabernes Selskab* 5 (1948) 1–34.
- [28] M. J. Powell, *The BOBYQA algorithm for bound constrained optimization without derivatives*, Cambridge NA Report NA2009/06, University of Cambridge, Cambridge, UK, 2009.
- [29] R. Storn, K. Price, Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces, *Journal of Global Optimization* 11 (4) (1997) 341–359.
- [30] A. K. Qin, X. Li, Differential evolution on the CEC-2013 single-objective continuous optimization testbed, in: *2013 IEEE Congress on Evolutionary Computation*, 2013, pp. 1099–1106.
- [31] D. King, *Dlib c++ library*, Available at: <http://dlib.net>.
- [32] U. Chakraborty, *Advances in differential evolution*, Vol. 143, Springer, 2008.
- [33] J. Santamaría, S. Damas, J. M. García-Torres, O. Cordón, Self-adaptive evolutionary image registration using differential evolution and artificial immune systems, *Pattern Recognition Letters* 33 (16) (2012) 2065–2070.
- [34] E. B. van de Kraats, G. P. Penney, D. Tomazevic, T. Van Walsum, W. J. Niessen, Standardized evaluation methodology for 2-D-3-D registration, *IEEE Transactions on Medical Imaging* 24 (9) (2005) 1177–1189.
- [35] K. Pearson, Note on regression and inheritance in the case of two parents, *Proceedings of the Royal Society of London* 58 (1895) 240–242.
- [36] E. A. Gehan, A generalized wilcoxon test for comparing arbitrarily singly-censored samples, *Biometrika* 52 (1-2) (1965) 203–223.
- [37] W. J. Dixon, A. M. Mood, The statistical sign test, *Journal of the American Statistical Association* 41 (236) (1946) 557–566.
- [38] Q. Chen, B. Liu, Q. Zhang, J. Liang, P. Suganthan, B. Qu, Problem definition and evaluation criteria for CEC 2015 special session and competition on bound constrained single-objective computationally expensive numerical optimization, *Computational Intelligence Laboratory, Zhengzhou University, China and Nanyang Technological University, Singapore*.

# Ajuste de un modelo de crecimiento de tumores de vejiga mediante PSO y predicción con incertidumbre

Clara Burgos-Simón\*, Juan-Carlos Cortés\*, José-Ignacio Hidalgo<sup>†</sup>, David Martínez-Rodríguez\*, Rafael-J. Villanueva\*

\*Instituto Universitario de Matemática Multidisciplinar,  
Universitat Politècnica de València, Valencia, Spain

Email: clabursi@upv.es, jccortes@upv.es, damarro3@upv.es, rjvillan@upv.es

<sup>†</sup>Departamento de Arquitectura de Computadores y Automática  
Universidad Complutense de Madrid, Madrid, Spain

Email: hidalgo@ucm.es

**Resumen**—El cáncer de vejiga es uno de los cánceres más comunes en el sistema urinario, siendo una neoplasia altamente agresiva. Generalmente, el pronóstico no es bueno y es muy difícil determinar la evolución para un paciente determinado. En este trabajo proponemos un modelo matemático dinámico que describe el crecimiento del tumor de vejiga y la evolución de la respuesta inmune. Este modelo está personalizado para un solo paciente, y se determinan los valores de los parámetros del modelo a través de una calibración que utiliza el algoritmo de optimización Particle Swarm Optimization. Debido a la incertidumbre de la evolución del tumor, utilizando los parámetros del modelo calibrado podemos predecir el tamaño del tumor y la evolución de la respuesta inmune en los próximos meses asumiendo tres escenarios diferentes: favorable, desfavorable y neutro.

**Keywords:** Cáncer de vejiga, Modelo dinámico, Cuantificación de la incertidumbre, Calibración del modelo, Particle Swarm Optimization.

## I. INTRODUCCIÓN

El cáncer de vejiga es uno de los cánceres más comunes del sistema urinario y es una neoplasia altamente agresiva [1], [2]. El 80 % de los pacientes diagnosticados con este cáncer presenta un carcinoma no invasivo que puede ser extirpado mediante una Resección Trans Uretral (RTU), que es un procedimiento endoscópico para eliminar el tumor de la parte interna de la vejiga [3]. El cáncer de vejiga se caracteriza por su recurrencia en más de la mitad de los pacientes, pues el tumor reaparece después de un tiempo y puede progresar a cáncer músculo-invasivo. Es en este estadio es cuando los tratamientos se vuelven más agresivos, y pueden incluir la extirpación de la vejiga para evitar la diseminación del cáncer a otras partes del cuerpo [4].

