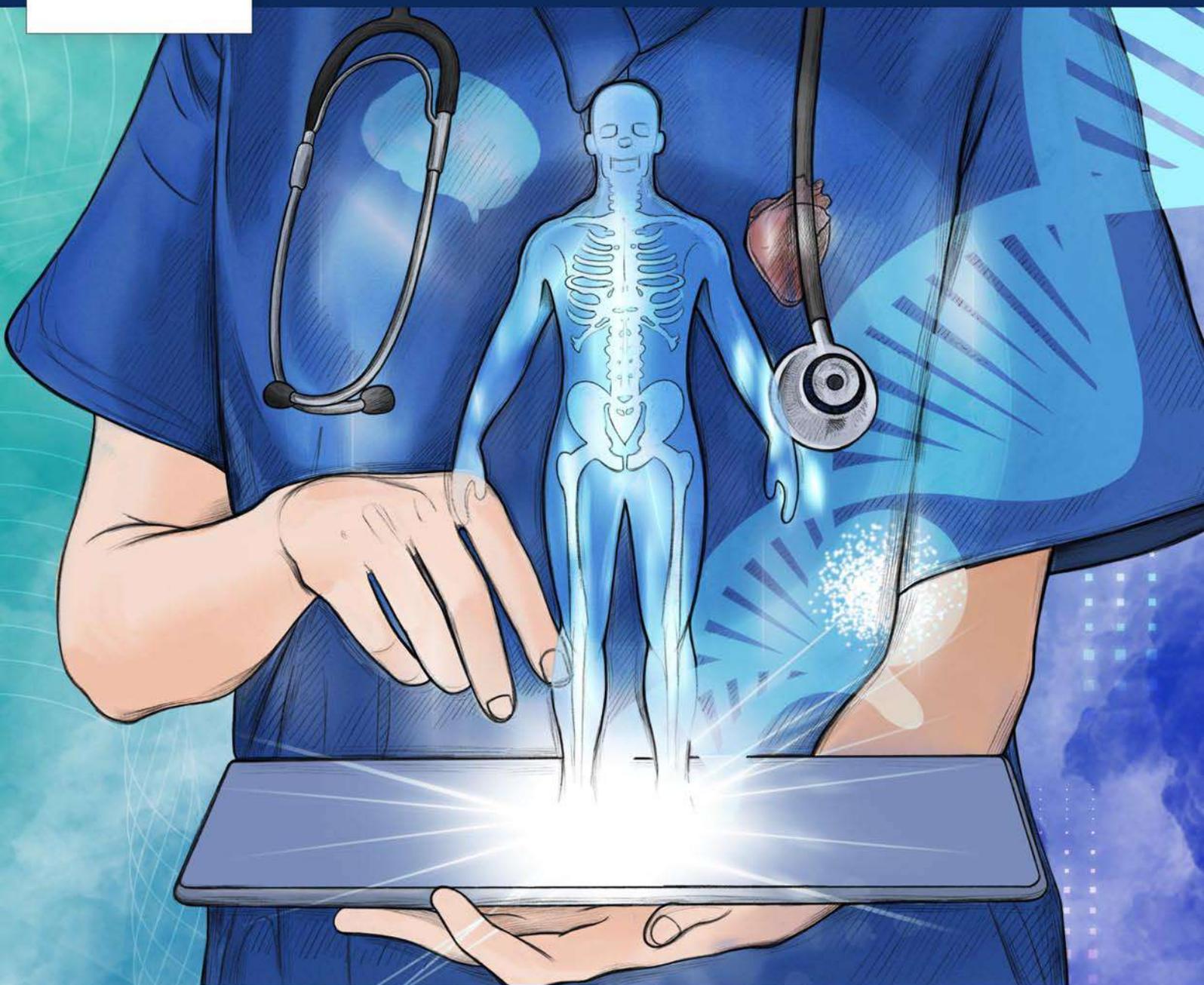




CASEIB
2021
XXXIX

XXXIX CONGRESO ANUAL DE LA SOCIEDAD ESPAÑOLA DE INGENIERÍA BIOMÉDICA



LIBRO DE ACTAS

ORGANIZAN:



UNIVERSIDAD
POLITÉCNICA
DE MADRID





CASEIB
2021
XXXIX

Congreso Anual
de la Sociedad
Española de
Ingeniería
Biomédica

LIBRO DE ACTAS

Madrid, 25 y 26 de noviembre de 2021

XXXIX Congreso Anual de la Sociedad Española de Ingeniería Biomédica

Presidente del Comité Organizador:

Enrique J. Gómez Aguilera

Secretaria del Comité Organizador:

Patricia Sánchez González

Editores:

Paloma Chausa Fernández

Gema García Sáez

Enrique J. Gómez Aguilera

M. Elena Hernando Pérez

Patricia Sánchez González

Diseño de portada:

Sandra Martínez

ISBN: 978-84-09-36054-3



Libro de Actas del XXXIX Congreso Anual de la Sociedad Española de Ingeniería Biomédica
Se distribuye bajo una licencia de Creative Commons Reconocimiento-NoComercial-Sin Obra
Derivada 4.0 Internacional.

ORGANIZADORES

Grupo de Bioingeniería y Telemedicina
Escuela Técnica Superior de Ingenieros de Telecomunicación
Universidad Politécnica de Madrid

Sociedad Española de Ingeniería Biomédica



SOCIEDADES INTERNACIONALES QUE AUSPICIAN EL CONGRESO

European Alliance for Medical and Biological Engineering & Science

International Federation for Medical and Biological Engineering



COLABORADORES



EMPRESAS COLABORADORAS DE LA SEIB



Medtronic

BIENVENIDA

Estimados compañeros y amigos de la Ingeniería Biomédica participantes en CASEIB2021:

En nombre del Comité Organizador, es un placer presentar el XXXIX Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2021). La nueva edición de nuestro congreso tiene un formato virtual, tal como hicimos el año pasado, pero en un formato más reducido que el congreso tradicional y con una serie de nuevas propuestas que esperamos sean de interés.

El congreso se celebra los días 25 y 26 de noviembre de 2021, y está organizado por el Grupo de Bioingeniería y Telemedicina de la Universidad Politécnica de Madrid y la Sociedad Española de Ingeniería Biomédica (SEIB), en colaboración con la Federación de Empresas de Tecnología Sanitaria (FENIN) y el nodo español del European Institute of Innovation & Technology in Health (EIT Health Spain).

El objetivo principal de esta nueva edición de CASEIB 2021 es potenciar aún más la participación de los estudiantes de Ingeniería Biomédica, tanto de grado, máster y doctorado. CASEIB 2021 tiene un enfoque muy dirigido hacia la innovación y el emprendimiento como áreas fundamentales en nuestra disciplina, en línea con el lema propuesto: “La Ingeniería Biomédica en la transformación digital de la sanidad”.

El primer día del congreso, organizado conjuntamente con FENIN, se centra en conocer y analizar la visión de la industria de la tecnología sanitaria sobre el papel presente y futuro de las nuevas generaciones de ingenieros biomédicos en la transformación digital de la sanidad. En la agenda de este primer día se incluyen las sesiones de las competiciones de los estudiantes de grado de los renombrados “Premios FENIN - Competición de estudiantes de Grado en Ingeniería Biomédica” y los “Premios Prof. José M^a Ferrero Corral” para estudiantes de máster y doctorado. El segundo día, organizado conjuntamente con EIT Health Spain, se desarrolla en inglés, y se centra en la innovación y emprendimiento en tecnologías sanitarias. Incluye una sesión dedicada al nuevo “Premio EIT Health Spain a la innovación y emprendimiento en Ingeniería Biomédica” para estudiantes de grado, máster y doctorado. Gracias a la fructífera colaboración establecida entre SEIB, FENIN y EIT Health Spain, se otorgan tres premios dotados económicamente a los mejores trabajos de las 3 competiciones del congreso.

CASEIB 2021, es una edición especial para todos los estudiantes de ingeniería biomédica, y en esta edición no se ha realizado la llamada habitual de trabajos ni la celebración de sesiones científicas de presentación de trabajos, con la excepción de las sesiones de finalistas de las distintas competiciones.

En este libro de actas se han incluido todos los trabajos presentados en los “Premios FENIN - Competición de estudiantes de Grado en Ingeniería Biomédica” (seleccionados por cada uno de los Grados), así como los trabajos aceptados por el Comité Científico para su participación en los “Premios Prof. José M^a. Ferrero Corral”, con independencia de su selección como finalistas. Finalmente, los trabajos presentados a este último, y aceptados por el Comité Científico, podrán optar, si lo desean, a ser valorados para su publicación en ediciones especiales de las revistas “Sensors” y “Entropy”, con las condiciones establecidas en congresos anteriores.

Nuestros sinceros agradecimientos a FENIN, EIT Health Spain, Universidad Politécnica de Madrid, y a todas las empresas colaboradoras de la SEIB que nos han brindado su apoyo en el congreso.

Con la ilusión de volver al formato habitual del congreso CASEIB en el próximo año 2022, y en el convencimiento que esta edición especial siga siendo el gran foro de conocimiento en Ingeniería Biomédica en España, os enviamos un afectuoso saludo.

A handwritten signature in blue ink, appearing to read 'Enrique J. Gómez Aguilera', with a long horizontal line extending to the right.

Enrique J. Gómez Aguilera

Presidente del Comité Organizador del CASEIB2021

COMITÉ ORGANIZADOR

Presidente

Enrique J. Gómez Aguilera

Secretaria

Patricia Sánchez González

Vocales GBT-SEIB

Paloma Chausa Fernández

José Manuel Iniesta Chamorro

Gema García Sáez

Ignacio Oropesa García

M. Elena Hernando Pérez

Julio Ignacio Piquero Cabañero

Vocales FENIN

Sergio Muñoz

Jorge Fuentes

Sofía Romero

Vocales EIT Health

Montse Delgado

Lilisha Burris

Mara Belluco

COMITÉ CIENTÍFICO

Presidente

Enrique J. Gómez
Aguilera

*Universidad Politécnica
de Madrid*

Secretaria

M. Elena Hernando
Pérez

*Universidad Politécnica
de Madrid*

Co-Secretaria

Patricia Sánchez
González

*Universidad Politécnica
de Madrid*

Vocales

Begoña Acha Piñero. *Universidad de
Sevilla*

Roberto Hornero Sánchez. *Universidad
de Valladolid*

Raúl Alcaraz Martínez. *Universidad de
Castilla-La Mancha*

Raimon Jané Campos. *Universidad
Politécnica de Cataluña*

Erik Alonso. *Universidad del País Vasco*

Pablo Laguna Lasasa. *Universidad de
Zaragoza*

Elisabete Aramendi Ecenarro.
Universidad del País Vasco

María Jesús Ledesma. *Universidad
Politécnica de Madrid*

José Luis Arce Diego. *Universidad de
Cantabria*

Maite Mujika Garmendia. *CEIT-TECNUN,
San Sebastián*

M^a Gloria Bueno García. *Universidad de
Castilla-La Mancha*

Jesús Poza Crespo. *Universidad de
Valladolid*

Manuel Desco Menéndez. *Universidad
Carlos III de Madrid*

Gema Prats Boluda. *Universidad
Politécnica de Valencia*

Félix Fanjul Vélez. *Universidad de
Cantabria*

Javier Reina Tosina. *Universidad de
Sevilla*

Javier García Casado. *Universidad
Politécnica de Valencia*

Ana de los Reyes Guzmán. *Servicio de
Salud de Castilla-La Mancha*

María García Gadañón. *Universidad de
Valladolid*

José Joaquín Rieta Ibáñez. *Universidad
Politécnica de Valencia*

Beatriz F. Giraldo Giraldo. *Universidad
Politécnica de Cataluña*

Isabel Román Martínez. *Universidad de
Sevilla*

Francisco Javier Sáiz Rodríguez.
Universidad Politécnica de Valencia

Francisco M. Sánchez Margallo. *C.C.M.I.*
Jesús Usón

Daniel Sánchez Morillo. *Universidad de*
Cádiz

Andrés Santos Lleó. *Universidad*
Politécnica de Madrid

Carmen Serrano Gotarredona.
Universidad de Sevilla

Abel Torres Cebrián. *Universidad*
Politécnica de Cataluña

PROGRAMA - JUEVES 25 DE NOVIEMBRE DE 2021

9:00 **Acto Inaugural**

- *- D. Guillermo Cisneros, Rector Magnífico de la UPM*

9:30 *- Dña. Margarita Alfonso, Secretaria General de FENIN*

- Dña. Cristina Bescós, Directora General de EIT Health Spain

- D. Enrique J. Gómez, Presidente CASEIB2021 y presidente SEIB

9:30 **Charla: «La Transformación Digital en la Sanidad»**

- Ponente:

10:30 *- D. Luis Javier Bonilla, Director General de la compañía CGM en España, multinacional del sector de salud digital*

Moderadora: Dña. M^a Elena Hernando Pérez, Catedrática Universidad Politécnica de Madrid

10:30 **Mesa redonda: «Ingeniería Biomédica y Transformación Digital»**

11:30 Ponentes:

- D. Javier Arribas Pastor, Digital Solutions Expert en Siemens Health, Siemens-healthineers

- Dña. Carmen Carrillo, Directora de ID&D, Matachana

- D. Lluís Blanch, Coordinador ITEMAS, Director Institut d'Investigació i Innovació Parc Taulí - I3PT

- D. Javier Reina, Catedrático Univ. Sevilla, Secretario General de SEIB

Moderador: D. Sergio Muñoz, Director Innovación de FENIN

11:30 *Descanso*

-

11:45

11:45 **Premios FENIN para alumnos GIB**

- Miembros del jurado:

13:30 - *D. Roberto Hornero*

- *D. Javier Reina*

- *D. Juan Carlos Antony*

Moderador: *D. Roberto Hornero*

13:30 *Descanso*

-

15:00

15:00 **Premio José María Ferrero Corral**

- Miembros del jurado:

18:00 - *Dña. Elisabete Aramendi*

- *Dña. Beatriz Giraldo*

- *Dña. Gema Prats*

Moderador: *Dña. Elisabete Aramendi*

PROGRAMA - VIERNES 26 DE NOVIEMBRE DE 2021

10:00 **Welcome & opening words**
- *By Enrique Javier Gómez, President of SEIB; and*
10:15 *Cristina Bescós, EIT Health Spain Managing Director.*

10:15 **Session to raise patients' voice in innovation**
- *By EUPATI Fellow.*
10:50 *Session followed by questions from the audience.*

10:50 **Roundtable: «How I became an Entrepreneur in Biomedical Engineering»**
- **Engineering»**
11:50 **With:**
- *Francesc Huguet París, Lainomedical*
- *Markus Wilhelms, Mowoot*
- *María González, Tucuvi*

11:50 *Break*
-
12:00

12:00 **Working'Health. Raising Life and Health Science Talent in Europe**
- *By Claire Nassiet*
12:30

12:30 **Acto de clausura**

- - *D. Enrique J. Gómez, Presidente CASEIB2021 y presidente SEIB*

14:30 - *D. Sergio Muñoz, Director Innovación de FENIN*

- *Dña. Cristina Bescós, Directora General de EIT Health Spain*

- *Dña. Patricia Sánchez, Secretaria del Comité Científico CASEIB2021*

Entrega de premios

- *D. Roberto Hornero (Premios FENIN-GIB)*

- *Dña. Elisabete Aramendi (Premios JM Ferrero)*

- *Dña. Lilisha Burris (Premios EIT Health)*

ÍNDICE DE SESIONES CIENTÍFICAS

Jueves 25 de Noviembre de 2021, a las 11:45

Premios FENIN para alumnos de Grado en Ingeniería Biomédica

Moderador: D. Roberto Hornero

Implementing an optical tweezers setup as a microrheology technique to study the mechanical properties of the cytoplasm	2
Diseño e implementación de algoritmos de machine learning para la caracterización objetiva del temblor patológico mediante señales cinemáticas y fisiológicas	6
Monitorización longitudinal de la compliancia pulmonar basada en la TIE en pacientes infectados por COVID-19	10
A Virtual Reality Application for Training Medical Students in a Pneumothorax Procedure	14
Reconstrucción en 3D del microambiente inmunológico de tumores de cáncer pancreático	18
Estudio de la arquitectura temporal de la actividad neuronal en la demencia por enfermedad de Alzheimer mediante técnicas de conectividad funcional dinámica	22
Optimización de hidrogeles inyectables y liberación de factores de crecimiento hepático.....	26
Discovering the drivers of aberrant splicing in cancer	30
Análisis de texturas como ayuda al diagnóstico neurooncológico en imágenes de resonancia magnética	34
Análisis de las variables biomecánicas para el estudio del índice de ruptura de la Triple A	38
SleepBCI: a platform for memory enhancement during sleep based on automatic scoring	42
Telerehabilitation of post-stroke motor impairment	46

Premios José María Ferrero Corral

Moderadora: Dña. Elisabete Aramendi

Prediction of short-term survival in glioblastoma patients with total tumor resection from multimodal pre-surgical MRI images using machine learning	51
Evaluación de la robustez de la red neuronal funcional en la enfermedad de Alzheimer	55
Análisis de las alteraciones en la organización jerárquica de la actividad neuronal debidas a la enfermedad de Alzheimer	59
Modelo de deep learning basado en la combinación de redes neuronales convolucionales y recurrentes para clasificar eventos de apnea e hipopnea mediante la señal de oximetría	63
Surgical phase recognition for non-intrusive surgical skills' assessment in educational contexts based in speech and video	67
La importancia dinámica como indicador de comportamiento de hub en la región parietal de la red neuronal por defecto	71
Sub-surface laser engraving application for hexagonal SiPM	75
Methodologies for the quantification and classification of nuclear medicine neuroimaging for clinical decision support	79
La importancia de la aurícula izquierda en la alteración del sustrato auricular tras ablación por catéter de Fibrilación Auricular Paroxística	83
Optimización preoperatoria de la configuración de los electrodos de estimulación con un modelo 3D de médula espinal personalizado al paciente	87
P-Wave detection using latent spaces from autoencoders and domain description classifiers	91
“Time-Warping” ponderado sobre la onda T para reducir el impacto de errores de delineación: implicaciones clínicas	95
Diseño e implementación de un sistema de simulación para el manejo de la fractura de pelvis	99
Cambio de FiO2 como alternativa al gadolinio para estudios de perfusión cerebral por resonancia magnética	103
Análisis de OCTA (angiografía por tomografía de coherencia óptica) mediante redes neuronales convolucionales para el diagnóstico de esclerosis múltiple	107

Desarrollo y validación de los bucles de control para las interfaces hápticas de un simulador de silla de ruedas manual	111
Modificación de un modelo de miocito ventricular humano para representar el papel de los canales SK en insuficiencia cardiaca	115
Análisis del impacto combinado del grupo sanguíneo y los antígenos leucocitarios humanos en la supervivencia del trasplante renal en Andalucía	119
Machine learning algorithms to predict adherence to treatment pathways in secondary prevention of patients with acute coronary syndrome	123
Deep learning-based segmentation of head and neck organs on CT for radiotherapy treatment: lessons learned with clinical data	127
Propuesta de modelos predictivos en salud mental para la personalización de terapias de rehabilitación en pacientes con adicción a la cocaína	131
Visualización de estudios preoperatorios en forma de nube de puntos volumétrica en dispositivos de realidad mixta	135
Evaluación de la calidad en registros electrocardiográficos para diagnósticos automáticos fiables mediante aprendizaje profundo y aumento de datos	139
Design, manufacture and validation of customized surgical guides for total knee replacement	143
Análisis paso a paso de registros de ablación por catéter en fibrilación auricular paroxística	147
Evaluación del riesgo de hipertensión a partir de registros fotopletimográficos utilizando clasificadores de deep learning	151
Diferencias entre patrones de ondas P y ondas de activación local del seno coronario durante ablación por catéter de fibrilación auricular paroxística	155
Reconocimiento automático de posiciones radiográficas en radiología	159
EEG-InceptionGen: una red convolucional de propósito general para la clasificación de señales EEG	163
Procesamiento de TAC de tórax para su uso en simulaciones de tomosíntesis.....	167
Electrophysiological modelling of left ventricle with heterogeneous scar to assess inducibility of reentry circuits	171
Inclusión de compensación de movimiento en un esquema deep learning basado en modelo para la reconstrucción de imágenes de cine cardíaco	175
Reconstrucción de imágenes para tomosíntesis de tórax mediante aprendizaje profundo	179
A simplified methodology to measure gait spatio-temporal parameters using low-cost inertial sensors	183

Wearable technology for predicting the quality of surgical performance and workload in laparoscopic surgery: preliminary results	187
Citometría de microflujo para estudios bacterianos	189
Información a priori obtenida con aprendizaje profundo para su uso en algoritmo de reconstrucción iterativo	193
Método para la caracterización de la señal de presión arterial invasiva continua durante la reanimación cardiopulmonar	197
Interfaz para el estudio del “jitter” neuromuscular	201
An immersive virtual reality training system for paediatric stroke code medical training	205

Premios Fenín Grado en Ingeniería Biomédica

Jueves 25 de Noviembre de 2021, 11:45

Implementing an optical tweezers setup as a microrheology technique to study the mechanical properties of the cytoplasm

A. Albajar Sigalés^{1,2}, A.E.M. Beedle², P. Roca-Cusachs Soulere^{1,2}

¹ Department of Biophysics and Bioengineering, University of Barcelona, Barcelona, Spain, aalbajsi7@alumnes.ub.edu, rocacusachs@ub.edu

² Institute for Bioengineering of Catalonia, The Barcelona Institute of Science and Technology, Barcelona, Spain, abcedle@ibecbarcelona.eu

Abstract

Cells are constantly sensing and generating forces through the cytoskeleton, a network of protein filaments that mechanically connects the cells to their substrate and to their nucleus. When mechanical forces reach the nucleus, they can trigger changes in gene expression and regulate cell behaviour. A recent finding indicates that fibrillar adhesions, a type of cell-substrate adhesions located under the nucleus, alter the transmission of mechanical stimuli from the membrane of the cell to the nucleus, making the nucleus less susceptible to mechanical perturbations. Here, we hypothesise that they do this by tuning the mechanical properties of the cytoplasm. To test this hypothesis, we first design a protocol to acquire microrheological measurements of the cytoplasm with an optical tweezers setup. Once the system is optimised, we study how fibrillar adhesions affect the mechanical properties of the cytoplasm and we find that cells with fibrillar adhesions exhibit a softer cytoplasm, suggesting a mechanism by which fibrillar adhesions can shield the nucleus from external mechanical perturbations.

1. Introduction

1.1. Cell mechanics

Tissues and organs are constantly submitted to and actively generating forces. Our lungs are stretched every time we breathe, our muscles contract every time we move, and our vessel walls experience shear stress due to the blood flow [1]. All these mechanical stimuli are sensed by the individual cells our tissues and organs are made of and can trigger responses as diverse as migration, proliferation and differentiation [2]. Moreover, the mechanisms by which cells transmit and sense these mechanical signals are finely tuned, and small alterations can lead to pathological processes, including fibrosis and cancer [3].

To respond to mechanical stimuli, cells need to convert a mechanical signal into a biochemical signal that can lead to a change in gene expression. One of the main players in this process is the cytoskeleton, a dynamic network of protein filaments that is linked to the cellular membrane through focal adhesions and to the nucleus through the LINC complex, thus allowing the transmission of mechanical stimuli throughout all the cell, including the nucleus. When mechanical forces reach the nucleus, they can induce the stretching of nuclear pores and affect the transport of molecules, or they can alter the conformation of the DNA and regulate its access to transcriptional machinery [4,5]. In essence, they can affect cell behaviour.

1.2. Fibrillar adhesions

Interestingly, by approximately three hours after being in contact with their substrate, cells form fibrillar adhesions, a mature form of focal adhesions that are associated with the remodelling of the fibronectin, an extracellular protein that is part of cells substrate, and with the formation of actin fibres, one of the main components of the cytoskeleton, beneath the nucleus (Figure 1).

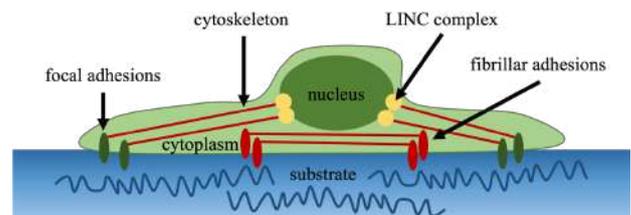


Figure 1. Sketch representing a cell with fibrillar adhesions and the associated fibronectin remodelling (blue curled lines) and actin fibres (red straight lines beneath the nucleus).

More importantly, a recent finding indicates that the presence of fibrillar adhesions alters the mechanical coupling between the cell and the nucleus. This is supported by experiments performed with a cell stretching device in which it was observed that, by stretching cells the same amount, with the presence of fibrillar adhesions, the nucleus deforms less, suggesting that fibrillar adhesions stop the deformations applied to the cell membrane from efficiently reaching the nucleus (Figure 2).

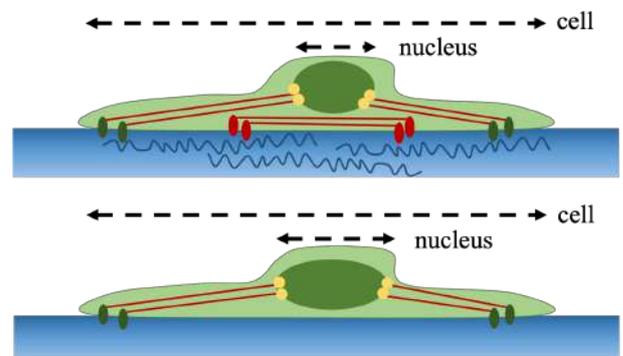


Figure 2. Cell stretching experiment with (top) and without (bottom) fibrillar adhesions.

A role of fibrillar adhesions in protecting the nucleus from deformations can have many implications in cell

physiology, but the mechanism remains unknown. Since it has been reported that the actin cytoskeleton is a major determinant of the mechanical properties of the cytoplasm [6], one of the possible explanations could be that the presence of fibrillar adhesions and, specifically, its associated actin fibres underneath the nucleus, modify the mechanical properties of the cytoplasm and alter the way in which mechanical stimuli are transmitted from the cellular membrane to the nucleus. To test this hypothesis, a non-invasive technique that can mechanically probe the interior of living cells and, particularly, the cytoplasm, is required.

1.3. Optical tweezers

Optical tweezers are a technique that use a highly focused laser beam to trap and move microscopic objects. Remarkably, unlike micropipette aspiration or Atomic Force Microscopy, which interact with the cell membrane, optical tweezers can directly probe the mechanical properties of intracellular structures by trapping micron-sized beads internalised in the cells [7]. Therefore, optical tweezers are a suitable technique to study the mechanical properties of the cytoplasm, and one of the ways in which it can be done is by performing microrheology, that is, by applying a small oscillatory displacement and measuring the resulting force (Figure 3A).

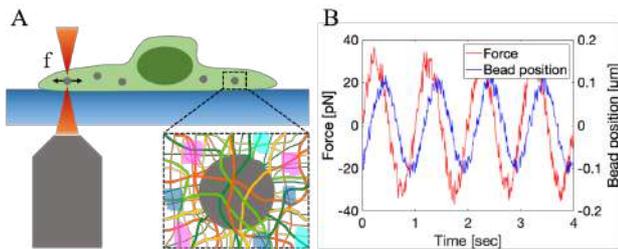


Figure 3. (A) Sketch representing a microrheological measurement of the cytoplasm with optical tweezers. (B) Typical force and bead displacement traces at 1 Hz.

Figure 3B shows the typical force and bead displacement traces obtained when performing microrheology of the cytoplasm. If the behaviour of the material is totally elastic, the response of the bead is immediate, and the two traces do not show any difference in phase, whereas if the material is totally viscous, the phase shift is 90° . In the case of the cytoplasm, a viscoelastic material, the phase shift is between 0 and 90° . Furthermore, since both the stress and the strain applied are known, the complex shear modulus, composed by the elastic and the viscous modulus, can be calculated (Equation 1) [7].

$$\sigma(t) = \gamma_0 \cdot [G'(\omega) \cdot \sin(\omega t) + G''(\omega) \cdot \cos(\omega t)]$$

Equation 1. Microrheology equation. G' is the elastic modulus, G'' is the viscous modulus, σ is the stress, γ_0 is the strain amplitude, t is time, ω is frequency.

Accordingly, **the objective of this project** is to design a protocol to study the mechanical properties of the cytoplasm using an optical tweezers setup as a microrheology technique and to apply it to study how fibrillar adhesions affect the mechanical properties of the cytoplasm.

2. Methods

Addressing the above-mentioned experimental question first required designing and testing a protocol to measure the mechanical properties of the cytoplasm. This process included three steps: (1) defining a protocol to internalise beads in the cytoplasm of cells, (2) optimising a protocol to acquire microrheological measurements of the cytoplasm with an optical tweezers setup, and (3) establishing a method to display and statistically test the data obtained. Once the system was functioning correctly, the mechanical properties of the cytoplasm of cells with and without fibrillar adhesions were measured.

2.1. Internalisation of beads

Measuring the mechanical properties of the cytoplasm with optical tweezers requires having single beads embedded in the cytoplasm of cells. With this purpose, T1FF fibroblast cells were seeded on top of coverslips functionalised with fibronectin and incubated overnight in cell media with $1 \mu\text{m}$ in diameter polystyrene beads at a concentration of $60 \mu\text{g/ml}$. This allowed cells to internalise beads by phagocytosis. To check that the beads were embedded in the cytoplasm, samples were fixed and stained for the nucleus and the actin cytoskeleton, which colocalises with the cytoplasm, with Hoechst and fluorescently labelled phalloidin, respectively. Then, samples were imaged with confocal microscopy, which allowed visualising simultaneously the beads and the cytoplasm in 3D and determining if the beads were embedded in the cytoplasm.

2.2. Acquisition of measurements

The microrheological measurements of the cytoplasm were acquired with the optical tweezers platform SENSOCCELL (Impetux Optics) and consisted in applying a 200 nm in amplitude oscillatory displacement at different frequencies and measuring the resulting force. This allowed the system to calculate the complex shear modulus at each frequency. Furthermore, to prove that the system was functioning correctly, a control that tuned the mechanical properties of the cytoplasm was implemented. This control consisted in treating cells with cytochalasin D $1 \mu\text{M}$ in cell media. Cytochalasin D is a cytoskeleton drug that disrupts actin filaments and inhibits actin polymerisation. Since the viscoelastic properties of the cytoplasm are known to be dominated by the cytoskeleton, it was expected that the addition of cytochalasin D would decrease the mechanical resistance of the cytoplasm [6].

2.3. Data analysis

To present the experimental data acquired with the optical tweezers setup in a user-friendly way that can be easily manipulated and analysed, a MATLAB program that is able to automatically display and statistically test the results was implemented. Specifically, the program reads the data files generated with the optical tweezers setup, plots the mean and the standard error of the complex shear modulus at each frequency, checks that the data follows a normal distribution and performs a 2-Way ANOVA test between two different cell conditions indicated by the user.

2.4. Fibrillar adhesions experiment

Once the protocol to measure the mechanical properties of the cytoplasm was well established, it was used to study the mechanical properties of the cytoplasm in cells with and without fibrillar adhesions. To create the two different cell conditions, cells were either seeded on top of coverslips functionalised with fibronectin and allowed to form fibrillar adhesions or seeded on top of fibronectin treated with glutaraldehyde or PFA, two treatments that are known to prevent the formation of fibrillar adhesions by crosslinking fibronectin and stopping its remodelling.

3. Results and discussion

Figure 4A and Figure 4B show a Z slice and a Z projection, respectively, of a bead embedded in the cytoplasm of a cell imaged with confocal microscopy. To further prove that the bead is inside the cell in the Z direction, Figure 4C shows the intensity profile of the bead and the actin signal along this direction. Both signals have a maximum intensity peak in the same position and, moreover, the actin signal is wider than the bead signal. Taken together, these results strongly suggest that the beads were embedded in the cytoplasm of the cells.

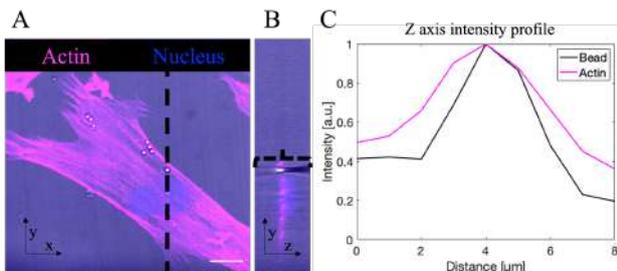


Figure 4. A) Z slice of a cell with a bead. (B) Z projection of the same cell along the dashed line. (C) Z intensity profile of the bead and the actin signals. The curly brackets indicate where the Z intensity profile was measured. Scale bar = 10 μm .

Figure 5 shows the organisation of the actin cytoskeleton of untreated and cytochalasin D treated TIFF fibroblast cells and verifies that cytochalasin D triggers large-scale changes to the actin cytoskeleton. Since it is known that the viscoelastic properties of the cellular cytoplasm are dominated by the contribution of the actin cytoskeleton [6], treating cells with cytochalasin D is a good control to prove that the microrheology experiments are directly probing the mechanical properties of the cytoplasm.

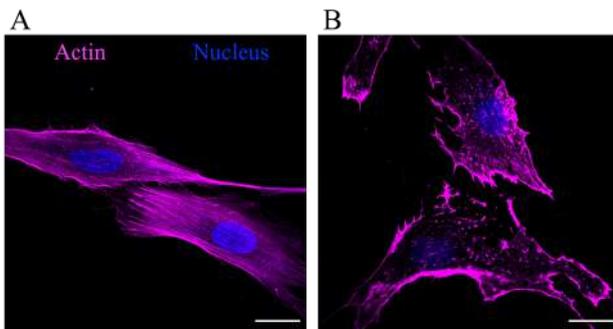


Figure 5. (A) Untreated TIFF fibroblast cells. (B) TIFF fibroblast cells treated with cytochalasin D. Scale bars = 20 μm .

Figure 6 presents the complex shear modulus of the cytoplasm of untreated and cytochalasin D treated TIFF fibroblast cells measured with optical tweezers. For the frequency range investigated, the elastic shear modulus dominates over the viscous shear modulus, indicating that the cytoplasm behaves more as a solid-like material than a fluid-like material. Furthermore, the complex shear modulus increases linearly in a loglog scale as a function of frequency. These findings are consistent with previous reports in the literature investigating the viscoelastic properties of other cell types [6,7]. Microrheological measurements on cytochalasin D treated cells revealed that these cells exhibited a softer cytoplasm compared to untreated cells, with a decrease of both the elastic and the viscous component of by approximately 2-fold. These results are consistent with previous reports in the literature for other cytoskeleton drugs and demonstrate that the system is sensitive to changes in the organisation of the actin cytoskeleton [6].

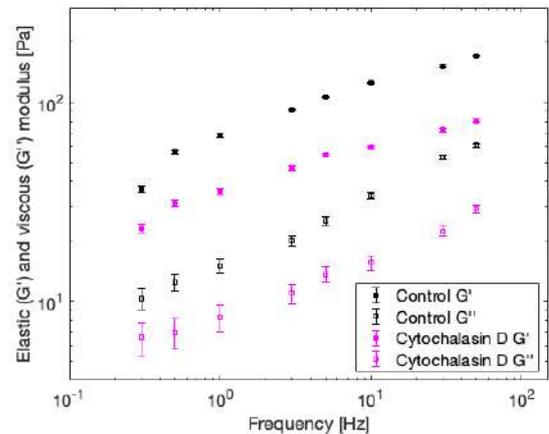


Figure 6. Complex shear modulus of untreated and cytochalasin D treated cells. The differences between the control and the condition are significant for both the elastic and the viscous modulus ($p < 0.001$). Error bars represent the standard error ($n = 24$ cells from 2 independent experiments for control cells, $n = 25$ cells from 2 independent experiments for cytochalasin D treated cells).

Once the system was optimised to perform active microrheology experiments inside the cytoplasm of living cells, the mechanical properties of the cytoplasm of cells with and without fibrillar adhesions were measured. Figure 7 shows the apical and the basal organisation of actin in control cells and in cells where the substrate is pre-treated with glutaraldehyde, which prevents the formation of fibrillar adhesions. Whereas the apical actin organisation is maintained in both conditions, cells that form fibrillar adhesions present additional actin fibres in the basal side. The current hypothesis is that this slight alteration in the actin organisation could change how deformations applied to the cell membrane are efficiently transmitted to the nucleus.

In fact, Figure 8 shows that cells that were able to form fibrillar adhesions exhibited a softer cytoplasm. The same experiment was performed with cells seeded on substrates treated with PFA, which also blocks the formation of

fibrillar adhesions, and similar results were obtained (data not shown).

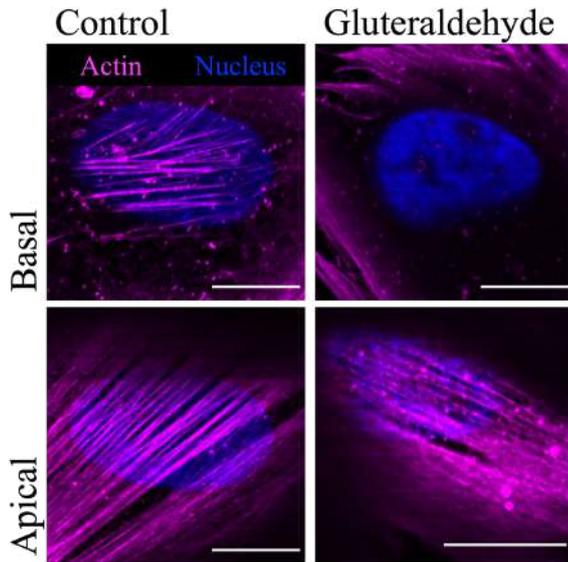


Figure 7. Actin organisation in control cells and in cells seeded on substrates treated with glutaraldehyde. Scale bars = 10 μm .