El pronóstico del cáncer de vejiga generalmente no es bueno y, después de la RTU, el tratamiento habitual es la instilación de bacilos Calmette-Guérin (BCG) dentro de la vejiga con la idea de que la BCG se fije a las células tumorales residuales, estimulando la respuesta inmune del paciente contra estas células infectadas para eliminarlas completamente [5].

En este trabajo proponemos un modelo matemático para estudiar la evolución del cáncer y la respuesta inmune desde el momento en el que se diagnostica. En este modelo consideraremos el protocolo terapéutico de tratamiento y lo adaptaremos a los datos disponibles que suelen utilizar los médicos para conocer el estado de la enfermedad y procurar un tratamiento adecuado y un seguimiento apropiado a cada paciente.

Una vez establecido el modelo, realizaremos un calibrado para determinar los parámetros del modelo que replican la evolución de un paciente particular. Para la calibración, utilizaremos una versión del algoritmo de optimización PSO.

Una vez calibrado el modelo, utilizaremos los parámetros para hacer una predicción a corto plazo de la evolución del tamaño del cáncer y de la respuesta inmune del paciente. Esta predicción, debido a la escasez de datos, se realizará teniendo en cuenta 3 posibles escenarios: favorable, neutro y desfavorable.

Aunque es natural en el contexto matemático-computacional realizar predicciones una vez obtenida la calibración, en el contexto médico resulta muy innovador debido al poco conocimiento que se tiene sobre la evolución de esta enfermedad en cada paciente.

## II. DATOS DISPONIBLES Y CONSTRUCCIÓN DEL MODELO

Habitualmente, un paciente va a urgencias hospitalarias o al médico de cabecera cuando orina sangre (hematuria). El paciente es derivado al servicio de urología donde, mediante ecografía o citoscopia, se confirma si hay un tumor. En caso de haberlo, empieza el seguimiento y el tratamiento. Este trabajo se ha realizado conjuntamente con doctores del Hospital Universitario y Politécnico La Fe de Valencia [6].

Los médicos han recolectado tejidos biológicos de pacientes que sufren cáncer de vejiga pT1G3, uno de los tipos llamados superficiales. El protocolo dicta la realización de una RTU para extirpar el tumor o tumores, y posteriormente administrar instilaciones de BCG, repitiendo el ciclo hasta que el paciente sane completamente (no hay más recurrencias) o reaparezca el



cáncer aumentando su tamaño y penetrando en el músculo de la vejiga de forma que haya que aplicar otros tratamientos más agresivos, entre los que puede haber incluso la extirpación de la vejiga.

Con la idea de investigar el posible interés de esta técnica para pronóstico y seguimiento de pacientes, hemos seleccionado de la base de datos del hospital los datos correspondientes a un paciente, que llamaremos paciente X, y que se resumen en la tabla I.

Fecha	Procedimiento	Tamaño (mm)	Células inflamatorias/campo
01/Mar/2012	Ecografía	3-5	-
14/Jun/2012	RTU	25	260
15/Feb/2015	Citoscopia	1-2	-
28/Abr/2015	RTU	5	515
30/Jan/2017	Citoscopia	2	-
14/Mar/2017	RTU	30-35	508

Tabla I

DATOS DEL PACIENTE X, QUE TUVO LOS PRIMEROS SÍNTOMAS EN MARZO DE 2012. DESDE ENTONCES, SE LE HAN PRACTICADO AL PACIENTE X TRES RTU.

Además de los procedimientos relacionados en la tabla I, al paciente X se le han administrado 3 tratamientos de instilaciones de BCG que empezaron los días 10/Ago/2012, 26/May/2015 y 27/Abr/2017. La dosis de las instilaciones que se administran en la vejiga consisten en 81 ml de colonias de BCG, instiladas cada semana durante un periodo de seis semanas.

Hemos adaptado el modelo introducido en [7] a los datos disponibles y a la práctica clínica y protocolos del Hospital La Fe. En una primera etapa, el tumor crece y no se administra ningún tratamiento hasta que se realiza la RTU. En esta etapa, sólo hay interacción entre las células tumorales y las inflamatorias, que constituyen la respuesta del sistema inmune global contra el tumor. Después de que en la RTU se reseccionen los tumores de la vejiga, al cabo de unas semanas empieza la administración del tratamiento de BCG. Desafortunadamente, la limpieza completa de células tumorales durante la RTU es imposible debido a la teoría de la siembra [8], por lo que las células tumorales residuales siguen creciendo. Sin embargo, con la administración de la BCG se pretende eliminar estas células residuales debido a que el sistema inmune genera una respuesta específica más agresiva y efectiva contra las células tumorales infectadas por la BCG.

La evolución de estas interacciones se modela mediante el siguiente sistema de ecuaciones en diferencias:

$$E(t+1) = E(t) - \mu_2 E(t) + \alpha T_u(t) + p_4 E(t) B(t) - p_5 E(t) T_u(t), \quad (1)$$

$$T_u(t+1) = T_u(t) - p_2 B(t) T_u(t) - p_6 E(t) T_u(t) + k T_u(t), \quad (2)$$

$$T_i(t+1) = T_i(t) - p_3 E(t) T_i(t) + p_2 B(t) T_u(t), \quad (3)$$

$$B(t+1) = B(t) - \mu_1 B(t) - p_1 E(t) B(t) - p_2 B(t) T_u(t) + b(t), \quad (4)$$

donde  $E(t)$  es el número medio de células inflamatorias por campo de visión del microscopio,  $T_u(t)$  es el diámetro

del tumor en milímetros,  $T_i(t)$  es el diámetro del tumor de células infectadas por la BCG,  $B(t)$  los mililitros de BCG en la vejiga y  $b(t)$  los mililitros de BCG instilados en una o varias dosis, en el instante temporal  $t$ .

El parámetro  $k$  describe la tasa de crecimiento del tumor,  $\alpha$  la tasa de crecimiento de las células inflamatorias,  $\mu_1$  es la tasa de mortalidad natural de la BCG y  $\mu_2$  es la tasa de mortalidad natural de las células inflamatorias.

Los parámetros que describen la interacción entre los distintos tipos de células son:  $p_1$  es la tasa de muerte de BCG debido a las células inflamatorias;  $p_2$  es la tasa de infección de las células tumorales por la BCG;  $p_3$  es la tasa de mortalidad de células tumorales infectadas, debido a la BCG;  $p_4$  es el crecimiento de células tumorales por la acción de la BCG;  $p_5$  es la tasa de mortalidad de células inflamatorias por su interacción con células tumorales no infectadas; y  $p_6$  representa la tasa de mortalidad de las células tumorales por las células inflamatorias.

### III. CALIBRACIÓN

Para encontrar los parámetros del modelo que hagan que las salidas del modelo estén lo más cerca posible a los datos de la tabla I en los instantes de tiempo correspondientes, definimos la siguiente función de ajuste  $F$ :

**Input.** Valor de los parámetros a sustituir en el modelo ( $k, \alpha, \mu_1, \mu_2, p_1, p_2, p_3, p_4, p_5, p_6$ ).

**Paso1.** Sustituir los valores de los parámetros en el modelo (1)-(4).

**Paso2.** Ejecutar el modelo y extraer los valores devueltos por el modelo para el tamaño del tumor y para el número de células inflamatorias por campo en los mismos instantes temporales en los que tenemos datos en la tabla I.

**Paso3.** Calcular el error cuadrático medio entre las salidas del modelo del Paso 2 y los datos de la tabla I.

La calibración del modelo se ha hecho en etapas porque el protocolo no es un proceso continuo debido a la extracción brusca del tumor mediante la RTU. Además, no todos los parámetros se consideran constantes en el tiempo. La respuesta inmune relacionada con la tasa de crecimiento  $p_4$  y la tasa de eficacia  $p_6$ , se va a considerar variable en el tiempo.

La primera etapa será desde 01/Mar/2012 hasta la primera RTU. La segunda etapa es hasta la segunda RTU y la tercera hasta la tercera RTU.

Vamos a minimizar la función  $F$  en cada etapa, determinando los valores de los parámetros del modelo utilizado el algoritmo rPSO [9]. Se trata de un algoritmo bio-inspirado basado en el intercambio de información entre las diferentes partículas que forman el enjambre. La función objetivo (en nuestro caso, la función a minimizar) es evaluada para cada partícula, donde las partículas son vectores formados por los valores de los parámetros del modelo. En cada iteración, actualizamos los valores de la posición de las partículas dentro del espacio de parámetros mediante las siguientes expresiones:

$$pos_{i+1} = pos_i + vel_i, \quad (5)$$

$$vel_{i+1} = U(0,1) * w * vel_i + U(0,1) * c_1 * (p_{i,mejor} - pos_i) + U(0,1) * c_2 * (g_{mejor} - pos_i). \quad (6)$$

donde:

- $U(0,1)$  es una muestra de una distribución uniforme entre los valores  $[0, 1]$ . Este valor es muestreado de forma independiente para cada meta-parámetro dentro de la función, para cada partícula dentro del enjambre y para cada iteración dentro del proceso.
- $p_{i,mejor}$  es la posición de la mejor evaluación de la función para cada partícula individual del enjambre (mejor local).
- $g_{mejor}$  es la posición de la mejor evaluación de la función realizada por cualquiera de las partículas que forman el enjambre durante todo el proceso (mejor global).