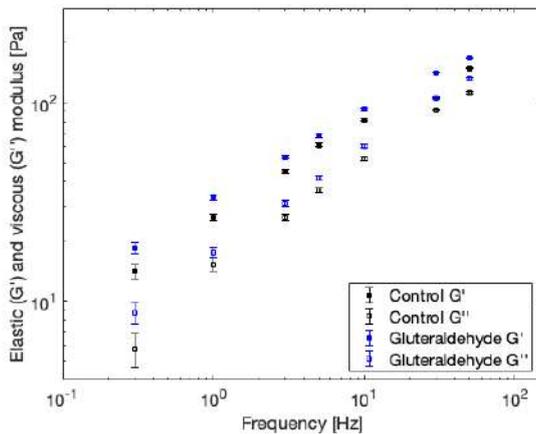


Figure 8. Complex shear modulus of cells seeded on untreated and glutaraldehyde treated substrates. The differences between the control and the condition are significant for both the elastic and the viscous modulus ($p < 0.001$). Error bars represent the standard error ($n = 44$ cells from 2 independent experiments for control cells, $n = 36$ cells from 2 independent experiments for cells seeded on glutaraldehyde treated substrates).

After having observed that the cytoplasm of cells with fibrillar adhesions is softer, the fact that the presence of fibrillar adhesions stops the deformations applied to the cell membrane from efficiently reaching the nucleus could be explained by the model presented in Figure 9. This model considers the cell as three springs in series, including the stiffness of the cytoplasm and the nucleus. When fibrillar adhesions are present, the stiffness of the cytoplasm is lower and, when submitting cells to a certain strain, the cytoplasm absorbs most of this deformation. In contrast, when fibrillar adhesions are not present, the stiffness of the cytoplasm is higher, the cytoplasm deforms less, and the nucleus deforms more to compensate it. Although this model is consistent with the experimental results, it is worth mentioning that it does not take into

account that fibrillar adhesions could also affect the stiffness of the nucleus, which is still unknown.

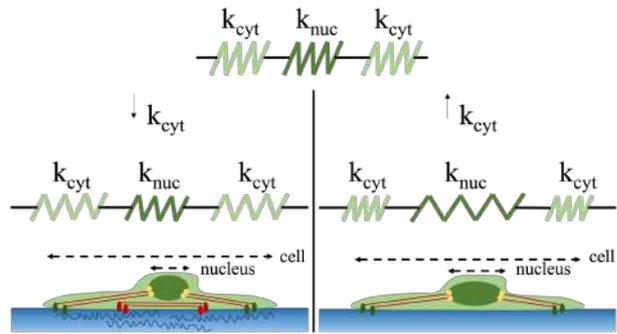


Figure 9. Spring model of a cell with (left) and without (right) fibrillar adhesions.

4. Conclusions

The protocol presented allows measuring the mechanical properties of the cytoplasm by performing microrheology with an optical tweezers setup. More importantly, it allowed studying how fibrillar adhesions affect the mechanical properties of the cytoplasm and revealed that cells with fibrillar adhesions exhibit a softer cytoplasm. These results suggest that one of the mechanisms by which fibrillar adhesions affect the mechanical coupling between the cell and the nucleus may be by tuning the mechanical properties of the cytoplasm through the presence of additional actin fibres underneath the nucleus. However, how the stiffness of the nucleus is affected by the presence of fibrillar adhesions remains to be investigated. Interestingly, it could also be studied using optical tweezers and a similar protocol to the one presented here.

References

- [1] D. E. Jaalouk and J. Lammerding, "Mechanotransduction gone awry," *Nature Reviews Molecular Cell Biology*, vol. 10, no. 1, pp. 63–73, 2009, doi: 10.1038/nrm2597.
- [2] E. C. Yusko and C. L. Asbury, "Force is a signal that cells cannot ignore," *Molecular Biology of the Cell*, vol. 25, no. 23, pp. 3717–3725, 2014, doi: 10.1091/mbc.E13-12-0707.
- [3] D. J. Tschumperlin and D. Lagares, "Mechano-therapeutics: Targeting Mechanical Signaling in Fibrosis and Tumor Stroma," *Pharmacology and Therapeutics*, vol. 212, p. 107575, 2020, doi: 10.1016/j.pharmthera.2020.107575.
- [4] A. Elosegui-Artola *et al.*, "Force Triggers YAP Nuclear Entry by Regulating Transport across Nuclear Pores," *Cell*, vol. 171, no. 6, pp. 1397–1410.e14, 2017, doi: 10.1016/j.cell.2017.10.008.
- [5] A. Tajik *et al.*, "Transcription upregulation via force-induced direct stretching of chromatin," *Nature Materials*, vol. 15, no. 12, pp. 1287–1296, 2016, doi: 10.1038/nmat4729.
- [6] M. Guo *et al.*, "Probing the stochastic, motor-driven properties of the cytoplasm using force spectrum microscopy," *Cell*, vol. 158, no. 4, pp. 822–832, 2014, doi: 10.1016/j.cell.2014.06.051.
- [7] S. K. Gupta and M. Guo, "Equilibrium and out-of-equilibrium mechanics of living mammalian cytoplasm," *Journal of the Mechanics and Physics of Solids*, vol. 107, pp. 284–293, 2017, doi: 10.1016/j.jmps.2017.07.007.

Diseño e implementación de algoritmos de machine learning para la caracterización objetiva del temblor patológico mediante señales cinemáticas y fisiológicas

V. Lopo-Martínez¹, A. Pascual-Valdunciel^{1,2}, F.O. Barroso², A. Gutiérrez¹

¹ E.T.S. Ingenieros de Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain

victorlopmartinez@gmail.com, alejandropv@cajal.csic.es, a.gutierrez@upm.es

² Neural Rehabilitation Group, Instituto Cajal, CSIC, Madrid, Spain

Resumen

El temblor esencial es un trastorno del sistema nervioso que provoca la aparición de temblor patológico en alguna parte del cuerpo o extremidad, pudiendo llegar a comprometer a la realización de actividades de la vida diaria del paciente. Actualmente, no existe ningún tratamiento eficaz y accesible para todos los pacientes contra el temblor esencial. Las nuevas líneas de investigación presentan la estimulación eléctrica de las vías aferentes junto con su estrategia de estimulación out-of-phase como una alternativa barata e inocua capaz de superar los efectos secundarios y desventajas de los tratamientos más convencionales. La estrategia out-of-phase se basa en la estimulación sobre un par de músculos antagonistas sincronizadas con la fase de la actividad muscular del temblor. Sin embargo, la capacidad predictiva de la estrategia out-of-phase es limitada, dando lugar a la aparición de desincronizaciones entre las estimulaciones y la actividad tremorogénica. Como solución a este problema, en este artículo se propone, explora y valida el uso de algoritmos de machine learning y deep learning para caracterizar de manera objetiva el temblor patológico y predecir los futuros ciclos de temblor del paciente. Se consigue por tanto optimizar la estrategia out-of-phase dentro de la estimulación eléctrica de las vías aferentes dando lugar a una mayor supresión del temblor patológico en pacientes con temblor esencial.

1. Introducción

El temblor se define como el movimiento involuntario de una parte del cuerpo o extremidad como consecuencia de oscilaciones neurales patológicas que se proyectan hacia los músculos, provocando la contracción alternante o sincrónica de músculos agonistas y antagonistas [1]. El temblor se hace más evidente en las extremidades superiores, especialmente en los movimientos de la mano, el último efector involucrado en las tareas de control fino como agarre. La enfermedad más común que produce temblor patológico es el temblor esencial (TE), con una incidencia estimada alrededor del 4 % de la población mundial por encima de los 65 años [2]. El TE se caracteriza por un rango de frecuencias que oscila entre 4 y 12 Hz, pudiendo variar en intensidad y con la edad [2].

1.1. Tratamientos convencionales

La medicación es la primera línea de tratamiento para el TE [3]. Los tratamientos basados en medicamentos pueden aportar una mejora sintomática y un aumento de la calidad

de vida de al menos 5-8 años. Pese a ello, se estima que los medicamentos son efectivos únicamente para el 50% de los pacientes que padecen TE y los efectos secundarios asociados a los mismos son comunes, llegando a provocar el abandono de este tipo de tratamientos en un tercio de las ocasiones [3]. En aquellos casos en los que la medicación es insuficiente o sus efectos adversos excesivos, se emplean tratamientos quirúrgicos como la estimulación cerebral profunda (DBS) de los ganglios basales o las talamotomías, entre las que se encuentran los novedosos Ultrasonidos Focalizados de Alta Intensidad (HIFU). La gran desventaja de la DBS es su bajo criterio de elegibilidad y los estrictos requisitos necesarios para que el paciente sea considerado apto para este procedimiento [4]. Las talamotomías provocan la destrucción total e irreversible de un núcleo talámico, lo que supone una gran desventaja frente a otro tipo de tratamientos [4].

1.2. Estimulación de las vías aferentes

La limitada eficacia de los fármacos, los restrictivos criterios de elegibilidad, los efectos secundarios y el elevado coste de los tratamientos quirúrgicos conduce a la búsqueda de soluciones alternativas dirigidas hacia la población general de TE, con efectos adversos reducidos, mínimamente invasivos y de coste asequible [4]. La estimulación eléctrica de las vías aferentes es una alternativa prometedora en la reducción del temblor patológico. Esta técnica consiste en la estimulación de las vías aferentes o sensitivas por debajo del umbral motor de las motoneuronas para producir una modulación a nivel espinal y supraspinal a través de las interneuronas y los circuitos reflejos. En este tipo de estrategia la precisa sincronización de los pulsos de estimulación eléctrica con la actividad fisiológica tremorogénica es vital para lograr la activación de los circuitos de neuromodulación [5]. La estrategia de estimulación *out-of-phase* de lazo cerrado [4] consiste en la estimulación sincrónica del músculo antagonista a la actividad temblorosa. Generalmente se ha implementado mediante señales de registro cinemáticas o de electromiografía (EMG), calculando la media del periodo de los ciclos de temblor durante una ventana de lectura y utilizando dicha media para la predicción de los futuros ciclos de temblor con los que se sincronizará la estimulación [6]. Como alternativa para superar los errores de predicción del análisis de fase, se diseñó la estrategia Selective and Adaptive Timely Stimulation (SATS),

basada en el uso de ventanas secuenciales de lectura y estimulación [7]. Si en la ventana de medición se supera un umbral adaptativo se aplica la estimulación sobre el antagonista. No obstante, esta estrategia está limitada por la latencia asociada a la resolución temporal de las ventanas y la posible presencia de artefactos.

1.3. Motivación

La predicción de futuros ciclos de temblor de la estrategia *out-of-phase*, basada en la frecuencia del mismo no es robusta, dando lugar a la aparición de desincronizaciones entre la fase del temblor muscular y las consecuentes estimulaciones. Este problema se puede resolver por medio de algoritmos de machine learning y deep learning. Para ello, en esta contribución se exploran una serie de clasificadores que permiten detectar la actividad tremorogénica durante una ventana de lectura para estimular únicamente en aquellos instantes en los que el paciente presente temblor patológico. Posteriormente, se aplican algoritmos de predicción de los futuros ciclos de temblor del paciente. El resultado final es la aportación de algoritmos que optimicen la estrategia *out-of-phase* con la estimulación eléctrica de las vías aferentes para alcanzar una mayor supresión del temor patológico.

2. Diseño y Metodología

Doce pacientes con TE de la Clínica de Desórdenes del Movimiento del hospital Gregorio Marañón (Madrid) fueron seleccionados entre abril de 2019 y enero de 2020 para la captura de las señales cinemáticas de temblor que forman el conjunto de datos necesario para desarrollar las tareas de clasificación y predicción. Los pacientes con una edad entre 62-82 años presentaban temblor postural con afección en al menos una de las extremidades superiores y con un temblor de flexo-extensión de muñeca prominente. El protocolo de experimentación fue aprobado por el comité de ética del hospital y todos los pacientes participaron en el estudio de manera voluntaria firmando el consentimiento informado. Para complementar el conjunto de datos de clasificación se realizaron los registros de cinco sujetos sanos replicando el protocolo experimental de los pacientes. Se realizó la captura de señales cinemáticas mientras los sujetos mantenían durante 60 segundos una postura que facilitase la aparición de temblor, con los brazos estirados al frente y en pronación, o bien con los brazos elevados, los codos flexionados y los dedos enfrentados entre sí a la altura del rostro. La adquisición de las señales cinemáticas se realizó con un sistema de captura del movimiento (Technaid S.L, España). Se emplearon 2 sensores de medida inercial (IMU) colocados sobre la mano y el antebrazo afectado por el temblor. Los datos cinemáticos fueron capturados en forma de cuaterniones para la articulación de la muñeca con una frecuencia de muestreo de 50 Hz y almacenados para su posterior conversión a ángulos de Euler. En total se emplearon para la el conjunto de datos de clasificación un total de 18.000 secuencias temporales de 1 segundo cada una, de las cuales 9.000 fueron etiquetadas como señales de temblor y las 9.000 restantes como señales de no temblor. Para la base de datos de predicción se emplearon 7000 segmentos temporales de 2 segundos cada uno y se eliminaron las

secuencias temporales correspondientes a pacientes sanos. Los datos fueron normalizados entre 0 y 1. En la *Figura 1* se observa una comparativa de la variación de los ángulos de flexo-extensión de muñeca de una señal con temblor patológico y una señal sin temblor patológico. Con el fin de eliminar las componentes frecuenciales no relativas al TE se aplicó sobre las señales un filtro en ambas direcciones tipo Butterworth, paso banda de orden 3 entre 4 y 10 Hz. Para el etiquetado del conjunto de datos de clasificación se empleó el valor de la Densidad Espectral de Potencia (DEP) de las señales procesadas, calculada en base al método de Welch. [8].

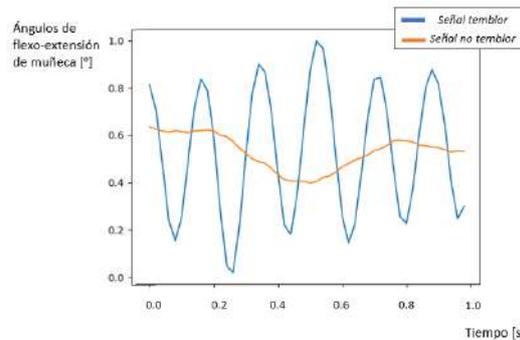


Figura 1. Comparativa señal temblor / señal no temblor

Tanto el conjunto de datos de clasificación como el de predicción se dividieron en 3 subconjuntos de datos cada uno. El 70% de las secuencias temporales se utilizó para entrenar los algoritmos (conjunto de entrenamiento), el 15% para validarlos (conjunto de validación) y el 15% restante para evaluar la calidad de los modelos (conjunto de prueba). Los clasificadores empleados son el K-Nearest Neighbours (KNN), el Random Forest, la Support Vector Machine (SVM) y una red neuronal Long Short-Term Memory (LSTM) mientras que para la tarea de predicción se diseñó e implementó una red neuronal LSTM.

El clasificador KNN parte de un conjunto de datos X en el que x es una instancia del conjunto total e y es la etiqueta de x . Siendo k el número de vecinos a tener en cuenta, el KNN busca en las k secuencias temporales más cercanas a x por medio de la distancia euclídea y calcula la probabilidad condicionada de que x pertenezca a la clase j [9].

$$P_r(Y = j|x) = \frac{1}{k} \sum_{i \in N} I(y_i = j)$$

El Random Forest es un ensamble de varios árboles de decisión. Un árbol de decisión consiste en un conjunto de ramas y nodos, donde los nodos son valores de los instantes temporales de la señal y las ramas reglas de decisión. Un nodo padre, se subdivide en dos nodos hijos. La elección del nodo se hace en base a la Ganancia de Información (IG) del nodo. Se escogerá aquel con mayor valor de IG. Queda definida la IG a continuación, donde E es el valor de la entropía de los datos de cada nodo y w es el tamaño relativo del nodo hijo con respecto al nodo padre [9].

$$IG = E(\text{Nodo padre}) - \sum (w_i E(\text{Nodo hijo}))$$

Las SVM son clasificadores que toman la secuencia temporal y la interpretan como un vector de características dentro de un espacio Ω -dimensional, siendo Ω el número de instantes temporales de la secuencia temporal ($\Omega=50$). Dentro este espacio la SVM crea un hiperplano ($w\mathbf{x} + b = 0$). Una vez definido el hiperplano se emplea la función de hipótesis ($h(x)$) para realizar las predicciones binarias [9].

$$h(x) = \begin{cases} 0 & : wx + b < 0 \\ 1 & : wx + b \geq 0 \end{cases}$$

Las redes neuronales LSTM superan las limitaciones teóricas de las redes neuronales artificiales simples y de las redes neuronales recurrentes, ya que tienen dependencias temporales tanto a corto como a largo plazo gracias a la solución al problema del desvanecimiento de gradientes, lo que las hace idóneas para el tratamiento de series temporales. Una celda LSTM está formada por 4 estructuras básicas llamadas puertas: puerta de olvido, puerta de entrada, puerta de estado y puerta de salida [10].

De cada uno de los clasificadores tradicionales se crearon distintos modelos variando los parámetros específicos de cada clasificador. Para el KNN se crearon un total de 6 modelos, 12 para el SVM y 18 para el Random Forest. El modelo óptimo de cada clasificador se obtuvo por medio de la técnica de validación *k-fold cross validation*. Se estableció un valor de $k = 10$, donde el grupo de validación se divide en $k-1$ subconjuntos para entrenar y el subconjunto k para validar el clasificador. Una vez obtenido el modelo óptimo de cada clasificador, se entrenó con el conjunto de entrenamiento para evaluarlo finalmente sobre el conjunto de prueba. Para la red neuronal LSTM de clasificación se crearon un total de 9 modelos, variando el número de neuronas ocultas de cada capa (20, 35 y 50) y el learning rate (0.0001, 0.001 y 0.005). El clasificador consta de 3 capas, 2 capas LSTM y una capa lineal. La salida de la red se compara con la salida real para calcular la función de pérdidas (L). Los pesos y bias de la red se actualizan con el objetivo de alcanzar el mínimo de la L gracias al descenso de gradientes y al valor del learning rate, que actúa en la velocidad de convergencia de la red dentro de la búsqueda del mínimo de L [11].