Junto con las ecuaciones descritas, al proceso se le añade una componente de mutación para evitar que las partículas converjan en un mínimo local en lugar del global. Este proceso se realiza de la siguiente manera:

- una partícula se rechaza y se genera una completamente nueva con un 10 % de probabilidad;
- si la partícula no ha sido rechazada, con un 10 % de probabilidad se le aplica una mutación.

Se han realizado diversas pruebas con números diferentes de partículas, siendo el número que mejores resultados nos ha dado el de 60 partículas. También se han realizado diversas pruebas con el valor de los meta-parámetros  $w$ ,  $c_1$  y  $c_2$ , siendo los valores finalmente utilizados  $w=0.73$ ,  $c_1=1.5$  y  $c_2=1.5$ .

#### IV. RESULTADOS

Los parámetros del modelo obtenidos en la calibración son los que se muestran en la tabla II.

Parámetro	Valor	Parámetro	Valor
$k$	0.01563	$T_u(RTU_1)$	0,02
$\alpha$	0.005	$p_4(RTU_1)$	1e-4
$\mu_1$	0.8	$p_6(RTU_1)$	1.62e-5
$\mu_2$	0.0005	$T_u(RTU_2)$	0.02
$p_1$	8e-5	$p_4(RTU_2)$	5.6e-4
$p_2$	1e-3	$p_6(RTU_2)$	6.3e-6
$p_3$	1e-4		
$p_4$	0.0		
$p_5$	0.0		
$p_6$	0.0		
$T_u(0)$	5.0		
$E(0)$	260		

Tabla II

VALORES DE LOS PARÁMETROS DEL MODELO CALIBRADOS PARA EL PACIENTE X. NÓTESE QUE LOS PARÁMETROS  $p_4$  Y  $p_6$  CAMBIAN CON EN CADA ETAPA. LOS VALORES QUE ACOMPAÑAN A  $RTU_1$  Y  $RTU_2$  CORRESPONDEN A PARÁMETROS VARIABLES TRAS LAS  $RTU_1$  Y  $RTU_2$ , RESPECTIVAMENTE.

En la figura 1, se puede ver el resultado de la calibración, en la de arriba la respuesta inmune y en la de abajo el tamaño del tumor. Las subidas bruscas en forma de dientes de sierra en la gráfica de la respuesta inmune son debidas a las dosis semanales de instilaciones de BCG. Las caídas en la gráfica del tamaño del tumor, corresponden a cada RTU.

Debido a la incertidumbre en la evolución futura, con el fin de predecir la dinámica del cáncer de vejiga, se han considerado tres escenarios diferentes: favorable, neutral y desfavorable. En el escenario favorable la respuesta del sistema inmune es mayor que la segunda etapa. El escenario neutral

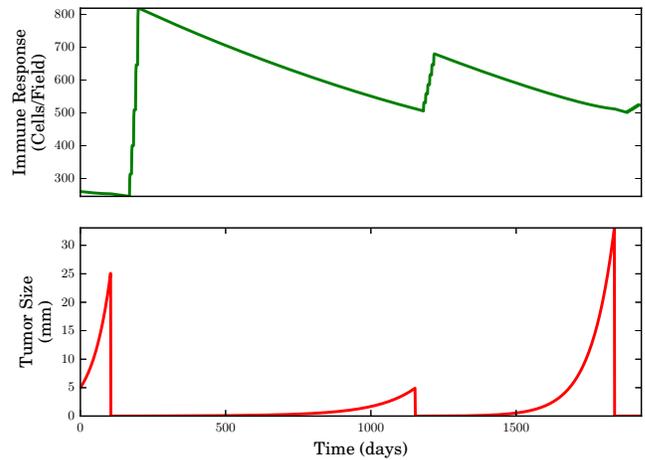


Figura 1. Gráficas con la evolución de la respuesta inmune (arriba) y del tamaño del tumor (abajo) con los parámetros obtenidos en la calibración del modelo.  $t = 0$  corresponde al 01/Mar/2012.

consiste en la misma respuesta del sistema inmune como en la tercera etapa, es decir, la respuesta del sistema inmunitario permanece constante. El escenario desfavorable se basa en el hecho de que la respuesta del sistema inmunitario disminuye en la misma proporción que disminuyó desde la segunda a la tercera fase.

En las figuras 2, 3, 4, se muestran los diferentes escenarios. La predicción empieza desde la línea negra vertical.

En el escenario favorable, el tratamiento con BCG tiene un resultado positivo y el tumor ya no crece más. En el escenario neutral, se estima que ocurrirá una recurrencia del cáncer con un tumor del tamaño de 5 mm que se prevé para el 15 de agosto de 2018. En el escenario desfavorable, la recurrencia del cáncer de vejiga con un tumor del tamaño de 5 mm se prevé para el 25 de mayo de 2018.

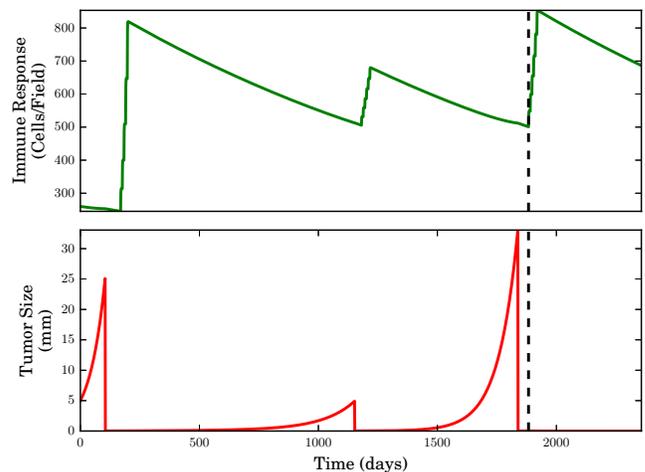


Figura 2. Escenario favorable. La BCG acaba con el tumor.

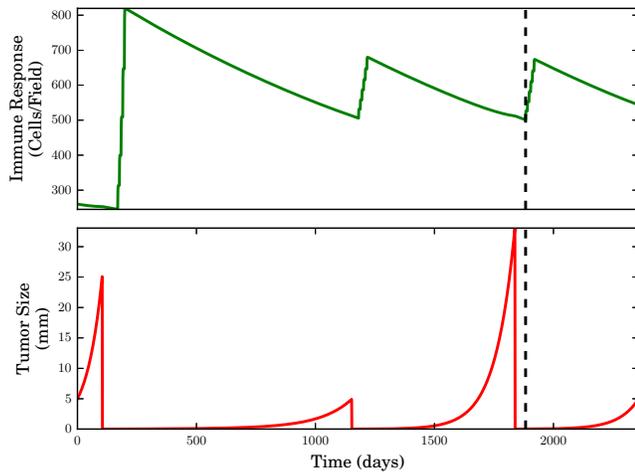


Figura 3. Escenario neutro. Se espera un tumor de 5mm para el 15 de agosto de 2018.

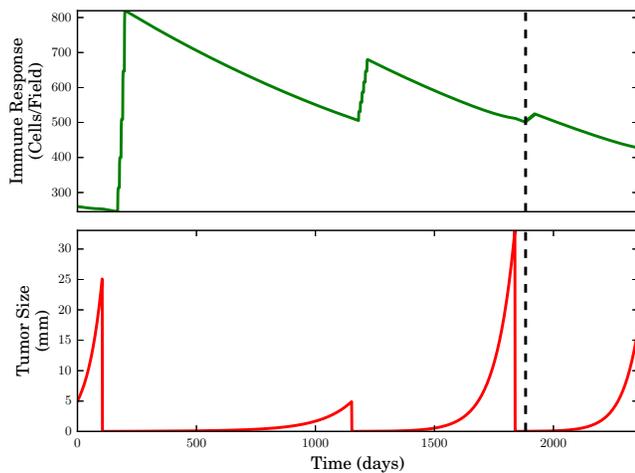


Figura 4. Escenario desfavorable. Se espera un tumor de 5mm para el 25 de mayo de 2018.

## V. CONCLUSIONES

En este trabajo presentamos un modelo dinámico para describir la evolución de un tumor de la vejiga en un paciente. El modelo se ha calibrado en varias etapas y los parámetros obtenidos nos permiten dar un pronóstico sobre el crecimiento del tumor en los próximos meses en tres escenarios: favorable, desfavorable y neutro. Que nosotros sepamos, éste es el primer acercamiento de este tipo. Los resultados parecen prometedores y ha creado expectación en el equipo médico con el que colaboramos debido al desconocimiento de la evolución de este tipo de cáncer en cada paciente.

Cuando tengamos más datos disponibles, seremos capaces de verificar la validez de nuestro enfoque, de realizar predicciones personalizadas a cada paciente más precisas y mejores, dando herramientas a los médicos para administrar tratamientos, para programar las revisiones de seguimiento y, en el futuro, para determinar mejores estrategias para mejorar la salud del paciente, su calidad de vida y evitar las recidivas

del tumor.