Durante la etapa de entrenamiento se validaron los modelos con el conjunto de validación para evitar un sobreajuste de la red. Los modelos con los pesos óptimos se evaluaron sobre el conjunto de prueba y se escogió aquel que mejor rendimiento mostró sobre dicho conjunto. El rendimiento de los clasificadores se midió en base a los resultados en exactitud, precisión, sensibilidad y especificidad. Estas métricas se obtienen a partir de la matriz de confusión, que indica el número de verdaderos positivos (VP), falsos positivos (FP), verdaderos negativos (VN) y falsos negativos (FN).

$$\begin{aligned} \text{Precisión} &= \frac{VP}{VP + FP} & \text{Especificidad} &= \frac{VN}{VN + FN} \\ \text{Exactitud} &= \frac{VP + VN}{VP + VN + FP + FN} & \text{Sensibilidad} &= \frac{VP}{VP + FN} \end{aligned}$$

Para la red neuronal LSTM de predicción se crearon un total de 180 modelos. Los parámetros de la red son el número de capas (1 y 2), número de neuronas ocultas de cada capa (20, 35 y 50), learning rate (0.0001, 0.001 y

0.0005), número de muestras del horizonte de predicción (5, 10, 20, 30 y 50) y número de muestras de entrenamiento (30 y 50). En la red neuronal LSTM de predicción (ver *Figura 2*) la señal cinemática del temblor es la entrada de la primera capa LSTM, pasa a través de la segunda capa LSTM para acabar en la capa lineal. La salida de la capa lineal (y) será el valor de la señal cinemática en las próximas z muestras, siendo z el horizonte de predicción.

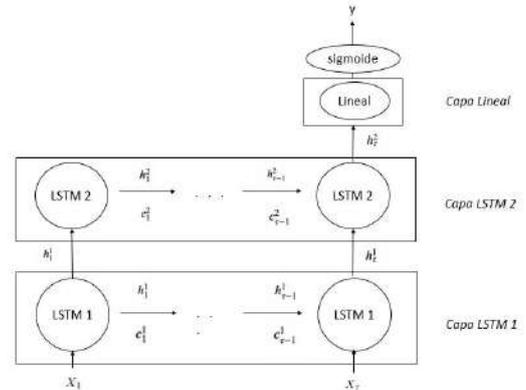


Figura 2. Arquitectura red neuronal LSTM de predicción

Para cada combinación de muestras de entrenamiento y horizonte de predicción se evaluó la calidad de los distintos modelos, siguiendo la misma dinámica de trabajo especificada en la red neuronal LSTM de clasificación. La métrica empleada para evaluar la calidad de los modelos es el coeficiente de correlación de Pearson que indica la similitud entre la señal real y la señal predicha.

$$r_{t,y} = \frac{\sum_{i=1}^N (t_i - \bar{t})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (t_i - \bar{t})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}}$$

3. Resultados

El rendimiento en base a las métricas definidas previamente de los modelos óptimos para cada tipo de clasificador queda recogido en de la *Tabla 1*. Las métricas a evaluar son la exactitud (Ex), precisión (Pr), sensibilidad (Se) y especificidad (Es).

	Ex	Pr	Se	Es
KNN	94 %	90 %	99 %	89 %
SVM	91 %	86 %	98 %	84 %
Forest	96 %	97 %	95 %	96 %
LSTM	97 %	97 %	97 %	97 %

Tabla 1. Resultados de los clasificadores

El rendimiento de las predicciones se muestra a través de boxplots, métodos estandarizados para representar gráficamente la correlación entre las predicciones y las señales reales a través de sus cuartiles. Pese a haber calculado los boxplots de las correlaciones para todas las posibles combinaciones de muestras de entrenamiento y horizontes de predicción, en la *Figura 3* se muestran los boxplots obtenidos para 50 muestras de entrenamiento y un

horizonte de predicción de 50 muestras, que se corresponde con 1 segundo futuro.

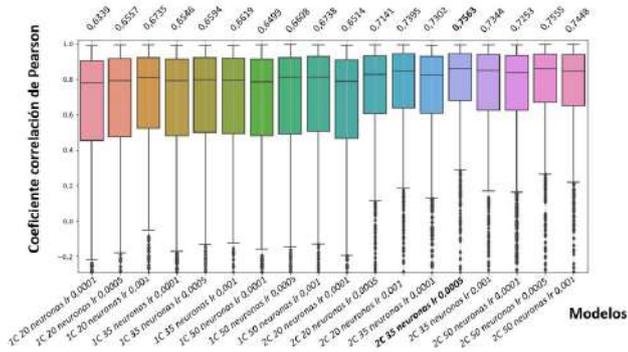


Figura 3. Boxplot 50 entrenamiento / 50 predicción

La red neuronal con 2 capas, 35 neuronas y un learning rate de 0.0005 es capaz de predecir 1 segundo de la señal de temblor patológico con una correlación del 75.63 %. Se muestra en la Figura 4 un ejemplo de la predicción de señal realizada por este modelo LSTM. A medida que disminuye el horizonte de predicción la correlación es mayor, llegando a alcanzar valor del 99.87% para la predicción de 5 futuras muestras (100 ms).

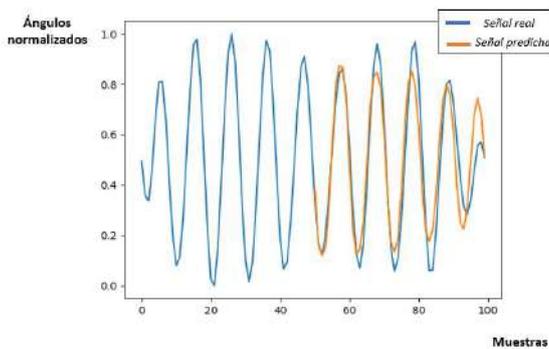


Figura 4. Ejemplo de predicción 1 segundo a futuro

4. Discusión

Todos los clasificadores presentan valores de exactitud superiores al 90 % por lo que son capaces de realizar de adecuadamente la clasificación binaria de señales de temblor y no temblor. Sin embargo, los resultados obtenidos por la red neuronal LSTM son ligeramente superiores al resto de clasificadores con unos valores del 97 % en cada una de las métricas evaluadas. Sus porcentajes de precisión y especificidad son notablemente superiores a la SVM y al KNN. La tarea de predicción ha sido satisfactoriamente abordada por la red neuronal LSTM. Su capacidad en el tratamiento de series temporales y señales cinemáticas de temblor le han permitido alcanzar valores de correlación del 99.87 %. Pese a que a medida que aumenta el horizonte de predicción disminuye la correlación, horizontes de predicción mayores (1 segundo) ofrecen un mayor rango temporal para la aplicación de las estimaciones en la estrategia *out-of-phase*. Por ese motivo, el modelo formado por 2 capas LSTM, 35 neuronas y 0.0005 de learning rate es el modelo óptimo, ya que presenta una mayor correlación (75.63 %) dentro de los modelos que predicen 1 segundo a futuro, siendo capaz

de adaptarse a las variaciones de frecuencia y amplitud propias del temblor patológico.

5. Conclusión

Tanto los algoritmos de clasificación como los algoritmos de predicción han conseguido adaptarse al problema del tratamiento de señales cinemáticas y fisiológicas del temblor patológico. Destacan los resultados obtenidos por las redes neuronales LSTM tanto en la tarea de clasificación como en la de predicción, todo ello gracias a su facilidad para el reconocimiento de patrones tanto a corto como a largo plazo. Las redes LSTM han demostrado una auténtica capacidad predictiva de las señales futuras del temblor, permitiendo optimizar la estrategia *out-of-phase* dentro de la estimulación eléctrica de las vías aferentes para conseguir una mayor supresión del temblor y una precisa activación de los circuitos de neuromodulación.

Referencias

- [1] R.C. Helmich, I. Toni, G. Deuschl, and B. R. Bloem. The pathophysiology of essential tremor and Parkinson's tremor. *Curr Neurol Neurosci Rep*, 13(9):378, Sep 2013.
- [2] J. A. Gallego, J. L. Dideriksen, A. Holobar, J. Ibáñez, J. L. Pons, E. D. Louis, E. Rocon, and D. Farina. Influence of common synaptic input to motor neurons on the neural drive to muscle in essential tremor. *J Neurophysiol*, 113(1):182-191, Jan 2015.
- [3] P. Hedera, F. Cibulcik, and T. L. Davis. Pharmacotherapy of essential tremor. *J Cent Nerv Syst Dis*, 5:43-55, Dec 2013.
- [4] Pascual-Valdunciel A, Hoo GW, Avrillon S, Barroso FO, Goldman JG, Hernandez-Pavon JC, Pons JL. Peripheral electrical stimulation to reduce pathological tremor: a review. *J Neuroeng Rehabil*. 2021 Feb 15;18(1):33.
- [5] A. Pascual-Valdunciel, F.O. Barroso, and J.L. Pons, Motor inhibition elicited by electrical stimulation of afferent pathways and its application in tremor suppression.
- [6] S. Dosen, S. Mucelli, J. L. Dideriksen, J.P. Romero, E. Rocon, J. Pons, and D. Farina. Online tremor suppression using electromyography and low-level electrical stimulation. *IEEE Trans Neural Syst Rehabil Eng*, 23(3):385-395, May 2015.
- [7] A. Pascual-Valdunciel, M. Gonzalez-Sanchez, S. Muceli, B. Adan-Barrientos, V. Escobar-Segura, J.R. Perez-Sanchez, M.K. Jung, A. Schneider, K. P. Hoffmann, J.C. Moreno, F. Grandas, D. Farina, J.L. Pons, and F.O. Barroso. Intramuscular stimulation of muscle afferents attains prolonged tremor reduction in essential tremor patients. *IEEE Trans Biomed Eng*, 68(6):1768-1776, Jun 2021.
- [8] A. Oppenheim and A. Wilsky. Signals and Systems. MIT, 2 edition, 1997.
- [9] Andriy Burkov. *The Hundred-Page Machine Learning Book*. 2019.
- [10] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput*. 9 (8). (1997) 1735–1780.
- [11] R. Sendra-Arranz and A. Gutiérrez. A long short-term memory artificial neural network to predict daily HVAC consumption in buildings. *Energy and Buildings*, 216:378-7788, 2020.

Monitorización longitudinal de la compliancia pulmonar basada en la TIE en pacientes infectados por COVID-19

I. Isasa Reinoso¹, A. Alberdi Aramendi¹, M. Barrenechea¹, R. Chen², A. Lovas³, K. Möller²

¹ Departamento de Ingeniería Biomédica, Mondragon Unibertsitatea, 20500 Mondragón, España, imanol.isasa@alumni.mondragon.edu, {aalberdiar, mbarrenetxea}@mondragon.edu

² Institute of Technical Medicine, Furtwangen University, Villingen-Schwenningen, Germany, {chr, moe}@hs-furtwangen.de

³ Kiskunhalas Semmelweis Hospital, Department of Anaesthesiology and Intensive Therapy, H-6400, Dr. Monzpart L. u. 1, Hungary, anlovas@gmail.com

Resumen

El COVID-19 es una infección vírica que causa complicaciones en el sistema respiratorio. Los síntomas más comunes sugieren que las tecnologías de imagen médica pueden ofrecer información relevante sobre el diagnóstico, tratamiento, y posterior recuperación de este tipo de infecciones. Entre estas técnicas de imagen se encuentra la Tomografía de Impedancia Eléctrica (TIE), la cual se trata de una técnica no invasiva, libre de radiación ionizante y de uso continuo seguro que genera imágenes utilizando secuencias de inyecciones de corriente eléctrica y mediciones de voltaje alrededor del cuerpo. Por todo ello, la TIE podría considerarse como una herramienta adecuada para el estudio y monitorización del comportamiento regional de los pulmones infectados. Además, esta técnica podría también ser una herramienta válida para una clasificación preliminar entre los fenotipos del COVID-19. Este estudio se basa en la monitorización longitudinal de dos pacientes infectados por el nuevo coronavirus: los resultados indican que uno de los pacientes podría pertenecer al fenotipo H, mientras que el segundo podría clasificarse como tipo L. Se ha concluido que la TIE es una herramienta muy útil a la hora de recopilar información sobre el COVID-19, así como de sus diferentes fenotipos.

1. Introducción

En noviembre de 2019 una nueva cepa de coronavirus fue descubierta en Wuhan, provincia de Hubei, China. Aunque en un primer momento se pensó que se trataba de un caso puntual, la situación se descontroló rápidamente. Hoy, en octubre de 2021, más de 242 millones de casos positivos confirmados de COVID-19 y más de 4,9 millones de muertes han sido comunicados por la Organización Mundial de la Salud (OMS) [1].

La enfermedad, causada por el virus SARS-CoV-2, se transmite principalmente mediante el contacto de aerosoles respiratorios, pudiendo provocar síntomas como fiebre, tos seca, anosmia y ageusia [2]. Además, la infección también es detectable en Tomografías Computarizadas (TCs), sugiriendo alteraciones en el tejido pulmonar intratorácico (p. ej.: neumonía).

Debido a estas posibles alteraciones morfológicas o funcionales, diversas técnicas de imagen médica pueden ser de gran ayuda para el diagnóstico, tratamiento y

recuperación de pacientes infectados por COVID-19. Entre las diversas tecnologías conocidas hasta la fecha, la Tomografía de Impedancia Eléctrica (TIE) puede ser una de las más apropiadas para este tipo de aplicaciones, ya que se trata de una herramienta no invasiva, libre de radiación, de uso continuo seguro y rentable [3].

La TIE se basa en la inyección de débiles corrientes eléctricas alternas en el cuerpo del individuo, induciendo variaciones de voltaje en las barreras del cuerpo (piel) [4]. Estas variaciones son resultado directo de la conductividad eléctrica intrínseca de cada tejido corporal, las cuales varían de forma significativa en cuanto a los tejidos torácicos se refiere, ofreciendo así una mayor o menor resistencia al paso de la corriente. Dado que estas variaciones pueden llegar a ser mínimas y estar localizadas alrededor de todo el cuerpo, es necesario el uso de varios electrodos (8, 16 o incluso 32) que se colocan seccionalmente alrededor del cuerpo. Las mediciones captadas por estos electrodos son posteriormente utilizadas para reconstruir imágenes del interior del cuerpo humano y medir, entre otros parámetros, la compliancia pulmonar, término que hace referencia a la distensibilidad del tejido pulmonar.

Como las variaciones temporales en tejidos pulmonares son medibles mediante la TIE, se ha valorado la posibilidad de realizar una clasificación preliminar de diferentes fenotipos del COVID-19. La identificación de los fenotipos específicos en cada caso puede ayudar a determinar el tratamiento más adecuado para cada paciente, mejorando así su pronóstico. En 2020, Gattinoni *et al.* presentaron una primera clasificación de estos fenotipos basada en el reclutamiento alveolar, la cual es una técnica de reexpansión de áreas pulmonares previamente colapsadas mediante un incremento breve y controlado de la presión transpulmonar. Estos investigadores consiguieron distinguir los pacientes con un reclutamiento alveolar alto o tipo H, de los pacientes con un reclutamiento alveolar bajo o tipo L [5]. Mientras que los primeros presentaban una alta elastancia de los tejidos, una alta derivación pulmonar de derecha a izquierda, y un peso pulmonar alto, los últimos mostraron una baja

elastancia del tejido pulmonar, un índice ventilación-perfusión bajo y un bajo peso pulmonar.

La hipótesis principal de este trabajo es que los distintos fenotipos del COVID-19 pueden ser detectados a través de mediciones de compliancia pulmonar obtenidas mediante la técnica de TIE. La compliancia pulmonar se define como la “medición de la facilidad con la que se expanden los pulmones y el tórax durante los movimientos respiratorios, determinada por el volumen y la elasticidad pulmonar” [6] y ha resultado ser de utilidad a la hora de diagnosticar varias enfermedades (como pueden ser un enfisema, la fibrosis pulmonar o incluso edemas pulmonares). Teniendo todo esto en cuenta y para obtener información relevante en cuanto al COVID-19 y a sus diferentes fenotipos se refiere, realizamos una evaluación de los cambios de compliancia pulmonar regional de dos pacientes que fueron sometidos a estrategias ventilatorias diariamente durante al menos una semana. También se incluyeron en la investigación análisis de Gasometrías Arteriales (GA), cuyos resultados pueden ser utilizados para evaluar la eficiencia de los pulmones a la hora trasladar el oxígeno a la sangre y retirar el dióxido de carbono de esta, complementando así la información recogida mediante la TIE.

2. Metodología

2.1. Descripción de los sujetos y el protocolo

Para este estudio, se recogieron datos de cuatro pacientes infectados por COVID-19, dos de los cuales fueron excluidos por la insuficiente longitud de su seguimiento (inferior a 7 días). Por tanto, los sujetos analizados durante la investigación fueron dos pacientes de COVID-19 (de ahora en adelante, los pacientes A y B) profundamente sedados y mecánicamente ventilados, los cuales fueron evaluados durante al menos una semana. La Tabla 1 muestra las características de estos sujetos.

Paciente	A	B
Sexo	F	F
Edad (años)	67	81
Peso (kg)	90	80
IMC (kg/m ²)	29,7	31,2
Días evaluados	7	12
Comorbilidades	Hipertensión, hipotiroidismo	Hipotiroidismo

Tabla 1. Características de los pacientes analizados infectados por COVID-19.

En cuanto al protocolo respiratorio que se les fue aplicado a los sujetos, una maniobra de reclutamiento alveolar diaria fue llevada a cabo utilizando una Ventilación Controlada por Presión (VCP), seguida de un análisis volumétrico de la Presión Positiva Espiratoria Final (PEEP, *Positive End-Expiratory Pressure*).

La fase de reclutamiento constó de un aumento progresivo de la PEEP, siendo esta presión basal de 10 cmH₂O y subiendo escalonadamente hasta 25 cmH₂O utilizando

incrementos de 3 cmH₂O. En la PEEP más alta, la presión intratorácica máxima alcanzada fue de 40 cmH₂O. El mismo protocolo escalonado fue llevado a cabo para la fase de deflación, donde la presión basal de 10 cmH₂O fue alcanzada nuevamente. Teniendo en cuenta que la duración de cada parcial aplicado fue de 2 minutos, la duración de cada grabación resultante es de aproximadamente 20 minutos. Adicionalmente, una GA fue llevada a cabo antes y después de cada maniobra respiratoria, facilitando así el análisis del efecto de las maniobras mediante el cálculo de las presiones parciales de oxígeno (pO₂) y dióxido de carbono (pCO₂) en sangre.

La recopilación de los datos fue llevada a cabo en Hungría, mientras que el análisis de estos fue realizado en Alemania. Cabe destacar que la investigación cumple con todas las regulaciones locales relacionadas con el uso de datos clínicos y las políticas institucionales, y fue realizada de acuerdo con los principios de la Declaración de Helsinki. Además, el ensayo clínico fue aprobado por el Comité de Revisión de Investigación Humana de la Universidad de Szeged (número de aprobación 37/2020-SZTE), y fue registrado en *clinicaltrials.gov* bajo el número de referencia NCT04360837. La documentación de conformidad pertinente fue recopilada por parte de los pacientes o sus respectivos representantes legales.