## VI. AGRADECIMIENTOS

Este trabajo ha sido parcialmente financiado por el proyecto del Ministerio de Economía y Competitividad MTM2017-89664-P y GenObIA-CM con referencia (S2017/BMD-3773) financiado por la Comunidad de Madrid y cofinanciado con Fondos Estructurales de la Unión Europea y Ministerio de Economía y Competitividad TIN2014-54806-R (HERACLES).

## REFERENCIAS

- [1] Official Site for Spanish Medic Oncology Society, <https://www.seom.org>, accessed: 08/05/2018.
- [2] R. T. Greenlee, M. B. Hill-Harmon, T. Murray, M. Thun, *Cancer statistics, 2001*, CA: A Cancer Journal for Clinicians 51 (1) (2001) 15–36. doi: 10.3322/canjclin.51.1.15. URL <https://doi.org/10.3322/canjclin.51.1.15>
- [3] R. Devere White, E. Stapp, *Predicting prognosis in patients with superficial bladder cancer*, ONCOLOGY 12 (12) (1998) 1717–1723. URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0005479569&partnerID=40&md5=a865e77b1a2c890bf2985b1517d7e3b6>
- [4] S. Holmang, H. Hedelin, C. Anderstrom, S. L. Johansson, *The relationship among multiple recurrences, progression and prognosis of patients with stages TA and T1 transitional cell cancer of the bladder followed for at least 20 years*, The Journal of Urology 153 (6) (1995) 1823–1827. doi:10.1016/s0022-5347(01)67321-x. URL [https://doi.org/10.1016/s0022-5347\(01\)67321-x](https://doi.org/10.1016/s0022-5347(01)67321-x)
- [5] G. Redelman-Sidi, M. S. Glickman, B. H. Bochner, *The mechanism of action of BCG therapy for bladder cancer—a current perspective*, Nature Reviews Urology 11 (3) (2014) 153–162. doi:10.1038/nrurol.2014.15. URL <https://doi.org/10.1038/nrurol.2014.15>
- [6] Official Site for Hospital Universitari i Politècnic La Fe, Valencia, Spain, <http://www.hospital-lafe.com>, accessed: 07/05/2018.
- [7] S. Bunimovich-Mendrazitsky, E. Shochat, L. Stone, *Mathematical model of BCG immunotherapy in superficial bladder cancer*, Bulletin of Mathematical Biology 69 (6) (2007) 1847–1870. doi:10.1007/s11538-007-9195-z. URL <https://doi.org/10.1007/s11538-007-9195-z>
- [8] A. Wein, L. Kavoussi, A. Novick, A. Partin, C. Peters, *Campbell-Walsh Urology*, Campbell-Walsh Urology, Elsevier Health Sciences, 2011. URL <https://books.google.es/books?id=W1aeyJD46kIC>
- [9] C. Jacob, N. Khemka, *Particle swarm optimization in Mathematica. an exploration kit for evolutionary optimization*, IMS'04, Proc. Sixth International Mathematica Symposium, Banff, Canada.

# On the Use of Fitness Sharing in Studying the Genetic Code Optimality

José Santos

Department of Computer Science,  
University of A Coruña, A Coruña, Spain  
Email: jose.santos@udc.es

Ángel Monteagudo

Department of Computer Science,  
University of A Coruña, A Coruña, Spain  
Email: angel.minsua@udc.es

**Abstract**—Since the canonical genetic code is not universal several theories arose to explain the evolution to its present form. Different computational methods were applied to analyze the optimality level of the canonical code organization, including our previous works using evolutionary computing in the problem. We discuss here the possibilities that the use of the classical fitness sharing technique provides for obtaining knowledge about the fitness landscape involved in the optimization of the genetic code.

## I. INTRODUCTION

The canonical genetic code (CGC), present in most superior organisms, establishes the association between codons and amino acids. In the canonical code there are 64 codons of three bases that encode 20 amino acids. Thus, the genetic code is redundant, since several codons codify the same amino acid. However, the canonical code is not universal, since there are other different associations between codons and amino acids. Mitochondrial DNA is an exception example. The exceptions show that the code could evolve in its origin, and it is an open question how the code evolved to the present form (the CGC).

Since other codes could appear, several models of hypothetical genetic codes were considered, with different associations between codons and amino acids. Considering only the present form of a genetic code with codons of three bases, if the codon set of the CGC is maintained, allowing only swaps of amino acids between the 20 codon sets, the possible codes are  $20!$  ( $2.43 \cdot 10^{18}$ ). Without that restriction, in the sense that every codon can codify every amino acid, the number of alternative codes is really huge, larger than  $1.51 \cdot 10^{84}$ .

In this research there are basically three theories about the reasons regarding the evolution of the genetic code: i) The stereochemical theory establishes that codon assignments are dictated by physicochemical affinity between amino acids and the cognate codons, ii) the co-evolution theory states that a set of precursor amino acids passed part or all of their codon domain to the biosynthetically produced amino acids, and iii) the physicochemical or error minimization theory states that the main factor in the code's evolution was the minimization of the adverse effects of mutations.

In this last error minimization theory there are two different alternatives to assess the level of optimization of the CGC:

1. The statistical approach considers random alternative codes and compares the CGC against the average optimization

quality of those codes [1]. The results of the works in this alternative tend to indicate that the CGC is quite better optimized with respect to random codes.

2. The engineering approach, which compares the CGC against the average random codes and with respect to better codes (typically obtained with local search algorithms) [2]. This alternative shows that the CGC has a certain level of optimization, but is still far from optimal.

In these alternatives, the “quality” of a hypothetical code is defined taking into account the consequences of all possible mutations in the codons. A mutation in a codon base can change the amino acid codified, which can take a serious consequence on the resultant protein. Usually, it is considered a property (or group of properties) of the codified amino acid, like the polar requirement (hydrophobicity), since it is the most important one in defining the folding of the protein. The quality is defined as the Mean Square (MS) of the difference of the property values of the corresponding amino acids before and after the mutation. This is averaged for all base mutations of all codons. The best code would be the one that minimizes MS, that is, the consequences of the mutations (with the lowest consequence on the phenotype, i.e., the proteins).

We introduced the possibility of evolutionary computation to discover better genetic codes [3], with a classical genetic algorithm (GA) (with *ad hoc* operators depending on the genetic code model of alternative codes), studying the level of optimality reached by the CGC in the huge landscapes considered, and taking into account aspects like the different mutation probability of transversion and transition mutations in the codon bases, or the biases in mutation probability in the three codon bases. In [4] we extended the analysis using a model of possible genetic codes that considers the known codon reassignments. Depending on the genetic code model and those commented aspects, the optimization level of the CGC varies, but in all cases is far from optimal, and in agreement with the engineering approach.

Since there is a discussion about whether the CGC is located in a local minimum (or close to a minimum) in the quality landscape and whether the landscape has a multimodal nature, in a recent work [5] (which is summarized here) we introduced the use of the classical Fitness Sharing (FS) technique [6] in the evolutionary algorithm to inspect that possible multimodal nature. FS allows the division of the population into different

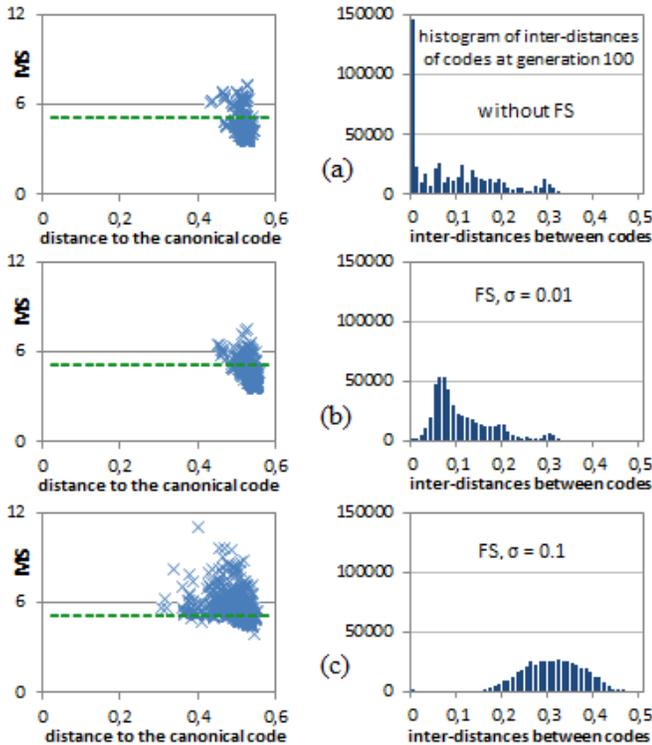


Fig. 1. Left: MS value of the hypothetical codes of the final genetic population vs. their distances to the canonical code and using restrictive codes. Right: Histograms of interdistances of final evolved codes.

subgroups according to the similarity of the individuals, groups that perform a simultaneous search in the best found promising areas (global or local minima) of the search landscape.