2.2. Adquisición de los datos de la TIE

Los datos relacionados a la conductividad eléctrica fueron captados mediante el dispositivo PulmoVista 500, de Dräger (Dräger, Lübeck, Alemania). De acuerdo con los estándares de posicionamiento de los electrodos, un cinturón de 16 electrodos se posicionó entre el 5° y 6° espacio intercostal, formando así un plano transversal alrededor del tórax. Las inyecciones de corriente eléctrica y las consiguientes mediciones de los voltajes resultantes fueron llevados a cabo utilizando electrodos adyacentes a una frecuencia de muestro de 50 Hz, obteniendo así una alta resolución temporal.

La reconstrucción de las imágenes derivadas de los voltajes eléctricos se llevó a cabo mediante el algoritmo *time-difference GREIT* [7].

2.3. Construcción de las imágenes de compliancia pulmonar

Las imágenes de compliancia pulmonar se basan en el procesamiento de imágenes funcionales de TIE (fTIE), concretamente, en imágenes funcionales de variación tidal (fTIE-VT), las cuales ofrecen información relacionada con los cambios de volumen entre el final de inspiración y del final de espiración durante ciclos respiratorios normales. Con el fin de obtener estas imágenes fTIE-VT, el primer paso fue el de determinar el contorno de los pulmones en las imágenes, de manera que las mediciones estuviesen centradas únicamente en el tejido pulmonar. Para ello, un valor umbral del 20 % fue aplicado en las imágenes fTIE-VT. Como los contornos pulmonares obtenidos para cada valor de PEEP variaban ligeramente los unos de los otros, el contorno final fue calculado mediante la superposición de todos los contornos generados por día y paciente. Finalmente, las fTIE-VT fueron obtenidas mediante el

cálculo de la diferencia entre los puntos inspiratorios y espiratorios finales de la señal de ventilación global.

Considerando la gran correlación que existe entre las variaciones tidales locales y las variaciones de las impedancias intratorácicas [8], la compliancia pulmonar por pixeles basada en la TIE puede calcularse mediante la siguiente expresión:

$$C_{px} = \frac{\Delta Z}{P_{meseta} - PEEP}$$

Donde ΔZ es la variación de la impedancia intratorácica y P_{meseta} es la presión meseta dentro de los conductos respiratorios. Una vez aplicada esta relación a cada pixel de las imágenes fTIE-VT, los mapas de compliancia pulmonar basada en TIE pueden ser visualizados.

Finalmente, para poder evaluar la reacción de los pacientes a las maniobras de reclutamiento alveolar, se calcularon las diferencias entre los mapas de compliancia provenientes de los PEEP iniciales y finales substrayendo las matrices correspondientes.

3. Resultados

Los mapas de compliancia presentan claras diferencias entre las evoluciones de ambos pacientes, sugiriendo que la reacción a las maniobras respiratorias del paciente B fue mejor que la del paciente A.

En el caso del paciente A, los mejores valores de compliancia se muestran el primer día de ensayo, seguido de un progresivo deterioro que termina con una ligera mejora durante el último día (Figura 1). En términos de presiones parciales de gases en sangre arterial, el paciente A sufrió un aumento progresivo de los valores de pCO_2 y una drástica disminución en los valores de pO_2 (Figura 2), hasta llegar al punto en el que no se aprecia mejora alguna en la oxigenación (día 6).

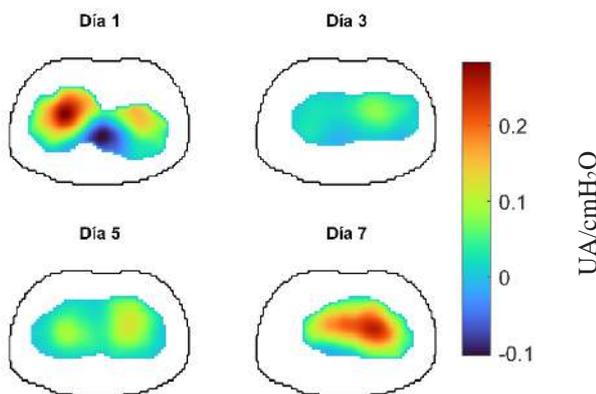


Figura 1. Resultados de los mapas de compliancia pulmonar basada en la TIE del paciente A. (UA: Unidad Arbitraria)

Por otro lado, se puede apreciar que los valores de compliancia resultantes del paciente B son generalmente más altos, los cuales siguen una tendencia más homogénea con el paso de los días (Figura 3). Además, no existe una clara tendencia en cuanto a los valores de la pO_2 obtenidos de las GAs, aunque los datos de los últimos cuatro días sugieren un claro deterioro fácilmente identificable. Además, los valores de la pCO_2 aumentan gradualmente

durante el paso de los días (Figura 4), lo cual es señal de una pobre ventilación pulmonar.

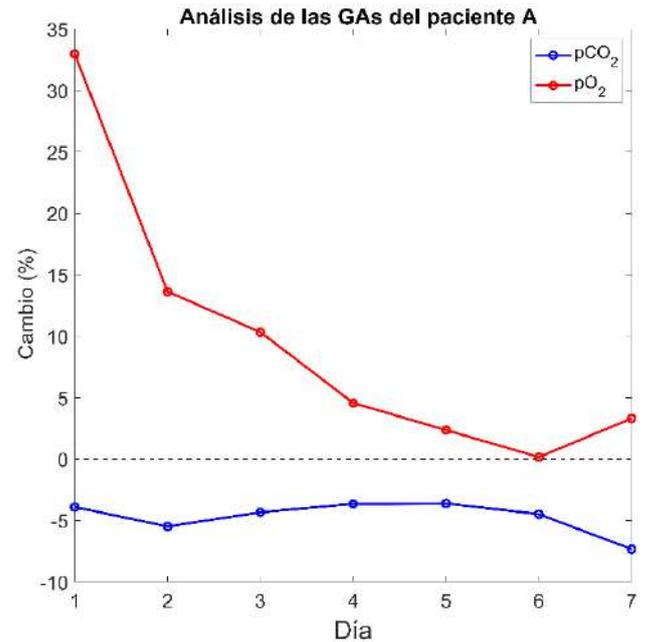


Figura 2. Evolución de los valores de la GA del paciente A.

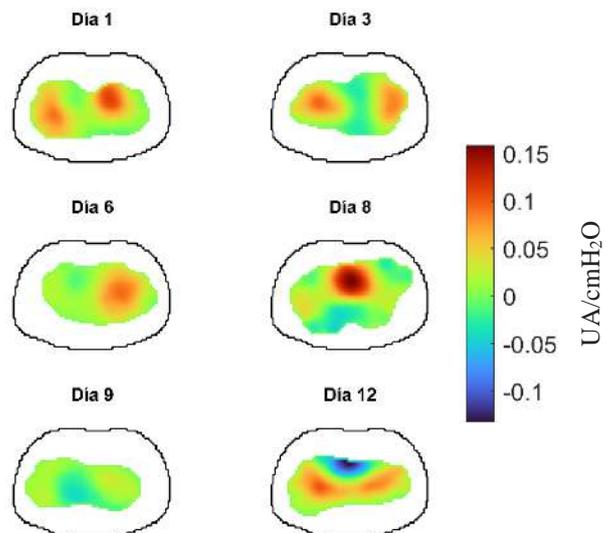


Figura 3. Resultados de los mapas de compliancia pulmonar basada en la TIE del paciente B.

4. Discusión

A lo largo de este trabajo se ha recopilado el proceso que se ha llevado a cabo para el análisis de la evolución de la compliancia pulmonar en pacientes infectados por COVID-19. Debido a la relativa novedad de la enfermedad, esta investigación es una de las pioneras en el ámbito, la cual se considera que será de gran ayuda para la toma de decisiones clínicas en el futuro.

En el estudio encontramos que los dos pacientes que fueron estudiados reaccionaron de manera diferente a las maniobras de reclutamiento alveolar, apreciable en la diferencia entre los valores de la Figura 1 y la Figura 3. Esta diferencia sugiere que el paciente A podría pertenecer al tipo L debido a la baja compliancia regional que presenta y, por otro lado, que el paciente B se podría clasificar como

tipo H, ya que la compliancia pulmonar que presenta es más alta y generalmente más estable. Es digno de destacar, además, que la clasificación preliminar presentada en este trabajo concuerda con la valoración realizada por el personal médico responsable de proveer los datos de los pacientes. Por otro lado, también cabe mencionar la posible transición desde el fenotipo L al fenotipo H del paciente A, conclusión basada en el cambio de compliancia apreciable en la Figura 1 (la compliancia del día 1 es considerablemente más alta que la de los siguientes días). En cuanto al paciente B, se sospecha un posible desajuste ventilación-perfusión basado en la deterioración progresiva de los resultados de las GAs y los valores mantenidos de compliancia pulmonar. En base a estos valores, los cuales aportan información sobre la ventilación pulmonar, se sugieren posibles disfuncionalidades en la perfusión alveolar.

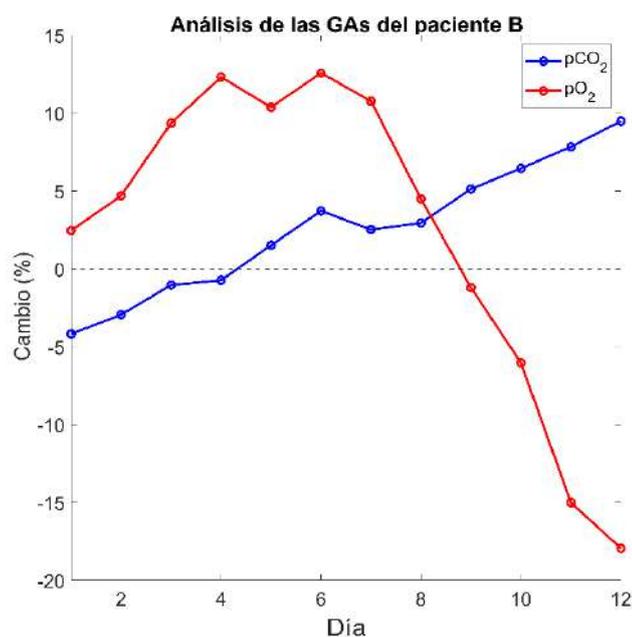


Figura 4. Evolución de los valores de la GA del paciente B.

Somos conscientes de que el presente trabajo tiene algunas debilidades. Por un lado, aunque el seguimiento realizado a los pacientes analizados en esta investigación fuese aceptable, la cantidad de sujetos analizados es demasiado bajo. Con el fin de obtener resultados estadísticamente significativos, sería necesario monitorizar más pacientes de esta enfermedad.

Además, el uso de un solo parámetro (p. ej.: compliancia pulmonar) para clasificar diferentes fenotipos del COVID-19 es posiblemente insuficiente. La combinación de distintos parámetros obtenidos mediante esta técnica de imagen puede incrementar la precisión de esta clasificación en el futuro.

Sin embargo, aunque la TC sea considerada el “gold standard” en cuanto al diagnóstico del COVID-19, se ha comprobado que la TIE es una opción prometedora para monitorizar posibles cambios en la infección, ya que ha sido comprobada la rapidez con la que esta evoluciona en el tiempo.

La recopilación de nuevos datos de TIE, así como la combinación de diferentes parámetros de medición y la utilización de técnicas más avanzadas de análisis de datos, ayudarán a comprender mejor la evolución de la enfermedad en diferentes tipos de pacientes, a optimizar la toma de decisiones clínicas en éstas y a mejorar su pronóstico.

5. Conclusiones

La TIE es una herramienta válida para la obtención de información relevante sobre los pacientes infectados por COVID-19 y sus respectivos tipos de progresión temporal (tipo L y tipo H), haciendo posible la monitorización y evaluación del comportamiento pulmonar regional en vivo, sugiriendo posibles transiciones entre fenotipos, e incluso planteando posibles desajustes entre la ventilación y perfusión de los tejidos respiratorios.

Agradecimientos

Este proyecto ha recibido financiación por parte del Plan de Acción de la Región Europea para la Movilidad de Estudiantes Universitarios (ERASMUS+, subvención 2019-1-ES01-KA103-060931), el Gobierno Vasco, el Ministerio Federal de Educación e Investigación Alemán (MOVE, subvención 13FH628IX6), y el marco de financiación Horizon 2020 MSCA RISE (#872488 DCPM).

Referencias

- [1] WHO, ‘WHO Coronavirus (COVID-19) Dashboard’, World Health Organization, 2021. [En línea]. Disponible en: <https://covid19.who.int/> [Accedido: 22-oct-2021].
- [2] R. Miller and K. Englund, ‘Clinical presentation and course of COVID-19’, *CCJM*, vol. 87, no. 7, pp. 384-388, Jul. 2020, doi: 10.3949/ccjm.87a.ccc013.
- [3] B. Schulleke, Bo Gong, and K. Moeller, ‘Steps towards 3D Electrical Impedance Tomography’, Milan, Aug. 2015, pp. 5323-5326, doi: 10.1109/EMBC.2015.7319593.
- [4] I. Frierichs et al., ‘Chest electrical impedance tomography examination, data analysis, terminology, clinical use and recommendations: consensus statement of the TRanslational EIT developmeNt stuDy group’, *Thorax*, vol. 72, no. 1, pp. 83-93, 2017, doi: 10.1136/thoraxjnl-2016-208357.
- [5] L. Gattinoni et al., ‘COVID-19 pneumonia: different respiratory treatments for different phenotypes?’, *Intensive Care Med*, vol. 46, pp. 1099-1102, Jun. 2020, doi: 1.1007/s00134-020-06033-2.
- [6] “Compliance pulmonar”, *Diccionario Médico de la Clínica Universidad de Navarra*, 2020. [En línea]. Disponible en: <https://www.cun.es/diccionario-medico/terminos/compliance-pulmonar> . [Accedido: 18-oct-2021].
- [7] A. Adler et al., ‘GREIT: a unified approach to 2D linear EIT reconstruction of lung images’, *Physiol. Meas.*, vol. 30, no. 6, pp. S35-S55, Jun. 2009, doi: 10.1088/0967-3334/30/6/S03.
- [8] E. L. V. Costa et al., ‘Bedside estimation of recruitable alveolar collapse and hyperdistension by electrical impedance tomography’, *Intensive Care Med*, vol. 35, pp. 1132-1137, Mar. 2009, doi: 10.1007/s00134-009-1447-y.

A Virtual Reality Application for Training Medical Students in a Pneumothorax Procedure

I. Méndez de Vigo Iriarte¹, J. A. Santos Paz¹, R. García Carmona¹

¹Escuela Politécnica Superior, Universidad CEU San Pablo, Madrid, Spain
{i.mendezvigo,joseangel.santospaz@usp.ceu.es}
rodrigo.garciacarmona@ceu.es

Abstract

A virtual reality application for training medical students in a pericardiocentesis procedure was developed and tested in Universidad CEU San Pablo. This document describes the extension of this application to add a pneumothorax procedure and a diagnosis module. The combination of both procedures creates a more challenging environment, where the user must make the correct diagnosis before deciding how to proceed. After making the decision, the user interacts with the virtual environment to treat the patient in a way similar to a real procedure. During this process, the user obtains real time feedback that enables them to learn from their mistakes. This work establishes the foundations for the development of a generic virtual reality framework for developing training scenarios for medical students, focusing more specifically in learning the steps of each procedure and making a correct diagnosis.

1. Introduction

Medical errors are preventable adverse effects of healthcare that could be harmful to the patient. A recent research states that such errors are among the most common causes of injury or death in the United States. Most medical errors occur as a failure to follow the right procedure for identifying and treating an illness. Loss of reputation can make healthcare professionals reluctant to admit making these mistakes. Consequently, since the medical errors are not properly documented, this can lead to an evolving cycle of committing the same misstep [1].

The performance of clinicians during surgical procedures depends widely on both their knowledge of the steps involved in it and their surgical skills (mostly hand-eye coordination and fine motor skills). Knowledge is acquired by the process of studying and understanding reliable, accurate and current information regarding the diagnosis and treatment of a particular disease. On the other hand, skills require surgical practice to be acquired [2].

Clinical skills could be considered as the key to health education. However, students sometimes complete their educational programs armed with theoretical knowledge but without many of the clinical skills vital for their day-to-day work. To solve this issue, simulators could provide the training scenarios required to obtain the necessary competences to become an accomplished clinician.

Medical simulators involve the use of technology to replicate clinical scenarios which can provide their users with experimental learning. Experimental learning is an active process during which the learner constructs new

knowledge by linking new information and new experience with previous knowledge and understanding. One technology that is particularly suited for accurately simulating medical situations is virtual reality (VR) [3,4].

Several medical procedures are particularly suitable to be simulated by VR simulators, and some of them, as the pneumothorax procedure, are related to the respiratory system.

The respiratory system main functionality is to provide oxygen to the human body and remove carbon dioxide produced in the internal respiration. The respiratory system is composed of the airways, the lungs, and the blood vessels. The airways are composed of the nose and nasal cavity, the sinuses, the mouth, the pharynx, the larynx, the trachea, the diaphragm, the lungs, the bronchial tubes/bronchi, the bronchioles, and the alveoli.

Inside the lungs, the exchange of oxygen and carbon dioxide between the capillaries and the alveoli is produced due to the Henry's law and the difference between the partial pressure between both gases in the capillaries and in the alveoli generated through the respiration.

The lungs are located on the chest cavity and are protected by the rib cage. To reduce any friction that could be produced between the lungs, rib cage and chest cavity during respiration, the lungs are surrounded by the pleura. The pleura is made up of a two-layers membrane separated by a viscous lubricant denominated pleura fluid [5].

There are two main types of pressures involved in the respiration, the intraalveolar pressure and the intrapleural pressure. The contraction and relaxation of the muscles involved in the respiration expand and contract the lungs. This increment and decrement in the volume of the lungs modifies the pressures involved in the respiration that force the air to be sucked into and pulled out of the lungs.

The collapse of a lung receives the name of pneumothorax. This collapse is produced by the entrance of air in the pleural space. The passage of air into the pleural space creates an increment of the intrapleural pressure that forces the lung to decrease its volume. Therefore, the intraalveolar pressure will increase and the air will be pushed out of the lung. In a normal situation, the contraction of the diaphragm will produce a negative intraalveolar pressure that allows the air flow into the lungs. But with a collapsed lung the contraction of the diaphragm would not be enough, and the lung loses its functionality.

To be able to properly design a simulation of a pneumothorax procedure, both its symptoms and the treatment that should be used must be properly understood.