The use of fitness sharing serves for two purposes: i) to have an indirect view of the huge search landscape and to locate the CGC in relation to the best possible codes that could be obtained. ii) Given the nature of the landscape obtained in i), to determine the difficulty of evolution of hypothetical genetic codes in such a landscape.

## II. RESULTS WITH A RESTRICTIVE CODE MODEL

An example of a run of the evolutionary algorithm is presented here. All the details about the GA implementation (genetic operators, encoding of hypothetical genetic codes) are in [5]. We present results of an example run with and without the use of fitness sharing and with a restrictive model of hypothetical codes. In this model, the pattern of synonymous coding found with the CGC is maintained, that is, the 21 CGC non-overlapping sets of codons are fixed (20 sets correspond to the amino acids and one set to the 3 stop codons). The 20 amino acids are randomly assigned to one of the 20 sets while the same 3 stop codons are fixed as in the CGC.

Figure 1 corresponds to a run of the GA without FS and with FS with two values of the parameter “sharing radius” ( $\sigma_{share} = 0.01$  and  $\sigma_{share} = 0.1$ ), which controls the extent of sharing [6]. In Figure 1, subfigures at the left, the  $x$ -axis corresponds to the distance of each encoded code to the CGC whereas the  $y$ -axis corresponds to the code MS value. These graphs correspond to the final populations at generation 100

in 3 cases: (a) without FS, (b) FS with  $\sigma_{share} = 0.01$  and (c) FS with  $\sigma_{share} = 0.1$ . The figures at the right correspond to the histograms of interdistances between the encoded codes at that final generation. The distance  $d_{ij}$  between two codes  $i$  and  $j$  is measured taking into account the difference in polar requirement between the amino acids encoded in the same positions by both population codes, normalized in  $[0,1]$ .

The distances of the population (1000 individuals) to the canonical code vary in a range between 0.3 and 0.6. Many of the evolved codes are better adapted than the canonical code (the dashed line represents the MS value of the CGC), showing that these better codes are far from the CGC. There is not any clustering of individuals, independently of the value of  $\sigma_{share}$ , which indicates that there are not deep local minima and that the CGC is not located in an area of a deep local minimum.

Moreover, the CGC was introduced in the initial population in the 3 cases. In all runs the CGC disappears from the population in few generations (as shown in [5]), even with the use of fitness sharing, which is another evidence that the CGC is not located in an area corresponding to a deep local minimum. The interdistances (Figure 1, right part) show that, without FS, the individuals are close to the best solution, whereas the solutions are more spread in the search landscape with larger values of  $\sigma_{share}$  when FS is used. Since there are not interdistance values close to 0 in the histogram (i.e., no clusters), both graphs (with FS) indicate that the landscape does not present clear localized areas of deep local peaks, as well as that the CGC is not located in one of such areas.

## III. CONCLUSION

The main conclusion to be drawn from the results is that the fitness landscape, although is clearly rugged, does not have a multimodal nature with deep localized areas of low MS values and separated by barriers of high MS values. Therefore, the fitness landscape considered in the error minimization theory does not explain how the canonical code ended its evolution in an area that does not correspond to a deep local minimum.

## ACKNOWLEDGMENT

This work was funded by Xunta de Galicia (project GPC ED431B 2016/035), Xunta de Galicia (“Centro singular de investigación de Galicia” accreditation 2016-2019 ED431G/01) and the European Regional Development Fund (ERDF).

## REFERENCES

- [1] S. Freeland, R. Knight, and L. Landweber, “Measuring adaptation within the genetic code,” *Trends in Biochemical Sci.*, vol. 25(2), pp. 44–45, 2000.
- [2] M. Di Giulio, “The origin of the genetic code,” *Trends in Biochemical Sciences*, vol. 25(2), p. 44, 2000.
- [3] J. Santos and A. Monteagudo, “Study of the genetic code adaptability by means of a genetic algorithm,” *Journal of Theoretical Biology*, vol. 264(3), pp. 854–865, 2010.
- [4] J. Santos and A. Monteagudo, “Simulated evolution applied to study the genetic code optimality using a model of codon reassignments,” *BMC Bioinformatics*, vol. 12:56, 2011.
- [5] J. Santos and A. Monteagudo, “Inclusion of the fitness sharing technique in an evolutionary algorithm to analyze the fitness landscape of the genetic code adaptability,” *BMC Bioinformatics*, vol. 18:195, 2017.
- [6] D. Goldberg and J. Richardson, “Genetic algorithms with sharing for multimodal function optimization,” *Proceedings 2nd International Conference on Genetic Algorithms*, pp. 41–49, 1987.