The most common symptoms of a pneumothorax are sudden chest pain and shortness of breath. However, there are also other symptoms: as one lung will be collapsing less oxygen will enter the body, producing a decrement of oxygen saturation that can be detected by a pulse oximeter. Also, the respiratory sounds in the collapse lung will disappear. Regarding the cardiovascular system, blood pressure decreases and tachycardia occurs due to the increase in the speed of blood flow. With that said, a thorough diagnosis is still needed to verify that the patient suffers a pneumothorax. To properly identify it among other respiratory conditions, a technique to observe the lungs, such as an examination using ultrasound or X-rays, is still needed.

The treatment of a pneumothorax depends on its type and magnitude. In a minor pneumothorax a clinical intervention it is not usually required, while in more severe cases the air or liquid accumulated in the pleural space is needed to be drained. There are distinct technics to drain the fluid of the pleural cavity such as needle aspiration or chest tube insertion.

2. Materials and Methods

The aim of this work is to develop a VR application that faithfully reproduces the diagnosis and treatment of a patient with pneumothorax. This application will be used to train medical students and, therefore, can potentially have a sensible impact on the development of its user's aptitudes.

To develop and/or use this VR application, several components are required, both hardware and software. The hardware components, used both for development and actual usage, can be classified as input devices, used to interact with the virtual environment, output devices, that send feedback to the user depending on the information received from the input devices, and the actual computers required for running the VR software. Such computers could be part of the VR device itself or a powerful external personal computer.

A brief study to select the most appropriate VR hardware devices to fulfil the stated objectives was made. It concluded with the selection of the HTC Vive Pro Head Motion Display (HMD) device. One of the main reasons to select this HMD is its ability to track an area of up to 4.6 x 4.6 square meters, a condition that makes it ideal for our scenario, since the virtual emergency room must be big enough to let the user work on the procedure. Furthermore, this HMD features other attractive characteristics, as an ample field of view (110°), a resolution of 2880 x 1600 pixels and a weight of only 555g.

In addition to the HMD, other hardware components were required: a pair of HTC base stations, that work as beacons, and a pair of controllers. The HTC base stations, also called lighthouses, were therefore used for tracking both the HMD and the controllers.

Regarding the software, the final VR application will be a software component on itself. To be able to run it, a computer also needs SteamVR as an additional software component, since the application is compiled to work with the OpenVR library.

To develop the application, several software components were used, of which the most important were the VR modeling tools and the VR development tools. The former are used to create the 3D structures that represent the virtual objects. Tinkercad was used. The latter are used to define and implement the features of virtual environments [6]. Unity 2019.4.12f1 was selected. Unity is a platform that enables the creation of interactive content in real time. Other applications were also used in a minor way, as Photoshop, to create stains for the virtual objects.

All the elements that comprise the VR environment and are used to build the application receive the name of assets. They can be 3D models, image or sound files, among others. In Unity, all these assets are grouped in GameObjects. A single GameObject can be composed of different assets or even other GameObjects that realize different shapes, functionalities, or properties established by the components and the scripts attached to them.

Most of the assets used in Unity were extracted from a previous work, in which the author developed a virtual reality application for training medical students in a pericardiocentesis procedure. From this work, the hospital room, the patient and all the medical tools used in both procedures (pericardiocentesis and pneumothorax) were reused. Nevertheless, some utensils to treat the patient such as the scalpel, the chest tube or the male-to-male connector needed to be modeled from scratch. These tools were built by creating similar 3D shapes in Tinkercad. This tool allows for the combination of simple hollow or filled polyhedrons to obtain complex structures.

Once all the instruments were implemented and set up in their proper position inside the corresponding scene, several scripts were developed. Their aim is to define the functionality of each GameObject that could correspond to a tool required for the diagnosis or treatment of the patient. Some examples are an electrode, a syringe or a cloth.

Also, several scripts were created. They enable the user to grab or drop the different medical utensils by holding or releasing the trigger of the controllers. These controllers are shown as two human hands in the virtual scene. All of the GameObjects that realize these instruments must contain, in addition to their corresponding scripts and components, a collider, a Rigidbody and the "Pickupable" tag, which defines which objects can be grab and used. To let the user know that a particular object can be grabbed, a green aura is show when a controller approaches such utensil.

To simulate the interaction between the patient and the medical equipment, more than forty scripts were programmed in C#, which provided realistic feedback. As an example, a scalpel used to make an incision was programmed to show a stain created with Photoshop as soon as the user holds the scalpel on the area of the virtual patient that corresponds to the inner thoracic space.

3. Results

The virtual reality application is structured around several scenes, as is usual in Unity-based development. A scene is a virtual environment where the different virtual elements are displayed and can be interacted with.

3.1. The main menu scene

The evaluation scene, which also works as the starting scene, is composed of a canvas that will be displayed only on the monitor of the computer, to allow the professor that is supervising the activity to select a clinical situation without letting the student know which is. This canvas will just contain a text and a pair of buttons.

The buttons, that can be selected with the mouse of the computer, are disposed on two different GameObjects, one indicating that is a pneumothorax procedure with an image and a text, and another one doing the same for a pericardiocentesis procedure (see Figure 1).



Figure 1: Evaluation Scene.

3.2. The Pneumothorax scene.

This is the main scene of this application. At the start of this scene, the user can see the patient lying in the hospital bed at the center of the room and all the equipment that can be used for diagnosing and treating the pneumothorax disposed around it on trolleys (see Figure 2; **Error! No se encuentra el origen de la referencia.**).

The user is expected to start the diagnosis of the patient immediately, since its condition will worsen with time. The application is programmed so the user has a 10-minute time limit to finish with the whole diagnosis and procedure.

For the diagnosis of the pneumothorax, the user can use three electrodes (for an Electrocardiogram signal), a blood pressure cuff, a pulse oximeter, an stethoscope and two different medical imaging techniques: X-ray or ultrasound



examination. Though the utilization of these tools the user receives specific feedback that helps them identify the condition of the patient. For example, by using the electrodes, the blood pressure cuff and the pulse oximeter, the patient's vital signs can be measured. They are shown through different videos projected on a monitor placed next to the patient.

Before beginning the treatment, the user must use several menus disposed in the wall farthest from the patient (used as a kind of virtual interactive board) to determine the diagnosis that have been performed and to select the treatment that is going to be used for a posterior evaluation of its performance.

Regarding the treatment, there are distinct techniques to drain the fluid in the pleural cavity such as needle aspiration or chest tube insertion. Depending on the type of treatment selected, the user must use the correct virtual equipment and follow the proper steps.

With needle aspiration the user is expected to introduce a catheter into the pleural space, drain the accumulated fluid using a needle and expel that fluid to the environment through a stopcock. The catheter will be connected to the needle with a stopcock. In this procedure the user must use a syringe to introduce the catheter, introduce the catheter and utilize the last syringe, which can be distinguished easily by a label, to drain and expel the accumulated air in several steps.

With chest tube insertion the user is expected to insert a chest tube into the patient, connect it to a draining machine and make an incision at the same level of the nipple past the anterior axillary line between the ribs. This is done to ensure that the chest tube is not placed too low and that it reaches the pleural space.

Once the incision is done, the clamp is used to dissect into the pleural space and hold the dissection to facilitate the introduction of the chest tube into the pleural cavity. As the chest tube could be displaced with the involuntary movements of the patient, it will also be immobilized to the skin using a suture.

Finally, the chest tube will be connected to the collection device through a male-to-male connector that will collect the excess fluids. Usually, the chest tube is kept inserted for two days, until there is no further leakage of fluid or air in the pleural space. Finally, a chest radiograph is performed and, if there is not any reaccumulating of fluid, the chest tube is removed.



Figure 2: Hospital Room.

Both procedures require the sterilization of all the utensils used during the treatment of the patient. Also, the field of the patient exposed to the treatment should be sterilized. However, this sterilization process is not present in the virtual procedure, since the user can assume that it has happened beforehand. However, the patient must be covered with a surgical cloth that does not cover the landmarks needed for the treatment as the nipple. The nipple is an important landmark because it determines the fourth inner thoracic space. This is a process that the user must perform.

To guarantee the safety and comfort of the patient a local anesthetic should also be injected. This must be done in all the cases that require a surgical procedure.

After both the diagnosis and the treatment of the patient are finished, the user presses a button to indicate as such. This action creates a canvas that shows the user if the diagnosis was performed correctly or not, and a comparison between the instruments that were and were not required and the ones that the user employed. This lets the user learn from their mistakes.

3.3. The Pericardiocentesis Scene

The last step of this work was its integration with an already existing virtual reality application, similar to this one, and initially developed for training medical students in a pericardiocentesis procedure. By putting together both applications, the user does not know which ailment the patient is suffering and must do a proper diagnosis.

This task required the expansion of the diagnosis of the pericardiocentesis procedure simulator, to include diagnosis based on imaging systems (X-ray and ultrasound). This is important to make both diagnosis procedures (pericardiocentesis and pneumothorax) indistinguishable, with all the components disposed in the same way on both scenes, even if they were not required. This fact ensure that the diagnosis should be performed correctly to proceed with an accurate treatment.

4. Discussion

The aim of this work was to develop a VR application for training medical students in the diagnosis and treatment of a pneumothorax, and its integration with an already existing application for the treatment of a pericardiocentesis.

The use of this application could potentially help medical students to internalize the steps that should be taken in these procedures and reduce the abundance of medical errors that are currently being made. However, it is important to note that this is a hypothesis that should be tested with real students in a pilot program, which is programmed to be run shortly after the publication of this work.

Additionally, the VR application is putting the users of this application in a position to make decisions that directly affect the health of the patient, albeit virtual. We hope that this simulated sense of responsibility could make them

internalize the responsibilities they will have to assume in their future medical practice [7].

Concerning the usability of this application, we are planning to conduct a thorough testing using standard usability questionnaires [8].

Finally, in the same way that this application extended and was integrated with an already existing one (pericardiocentesis), we believe that this work could in turn lay the foundations for a virtual reality application suite with many procedures. This way, medical education professional could expose their students to a simulation of a real hospital without risking the safety of real patients.

5. Acknowledgements

The development of this project would not have been able without the installations and equipment supplied by the Universidad CEU San Pablo.

Also, Dr. Alberto Rubio López deserves a special mention for his help during the development of this application, as well as for supplying several clinical resources to extract information relevant to the development of this work.

This research was funded by the Ministry of Science, Innovation, and Universities of Spain and by the European Regional Development Fund of the European Commission under grant number RTI2018-097122-A-I00.

6. References

1. Rodziewicz TL, Houseman B, Hipskind JE. Medical Error Reduction and Prevention. [Updated 2021 Jan 4].
2. Kneebone, R. (2003). Simulation in surgical training: educational issues and practical implications. *Medical education*, 37(3), 267-277.
3. Al-Elq, A. H. (2010). Simulation-based medical teaching and learning. *Journal of family and Community Medicine*, 17(1), 35.
4. Pottle, J. (2019). Virtual reality and the transformation of medical education. *Future healthcare journal*, 6(3), 181.
5. Charalampidis, C., Youroukou, A., Lazaridis, G., Baka, S., Mpoukovinas, I., Karavasilis, V., ... & Zarogoulidis, P. (2015). Pleura space anatomy. *Journal of thoracic disease*, 7(Suppl 1), S27.
6. Oluleke Bamodul and Xuming Ye¹, Virtual Reality and Virtual Reality System Components, Proceedings of the 2nd International Conference on Systems Engineering and Modeling (ICSEM-13), Published by Atlantis Press, Paris, France.2013.
7. Riva, G., & Wiederhold, B. K. (2015). The new dawn of virtual reality in health care: medical simulation and experiential interface. *Annual Review of Cybertherapy and Telemedicine* 2015, 3-6.
8. Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *Intl. Journal of Human-Computer Interaction*, 24(6), 574-594.

Reconstrucción en 3D del microambiente inmunológico de tumores de cáncer pancreático

C. Almagro-Pérez¹, A. Kiemen², P. Wu², A. Braxton³, L. Nicolás-Sáenz^{1,4}, L. Wood³, R. Hruban³, D. Wirtz^{2*}, A. Muñoz-Barrutia^{1,4*}

¹ Departamento de Bioingeniería e Ingeniería Aeroespacial, Universidad Carlos III de Madrid, Leganés, España.

² Department of Chemical and Biomolecular Engineering, The Johns Hopkins University, Baltimore, Maryland, USA.

³ Department of Pathology, The Sol Goldman Pancreatic Cancer Research Center, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA.

⁴ Instituto de Investigación Sanitaria del Hospital Gregorio Marañón, Madrid, España.

* Correspondencia: wirtz@jhu.edu, mamunozb@ing.uc3m.es

Resumen

La composición del microambiente inmunológico es crucial para el desarrollo del cáncer, entre ellos del cáncer pancreático, desgraciadamente uno de los más letales [1]. Sin embargo, su entendimiento sigue siendo escaso ya que los estudios realizados hasta el momento se basan en una única sección histológica, a partir de la cual es imposible inferir la alta complejidad e inherente estructura tridimensional del microambiente del tumor. Utilizando ~1,000 histologías seriadas teñidas de manera alterna, así como métodos de aprendizaje profundo y técnicas de procesamiento de imagen, este trabajo establece por primera vez una estrategia para el estudio del microambiente inmune del tumor en 3D y a resolución unicelular. Asimismo, incluye un estudio exhaustivo de la inflamación tumoral en 3D. El desarrollo de mapas inmunes tridimensionales revela una alta heterogeneidad de la inflamación, ilustrando que esta varía hasta en tres órdenes de magnitud entre unas regiones y otras del tumor. Como apuntan los análisis realizados, este estudio presenta una herramienta capaz de proporcionar información valiosa sobre el entorno del tumor que podría ayudar en el desarrollo de nuevas inmunoterapias para el cáncer de páncreas.

1. Introducción

El cáncer pancreático es uno de los tipos más letales de cáncer. La forma más común de cáncer pancreático es el adenocarcinoma pancreático ductal, el cual supone el 90% de los casos. Este cáncer tiene una desoladora tasa de supervivencia a cinco años inferior al 6% ya que la mayoría de los pacientes (~80%) son diagnosticados demasiado tarde con la enfermedad localmente avanzada o metastásica [2].

La neoplasia intraepitelial pancreática (PanIN) es una lesión precursora de este cáncer. Una célula pancreática con una mutación inicial tarda aproximadamente dos décadas hasta que se convierte en un cáncer pancreático metastásico. Esta es la razón del creciente interés en el estudio de estas lesiones para frenar el desarrollo del cáncer pancreático [3].

Asimismo, los tratamientos tradicionales de cáncer tales como la quimioterapia, la radioterapia o la cirugía no han

mejorado significativamente la supervivencia de pacientes con cáncer pancreático. Por esta razón, en los últimos años, ha habido un rápido aumento en el estudio de estrategias basadas en inmunoterapia para abordar el cáncer de páncreas, ya que, además, este tipo de tratamientos han demostrado ser altamente efectivos para otros tipos de cánceres, como el cáncer de mama [3].

Hasta el momento, el estudio del microambiente del tumor y su inflamación se ha enfocado principalmente en el análisis de entre 2 y 5 secciones histológicas por paciente [4]. Sin embargo, aspectos como la conectividad de los tejidos, la distribución de las células o la organización de estructuras biológicas complejas tales como la vasculatura tumoral son difíciles, o incluso imposibles, de ser inferidos a partir de estudios realizados en dos dimensiones. Estos hechos son ilustrados en la Figura 1.

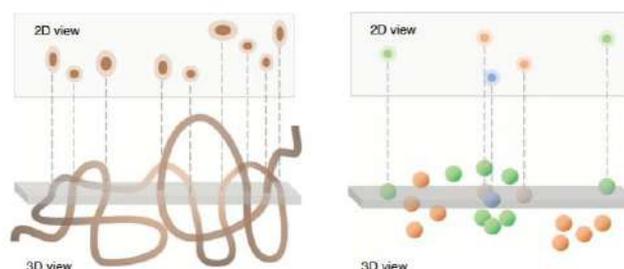


Figura 1. Las vistas en secciones transversales en 2D pueden inducir a error sobre la conectividad de los tejidos, la organización de estructuras biológicas complejas (Izqda.) y la distribución celular (Dcha.) (Imágenes extraídas de [5]).

2. Materiales

Para el desarrollo de este proyecto, 1.145 cortes histológicos fueron proporcionados por el departamento de Patología del Hospital Johns Hopkins. Las secciones de tejido se obtuvieron a partir de cortes en serie de una muestra de páncreas extirpado que contenía lesiones PanIN. La muestra tenía un tamaño de ~ 2.5 cm x 2.5 cm x 0.5 cm.

Las secciones se tiñeron siguiendo un patrón alternante de tres tinciones (véase Figura 2). La primera tinción es Hematoxilina y Eosina (H&E), la cual tiñe los núcleos de las células en azul y el citoplasma y las proteínas extracelulares en rosa. La segunda es la tinción de inmunohistoquímica CD45, que tiñe el citoplasma de los leucocitos (tipo de célula inmune) en marrón. Por último, la tinción de inmunohistoquímica CD3/Foxp3 fue utilizada para la identificación de otros dos tipos de células inmunes, las células tipo T (CD3+) y tipo T regulatorias (CD3+ y Foxp3+).

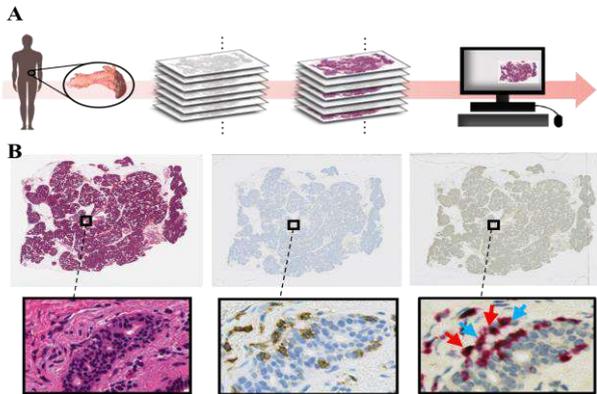


Figura 2. A Proceso para la digitalización de tejidos: Cortes seriados, tinción alternante y escaneo a alta resolución. B (Izqda.) Histología de Hematoxilina y Eosina (H&E). (Centro.) Histología de inmunohistoquímica CD45. El citoplasma de los leucocitos aparece en marrón. (Dcha.) Histología de inmunohistoquímica CD3/Foxp3. La flecha azul indica las células T (citoplasma rosa) y la flecha roja las células T regulatorias (citoplasma rosa y núcleo negro).

3. Métodos

Para reconstruir en 3D el microambiente del tumor a partir de las histologías descritas en la sección anterior, se siguió el proceso ilustrado en la Figura 3. Todo el proyecto se realizó empleando el software MATLAB.



Figura 3. Proceso para la reconstrucción en 3D del microambiente de tumores pancreáticos.

3.1. Reconstrucción 3D del microambiente inmune

Detección de células inmunes en imágenes IHC

En las imágenes IHC, las tinciones de anticuerpos (CD45, CD3/Foxp3) van acompañadas de hematoxilina para generar contraste (véase en la Figura 1B el núcleo azul de las células generado por esta tinción secundaria). Por ello, en primer lugar, los canales de hematoxilina y anticuerpos se separaron utilizando deconvolución de color siguiendo el método presentado en [6]. Seguidamente, los canales resultantes se suavizaron utilizando un filtro gaussiano

pasa bajos ($\sigma = \sqrt{2}$). A continuación, se localizó el núcleo de todas las células utilizando el canal de hematoxilina y un algoritmo que detecta máximos locales de intensidad en 2D de un determinado diámetro [7].

Finalmente, se construyeron histogramas conjuntos considerando la intensidad de las células en el canal de hematoxilina y el canal de anticuerpo. El algoritmo K-medoids [8] se utilizó para separar las células inmunes del resto de células del tejido. Este algoritmo se empleó para la detección de los tres tipos de células inmunes consideradas en este proyecto (leucocitos, células T y células T regulatorias).

Clasificación de tejidos en imágenes H&E

El siguiente paso para la reconstrucción en 3D del microambiente inmunológico de tumores pancreáticos es la identificación de los distintos tejidos presentes en el páncreas (incluido PanIN, las lesiones precursoras al cáncer pancreático). Con este objetivo, las imágenes H&E se separaron del resto de imágenes de la muestra (470 imágenes del total de 1.145). Seguidamente, siete de estas imágenes se extrajeron para anotar manualmente los tejidos empleando el software ImageScope (Leica Biosystems). En cada una de estas imágenes, se realizaron 50 anotaciones por cada uno de los siete tipos de tejidos pancreáticos presentes: grasa, colágeno, acinos exocrinos, islotes de Langerhans, músculo liso, epitelio ductal y PanIN.

A continuación, se empleó el método CODA, un algoritmo de deep learning basado en segmentación semántica desarrollado previamente por el grupo del Prof. Wirtz [9]. Para el entrenamiento del algoritmo, se crearon un total de 6.480 imágenes de 256 píxeles x 256 píxeles x 3 píxeles a partir de anotaciones superpuestas. Para la clasificación, las imágenes H&E de dimensiones 12,500 píxeles x 12,500 píxeles se dividieron en subimágenes de 256 píxeles x 256 píxeles.

Registro de imágenes y reconstrucción en 3D

Durante la adquisición de las secciones de tejido, estas se estiran, pliegan y dividen de manera impredecible generando una distorsión irregular entre los planos. Para alinear las 1145 imágenes y reconstruir digitalmente el volumen de tejido se desarrolló un método de registro de imágenes no lineal basado en intensidad.

Para realizar el registro, el canal común a los tres tipos de imágenes (canal hematoxilina) se obtuvo mediante deconvolución de color y la imagen central del bloque de secciones fue designada como la referencia de alineamiento. Cada imagen se registró con la imagen contigua más próxima al centro de acuerdo con el ángulo de rotación y el valor de translación que maximizaba la correlación cruzada entre las dos imágenes. A continuación, dicho registro global fue seguido de un registro elástico. Como puede observarse en la Figura 4, la imagen se dividió en imágenes de menor tamaño (250 x 250 píxeles) y en cada una de estas regiones se aplicó el registro afín. Finalmente, los mapas de desplazamiento de estas imágenes de menor tamaño se combinaron para crear mapas de desplazamiento de las imágenes enteras.

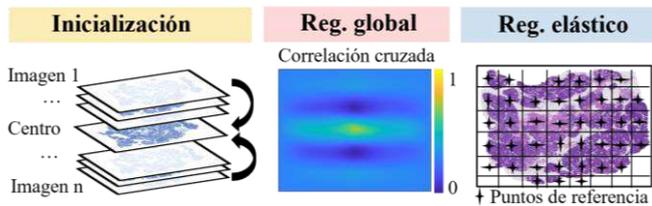


Figura 4. (Izqda.) La imagen central del bloque de tejido se utiliza como referencia. (Ctro.) El máximo de la correlación cruzada entre las imágenes proporciona el ángulo de rotación y el valor de translación. (Dcha.) Imagen de muestra con los intervalos discretos en los que se realiza el registro elástico.

Una vez finalizado el registro, las histologías alineadas se consolidaron en una matriz tridimensional. Asimismo, las imágenes con las etiquetas de los tejidos obtenidas mediante deep learning se consolidaron en una matriz de iguales dimensiones. Por último, las coordenadas de las células inmunes previamente detectadas se registraron utilizando los mapas de desplazamiento del registro y se consolidaron en una nueva matriz tridimensional.

3.2. Análisis del microambiente inmune en 3D

Además de reconstruir en 3D el microambiente inmune de tumores pancreáticos, uno de los objetivos clave de este proyecto era analizar en profundidad toda la información biológica contenida en los volúmenes.

En primer lugar, se estudió la correlación entre las matrices (3D) de coordenadas de células inmunes y matrices que se crearon a partir del volumen de etiquetas, conteniendo cada una uno de los siete tejidos pancreáticos.

A continuación, se realizaron estudios para determinar la distribución de células alrededor de las lesiones precursoras al cáncer (PanIN). Dichas lesiones se desarrollan a partir de los distintos tubos pancreáticos que transportan las enzimas digestivas. Así, las distintas lesiones PanIN se identificaron como objetos espacialmente independientes y seguidamente, para cada lesión, se calculó la densidad de células inmunes en función de la distancia (Euclídea) radial de su superficie.

Este último estudio permite determinar la distribución de células alrededor de PanINs. Sin embargo, al considerar las lesiones como un único objeto se pierde parte de la información 3D. Para analizar la inflamación de forma más exhaustiva se crearon mapas tridimensionales de células inmunes. De este modo, para cada lesión PanIN identificada, se calculó el número de células en el microambiente del tumor para cada punto de su superficie.

4. Resultados

4.1. Evaluación de la metodología.

Para evaluar la calidad del algoritmo de detección de células inmunes, las células detectadas automáticamente se compararon con células detectadas manualmente. Se obtuvo que el algoritmo presenta una precisión del $86 \pm 0.06 \%$, una exhaustividad del $88 \pm 0.07 \%$ y un valor-F del $87 \pm 0.03 \%$. Por su parte, el modelo de deep learning entrenado alcanzó una precisión y exhaustividad media del 95.2% y 95.3% respectivamente en la clasificación de

tejidos pancreáticos en imágenes H&E. Finalmente, para la evaluación del registro, el coeficiente Dice se midió entre imágenes contiguas. Se obtuvo un valor medio de 0.98 ± 0.01 , indicando que la superposición media entre imágenes contiguas es del 98%.

La Figura 5 ilustra los resultados de los distintos pasos seguidos para la reconstrucción 3D de la muestra de tejido.

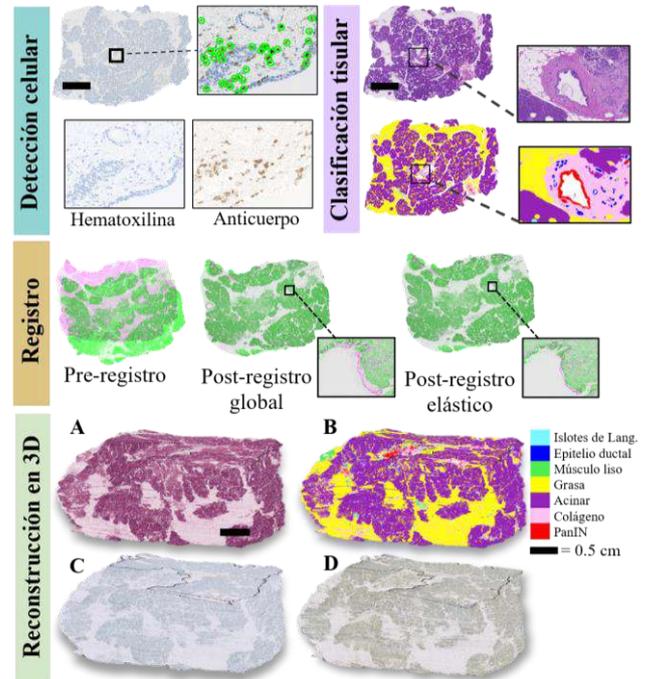


Figura 5. (Sup., Izqda.) Detección automática de células a partir de los canales resultantes de la deconvolución de color (Sup., Dcha.) Resultado de la clasificación de los tejidos pancreáticos en secciones H&E mediante deep learning. (Ctro.) Imágenes de muestra superpuestas antes del registro, después del registro global y el resultado final después del registro elástico. (Inf.) Reconstrucción digital en 3D de 1145 histologías seriadas de páncreas. A. Volumen H&E. B. Volumen con tejidos identificados. C. Volumen CD45. D. Volumen CD3/Foxp3.

4.2. Análisis del microambiente inmune en 3D

Inflamación en los tejidos pancreáticos

Distintos tejidos pancreáticos presentan diferentes niveles de inflamación. Como se muestra en la Tabla 1, los tres tipos de células inmunes tienen una correlación positiva con las lesiones precursoras al cáncer (PanIN) y con colágeno, el componente mayoritario del estroma de estas lesiones precancerosas. Por el contrario, tejidos como la grasa o los acinos secretores de enzimas digestivas están negativamente correlacionados con la inflamación.

Leucocitos	0.11	0.03	0.00	-0.17	-0.40	0.46	0.15
Cél. T	0.15	0.04	-0.04	-0.21	-0.22	0.52	0.18
Cél. T _{reg}	0.12	0.03	0.01	-0.11	-0.28	0.46	0.12
	Islotes de L.	Epitelio D.	Músculo liso	Grasa	Acinos	Colágeno	PanIN

Tabla 1. Correlación entre las células inmunes y los distintos tejidos pancreáticos.

Perfil de la inflamación alrededor de los tumores

En la muestra de tejido analizada, se identificaron 28 lesiones precursoras al cáncer (lesiones PanIN). La Figura 6A ilustra la lesión de mayor tamaño (16.7 mm³) con las células inmunes del entorno tumoral.

Como se puede observar en la Figura 6B, el nivel de inflamación alcanza el valor máximo en la superficie de la lesión precursora al cáncer. Seguidamente disminuye rápidamente en los primeros 100-200 μm para finalmente estabilizarse en un valor de inflamación constante.

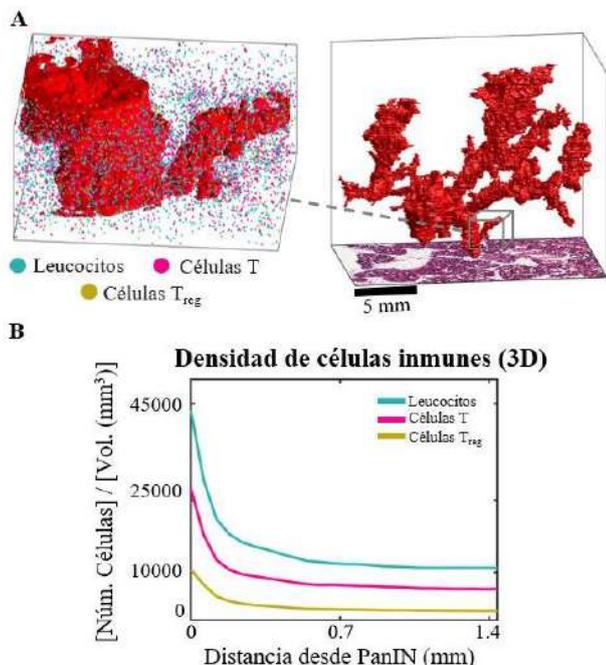


Figura 6. Análisis del perfil de inflamación alrededor de lesiones precancerosas (PanIN). **A.** Representación en 3D de una lesión específica (PanIN I) y células inmunes circundantes. **B.** Densidad de células en función de la distancia a la lesión.

Mapas tridimensionales de la inflamación

La Figura 7 ilustra los mapas 3D de células inmunes de la lesión precursora al cáncer de mayor tamaño.

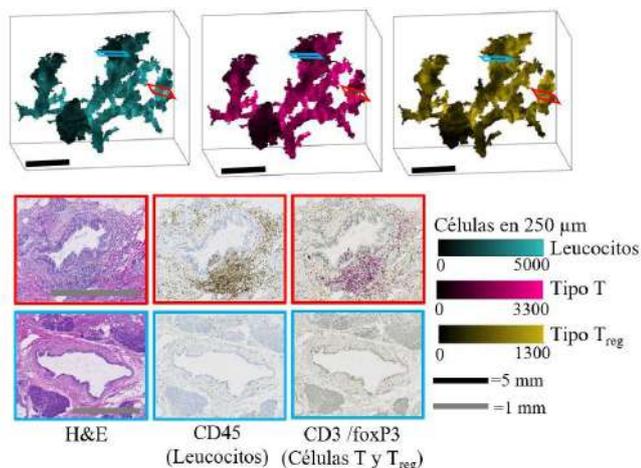


Figura 7. Mapas tridimensionales de la inflamación alrededor de lesiones precancerosas. Visualización de una lesión específica (PanIN I). Esta presenta áreas de baja inflamación (marco azul) y alta (marco rojo) con una diferencia de tres órdenes de magnitud en término de número de células.

Como se puede observar, en un mismo PanIN la inflamación varía hasta en tres órdenes de magnitud entre unas regiones y otras, indicando un grado de heterogeneidad imposible de captar en estudios en 2D.

5. Conclusiones

Este proyecto muestra que el microambiente inmune puede ser reconstruido en tres dimensiones y a resolución unicelular a partir de miles de secciones histológicas. Como apuntan los análisis realizados, es una herramienta capaz de proporcionar información valiosa sobre el entorno del tumor y su nivel de inflamación que podría ser útil para el desarrollo de nuevas inmunoterapias que contribuyan a la lucha contra el cáncer de páncreas.

Agradecimientos

Este proyecto ha sido financiado por la fundación Rolfe y el centro de investigación del Cáncer Pancreático Sol Goldman, perteneciente a la escuela de medicina de la Universidad Johns Hopkins.

Referencias

- [1] L. Zheng, J. Xue, E. M. Jaffee, and A. Habtezion, "Role of immune cells and immune-based therapies in pancreatitis and pancreatic ductal adenocarcinoma," *Gastroenterology*, vol. 144, no. 6, pp. 1230–1240, 2013.
- [2] E. Dell'Aquila *et al.*, "Prognostic and predictive factors in pancreatic cancer," *Oncotarget*, vol. 11, no. 10, 2020.
- [3] R. F. De Wilde, R. H. Hruban, A. Maitra, and G. J. A. Offerhaus, "Reporting precursors to invasive pancreatic cancer: Pancreatic intraepithelial neoplasia, intraductal neoplasms and mucinous cystic neoplasm," *Diagnostic Histopathology*, vol. 18, no. 1, pp. 17–30, 2012.
- [4] U. M. Mahajan *et al.*, "Immune cell and stromal signature associated with progression-free survival of patients with resected pancreatic ductal adenocarcinoma," *Gastroenterology*, vol. 155, no. 5, pp. 1625–1639, 2018.
- [5] J. T. Liu *et al.*, "Harnessing non-destructive 3D pathology," *Nature Biomedical Engineering*, vol. 5, no. 3, pp. 203–218 2021.
- [6] A. C. Ruifrok, D. A. Johnston, *et al.*, "Quantification of histochemical staining by color deconvolution," *Analytical and quantitative cytology and histology*, vol. 23, no. 4, pp. 291–299, 2001.
- [7] Pei-H. Wu, C. M. Hale, W.-C. Chen, J. S. Lee, Y. Tseng, and D. Wirtz, "High-throughput ballistic injection nanorheology to measure cell mechanics," *Nature Protocols*, vol. 7, no. 1, pp. 155–170, 2012.
- [8] Jin X., Han J. "K-Medoids Clustering" *Encyclopedia of Machine Learning*. Springer, 2011.
- [9] A. Kiemen *et al.*, "In situ characterization of the 3D microanatomy of the pancreas and pancreatic cancer at single cell resolution," bioRxiv, 2020.

Estudio de la arquitectura temporal de la actividad neuronal en la demencia por enfermedad de Alzheimer mediante técnicas de conectividad funcional dinámica

M. Bautista-Salineró¹, P. Núñez², R. Goya Esteban¹, M.A. Tola-Arribas³, M. Cano del Pozo⁴, J. Poza^{2,5,6}

¹ Universidad Rey Juan Carlos, Madrid, España, m.bautista.2017@alumnos.urjc.es

² Grupo de Ingeniería Biomédica, Universidad de Valladolid, Valladolid, España

³ Departamento de Neurología, Hospital Universitario Río Hortega, Valladolid, España

⁴ Departamento de Neurofisiología Clínica, Hospital Universitario Río Hortega, Valladolid, España

⁵ Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), España

⁶ IMUVA, Instituto de Investigación en Matemáticas de la Universidad de Valladolid, Valladolid, España

La enfermedad de Alzheimer (EA) es una patología asociada a una desconexión cerebral, que típicamente ha sido caracterizada en base al análisis de conectividad entre regiones corticales. Sin embargo, recientes estudios apuntan a que es necesario tener en cuenta la variabilidad temporal de la red cerebral para caracterizar de forma precisa las alteraciones funcionales y cognitivas asociadas. Por ello, en el presente estudio se analizan las alteraciones en la arquitectura temporal subyacente a los patrones de conectividad funcional dinámica asociadas a la progresión de la EA, mediante registros de electroencefalografía (EEG) en reposo. Se empleó una base de datos compuesta por 160 pacientes pertenecientes a tres grupos: 43 controles cognitivamente sanos, 67 pacientes con deterioro cognitivo leve por EA y 50 enfermos con demencia debida a EA. Tras la obtención de patrones cerebrales recurrentes, denominados meta-estados, mediante un método no supervisado previamente desarrollado, se aplicaron nuevas medidas para cuantificar el nivel de desactivación de estos. Asimismo, se estudió la distribución de combinaciones de los 3 meta-estados con mayor probabilidad de ocurrencia. Los resultados indicaron que los meta-estados están menos definidos conforme progresa la EA. Además, se observó que el avance de la enfermedad implica una desestructuración de la secuenciación temporal de los patrones de activación simbólicos de los meta-estados. Se puede concluir, por tanto, que la EA provoca alteraciones significativas de la estructura temporal de los patrones de conectividad que pueden estar ligadas a los déficits funcionales de la enfermedad.

1. Introducción

El cerebro humano es un órgano sumamente complejo, en el que hay una intrincada red de conexiones neuronales [1]. Esta red es susceptible a estímulos externos y presenta alteraciones dinámicas, tanto durante la realización de una función cognitiva como durante el reposo, asociadas a diversas enfermedades [2]. El electroencefalograma (EEG) es una técnica de neurofisiología que permite la medición de la actividad eléctrica neuronal [1]. Debido a su bajo coste, su amplia difusión clínica, su no-invasividad y su alta resolución temporal, se ha convertido en una excelente herramienta para evaluar las alteraciones cerebrales de neuropatologías [3].

Una de las patologías en las que se han realizado numerosos estudios de EEG con actividad neuronal en

reposo es la enfermedad de Alzheimer (EA). Esta enfermedad es la causa de demencia más común en países occidentales y representa del 50 al 60% de todos los casos [4]. Es una patología neurodegenerativa caracterizada por la pérdida progresiva de memoria, seguida por un deterioro del juicio, de la capacidad de toma de decisiones y de la orientación [4]. Su fenotipo clínico presenta un estado prodrómico, denominado deterioro cognitivo leve (DCL), que se caracteriza por un deterioro cognitivo superior al esperado demográficamente por sexo, edad y nivel de educación [2]. Actualmente, los pacientes con DCL predecesor de EA son el principal objetivo de numerosos estudios para encontrar marcadores tempranos [3].

Los estudios de EEG en estado de reposo permiten el análisis de patrones de fluctuaciones del cerebro en ausencia de actividad específica. Estos patrones se pueden medir mediante el acoplamiento de la activación funcional entre regiones del cerebro, permitiendo la cuantificación de la conectividad cerebral [5]. Habitualmente, los estudios de acoplamiento neuronal en estado de reposo se han centrado en analizar la conectividad funcional estática (sFC), que estudia los patrones de acoplamiento promediados a lo largo del tiempo [3]. Sin embargo, cada vez hay más evidencias que apuntan a que el estudio de la variabilidad del acoplamiento neuronal, mediante técnicas de conectividad funcional dinámica (dFC), es fundamental para entender el procesamiento y transmisión de la información neuronal [6], [7].

En un estudio previo [8] se introdujo un nuevo método de detección de patrones cerebrales recurrentes (meta-estados) basado en el cálculo de la conectividad funcional instantánea, *recurrence plots* y algoritmos de detección de comunidades en controles, pacientes con DCL y EA. Los resultados de este estudio confirmaron que tanto el DCL como la EA afectan a la secuenciación de los meta-estados, posiblemente debido a una menor flexibilidad, especialmente en la banda alfa. Considerando estos resultados, en el presente estudio se pretende dar un paso más allá. En este sentido, la hipótesis del trabajo es que la progresión de la EA provoca una desestructuración

paulatina de la arquitectura temporal subyacente a la secuenciación de los meta-estados funcionales. Con el objetivo de cuantificar estas alteraciones, en este estudio se plantea caracterizar la distribución de los meta-estados como patrones simbólicos mediante la introducción de nuevas medidas derivadas de los patrones de dFC.

2. Materiales

2.1. Participantes

La base de datos utilizada en el presente estudio está compuesta por 160 pacientes pertenecientes a tres grupos: 43 controles cognitivamente sanos, 67 pacientes con DCL por EA, y 50 enfermos con demencia debida a EA. En la Tabla 1 aparecen los datos clínicos y sociodemográficos más relevantes. Los pacientes fueron diagnosticados mediante los criterios NIA-AA (*National Institute on Aging and Alzheimer's Association*). Todos los participantes y familiares fueron informados sobre el protocolo del estudio y dieron su consentimiento escrito. El estudio fue aprobado por el comité ético del Hospital Río Hortega de Valladolid, siguiendo el Código ético de la Asociación Médica Mundial (Declaración de Helsinki).

Se realizaron análisis estadísticos para evaluar diferencias entre los grupos para las características sociodemográficas. No se hallaron diferencias significativas en cuanto a la edad (p -valor > 0.05 , test de Kruskal Wallis) ni sexo (p -valor > 0.05 , test Chi cuadrado). En cambio, los grupos sí presentaron diferencias significativas en el nivel de educación (p -valor < 0.05 , test Chi cuadrado), y en la puntuación del *Mini-Mental State Examination* (MMSE) (p -valor < 0.05 , test de Kruskal Wallis).

2.2. Registros EEG y preprocesado

Las señales EEG fueron registradas en el Departamento de Neurofisiología Clínica del Hospital Universitario Río Hortega, en Valladolid. El registro se realizó mediante un electroencefalógrafo de 19 canales, siguiendo las especificaciones del sistema internacional 10-20, y con una frecuencia de muestreo de 200 Hz. Se registraron un total de 5 minutos por paciente en estado de reposo.

Para evitar artefactos durante el estudio, las señales se sometieron a un preprocesado en tres pasos [8]: (i) análisis de componentes independientes para eliminar movimientos musculares; (ii) filtrado de respuesta finita (FIR) con ventana de Hamming para limitar los

componentes espectrales entre [1 70] Hz y eliminar la interferencia de red de 50 Hz; (iii) selección de los primeros 60 s sin artefactos mediante inspección visual.

Haciendo uso de sLORETA se reconstruyeron las series temporales de EEG a nivel de fuente, que posteriormente fueron distribuidas en las 68 regiones corticales (ROI) del atlas de Desikan-Killiany [9].

3. Métodos

3.1. Detección de los meta-estados cerebrales

En esta sección se explica brevemente la metodología introducida en [8], cuyo objetivo es identificar meta-estados cerebrales y generar las series temporales de la activación de los mismos.

- (i) Utilizando la *Instantaneous Amplitude Correlation* (IAC) [10] se construyó un vector de conectividad instantánea para cada una de las muestras del EEG ($68 \text{ ROI} \times 68 \text{ ROI} \times 12,000$ muestras). La IAC se calculó en la banda alfa (8-13 Hz), dada su importancia para caracterizar las alteraciones en la dFC conforme progresa la EA [8].
- (ii) A partir de las matrices IAC se construyeron *recurrence plots*.
- (iii) La extracción de meta-estados se llevó a cabo mediante el método de detección de comunidades Louvain-GJA aplicado sobre los *recurrence plots*.
- (iv) Se extrajeron la representación discreta del meta-estado dominante en cada instante (*Temporal Activation Sequence*, TAS) y el tensor de la correlación de Spearman de cada meta-estado con la IAC en cada instante (*Instantaneous Correlation Tensor*, ICT). La Figura 1 muestra los primeros 5 s del TAS y del ICT para un paciente con DCL.

3.2 Nuevas métricas para caracterizar la arquitectura temporal de la dinámica de los meta-estados

El objetivo fundamental de este artículo es caracterizar la arquitectura temporal dinámica de los meta-estados. Para ello se plantearon dos estrategias: i) cuantificar el balance entre las fuerzas de atracción y repulsivas de cada meta-estado, analizando la relación entre las correlaciones positivas y negativas del ICT; ii) analizar la distribución de la secuenciación temporal de los meta-estados que aparecen en el TAS.

Grupo	EA	DCL	C
Sujetos	50	67	43
Edad (años)	79 [76 , 82]	77 [72 , 81]	76 [74 , 79]
Sexo (H:M)	23:27	29 :38	13: 30
Educación (A:B)	35:15	41:26	16:27
MMSE	22 [20 , 24]	27 [26 , 28]	29 [28 ,30]

Tabla 1. Características clínicas y sociodemográficas de la base de datos. A: educación primaria o inferior; B: educación secundaria o superior. [Q1 Q3]: rango intercuartílico.

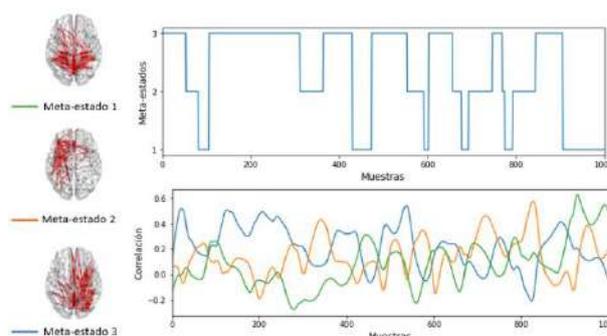


Figura 1. Representación del TAS (superior) y del ICT (inferior) de un paciente con DCL con 3 meta-estados.

Factor de antagonismo (FA)

Esta métrica permite calcular la cantidad de correlaciones negativas frente al total de las correlaciones por sujeto.

$$FA = \frac{1}{N} \sum_{n=1}^N \frac{|\sum CN^{ICT}(n)|}{\sum |CT^{ICT}(n)|}$$

donde $CN^{ICT}(n)$ representa las correlaciones negativas del ICT en la muestra temporal n , $CT^{ICT}(n)$ representa todas las correlaciones del ICT y N es el número de muestras.

Grado de antagonismo (GA)

El grado de antagonismo refleja la diferencia de fuerza ejercida entre la correlación más atractiva del ICT (CA), reflejada en la correlación positiva de mayor valor, y la más repulsiva (CR), identificada como la correlación negativa de mayor valor absoluto. En caso de ausencia de correlación negativa, se establece dicho valor como 0.

$$GA = \frac{1}{N} \sum_{n=1}^N [CA(n) - CR(n)]$$

Patrones simbólicos

El objetivo de esta métrica es estudiar la distribución de probabilidad de ocurrencia de las posibles combinaciones de los meta-estados. Para homogeneizar la medida entre los individuos de cada grupo, se consideraron únicamente los 3 meta-estados con mayor probabilidad de ocurrencia de cada sujeto. Se crearon 18 posibles combinaciones, ordenando el símbolo de cada meta-estado de mayor a menor valor del ICT. En caso de presentar valores del ICT negativos, el símbolo del meta-estado se estableció con un signo negativo. Por ejemplo, tras ordenar de mayor a menor las correlaciones del ICT en un paciente, se forma la combinación [2 1 3]. Si el meta-estado 3 presenta un valor negativo, la combinación final es [2 1 -3]. Para estudiar la dispersión de las distribuciones categóricas de meta-estados se utilizó el índice de variación cualitativa (IVQ) y la entropía de Shannon.

3.3 Análisis estadísticos

Para garantizar que todos los resultados obtenidos se deben a dFC real del cerebro y no a fluctuaciones aleatorias, se normalizaron las medidas derivadas del TAS y el ICT. Para ello, se generaron datos subrogados mediante AAFT (*Amplitude Adjusted Fourier Transform*) usando la misma aleatorización para cada ROI [8]. Todas las métricas se normalizaron dividiendo entre el valor medio de las 100 secuencias subrogadas. Se realizaron test no paramétricos para evaluar las diferencias entre grupos. Se usó el test de Kruskal Wallis para detectar interacciones globales entre los tres grupos; para investigar las diferencias entre pares de grupos, se empleó el test U de Mann-Whitney. Finalmente, se aplicó la corrección FDR (*False Discovery Rate*) para corregir por el número de bandas y grupos.

4. Resultados y discusión

4.1. Desequilibrio entre fuerzas de atracción y repulsivas

En la Figura 2 se observa la representación del GA y FA en la banda alfa. Se hallaron diferencias estadísticamente

significativas entre grupos en la banda alfa, con ambas métricas (grado: p -valor = 0.0092; factor: p -valor < 0.001; test de Kruskal Wallis, corrección FDR). Entre pares de grupos se encontraron diferencias entre controles y pacientes con DCL, así como entre controles y enfermos con EA (grado: p -valor = 0.0008; factor p -valor < 0.001; test U de Mann-Whitney, corrección FDR).

Los resultados del GA y del FA muestran valores superiores en los controles, frente a los otros dos grupos. Teniendo en cuenta que las correlaciones negativas están asociadas con una fuerza de repulsión respecto al meta-estado considerado, los meta-estados que permanecen activos están menos influenciados y por tanto atraen con mayor fuerza a la matriz de conectividad instantánea. De esta manera, los estados dominantes presentan una mayor separación en controles sanos, mientras que la reducción de las métricas de antagonismo en DCL y EA sugieren una separación más difusa de los meta-estados. Estos resultados también refuerzan los encontrados en el estudio previo [8], donde se observó que los controles presentaban una mayor modularidad, la cual es un indicador del grado de definición y separación de los meta-estados. Asimismo, los resultados normalizados apoyan la idea de que las diferencias encontradas son debidas a dFC genuina, y no a fluctuaciones aleatorias, siendo prueba de que las diferencias observadas entre grupos se deben a un deterioro de las propiedades de reconfiguración dinámica de los patrones conectividad conforme avanza la EA.

4.2. Alteraciones en la arquitectura temporal de la actividad EEG espontánea

En la Figura 3 se pueden observar los resultados de la probabilidad de ocurrencia de los patrones simbólicos. Se aprecian claras diferencias entre las combinaciones con mayor y menor presencia. Se encontraron diferencias estadísticamente significativas (p -valor = 0.0094, test de Kruskal Wallis, corrección FDR) con el IVQ. Entre pares de grupos, se hallaron diferencias significativas entre controles y pacientes con DCL, así como entre controles y pacientes con EA (control-DCL: p -valor = 0.0098; control-EA: p -valor < 0.001, test U de Mann-Whitney, corrección FDR). Para la entropía de Shannon no se encontraron diferencias significativas globales (p -valor > 0.05, test de Kruskal Wallis, corrección FDR), si bien, al comparar entre grupos se obtuvieron diferencias significativas entre controles y pacientes con EA (p -valor = 0.009, test U de Mann-Whitney, corrección FDR).

Al analizar la estructura temporal del EEG, es interesante observar que los patrones con mayor probabilidad son

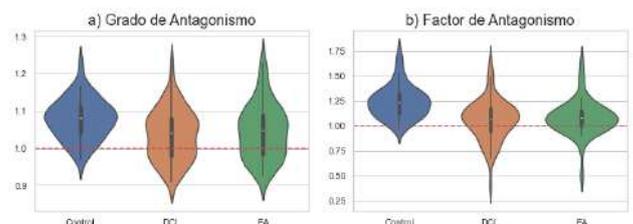


Figura 2. a) Distribución del grado y b) factor de antagonismo para los tres grupos bajo estudio (controles, pacientes con DCL y enfermos con EA) en la banda alfa.

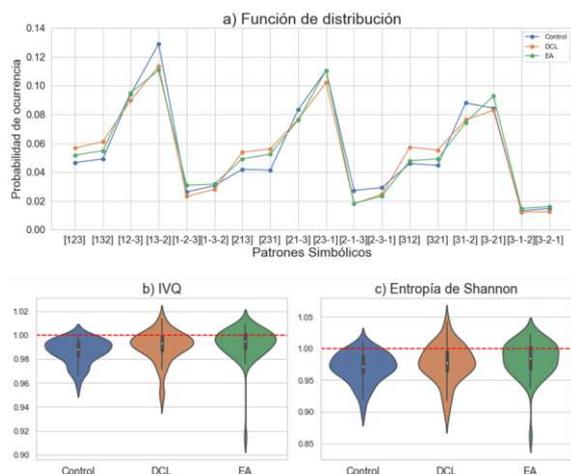


Figura 3. a) Función de distribución de los patrones simbólicos para los tres grupos bajo estudio (controles, pacientes con DCL y enfermos con EA) en la banda alfa. Distribuciones normalizadas de b) IVQ c) entropía de Shannon.

aquellos en los cuales existe un único meta-estado negativo desactivado, mientras que los otros permanecen activos. El patrón común en las combinaciones menos probables es la presencia de dos estados desactivados. Asimismo, los patrones en los que todos los meta-estados están activados tienen una baja probabilidad. Esto refuerza los resultados previos [8], donde se observó que los meta-estados no se activan de manera exclusiva y los patrones de conectividad instantánea se generan en base a la combinación dinámica de los meta-estados principales con diferente fuerza. Sin embargo, son necesarios nuevos estudios para confirmar esta idea. Los resultados del IVQ y de la entropía de Shannon muestran valores inferiores en los controles, lo cual indica que según la enfermedad progresa, la distribución de los patrones simbólicos es más uniforme. Esto estaría asociado con una pérdida paulatina de la estructura temporal subyacente, que refuerza los resultados presentados en [8].

4.3. Limitaciones y líneas futuras de investigación

Este estudio presenta ciertas limitaciones que pueden ser abordadas en futuras investigaciones. Para empezar, la base de datos utilizada consta únicamente de 160 sujetos, lo que limita la capacidad de generalización de los resultados. Para continuar, el grupo de pacientes con DCL se podía subdividir en dos, aquellos que evolucionan a EA y los que permanecen esa etapa. Con esta información, se podría realizar un estudio más detallado sobre la progresión de la enfermedad. Para finalizar, los meta-estados utilizados están numerados por frecuencia de aparición. Sería muy interesante asignar los meta-estados individuales en base a meta-estados grupales, comunes para todos los sujetos.

5. Conclusiones

Los resultados de este estudio indican que tanto el DCL como la EA provocan una asignación más difusa de los meta-estados de dFC que, una vez agregados, generan los patrones de sFC; esto se puede asociar con una peor definición de la secuenciación temporal de los patrones de dFC. Asimismo, se ha observado que la progresión de la EA da lugar a alteraciones en la estructura temporal de los patrones simbólicos de asignación a meta-estados. La

mayor ocurrencia de patrones con un meta-estado negativo sugiere que la conectividad instantánea no es exclusiva de un único meta-estado, sino que se construye mediante a la combinación de varios meta-estados "base". Por tanto, los resultados obtenidos en este artículo suponen un nuevo avance para entender los mecanismos neuronales implicados en la progresión del EA.

Agradecimientos

Este estudio ha sido financiado por el 'Ministerio de Ciencia e Innovación - Agencia Estatal de Investigación' y el Fondo Europeo de Desarrollo Regional (FEDER) bajo el proyecto PGC2018-098214-A-I00 y por el CIBER en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN) a través del Instituto de Salud Carlos III cofinanciado con fondos FEDER. P. Núñez disfruta de un contrato predoctoral 'Ayuda para contratos predoctorales para la Formación de Profesorado Universitario (FPU)' del 'Ministerio de Educación, Cultura y Deporte' (FPU17/00850).

Referencias

- [1] C. Maestú Unturbe, E. Gómez-Utrero Fuentes, R. Piñeiro Bolaño, and R. García de Sola, "Magnetoencefalografía: una nueva técnica de diagnóstico funcional en neurociencias," *Rev. Neurol.*, vol. 28, no. 11, p. 1077, 1999.
- [2] Z. Jafari, B. E. Kolb, and M. H. Mohajerani, "Neural oscillations and brain stimulation in Alzheimer's disease," *Prog. Neurobiol.*, vol. 194, no. June 2019, p. 101878, 2020.
- [3] P. M. Rossini *et al.*, "Early diagnosis of Alzheimer's disease: the role of biomarkers including advanced EEG signal analysis. Report from the IFCN-sponsored panel of experts," *Clin. Neurophysiol.*, vol. 131, no. 6, pp. 1287–1310, 2020.
- [4] K. Blennow, M. J. de Leon, and H. Zetterberg, "Alzheimer's disease," *Lancet*, vol. 368, no. 9533, pp. 387–403, 2006.
- [5] F. Miraglia, F. Vecchio, and P. M. Rossini, "Searching for signs of aging and dementia in EEG through network analysis," *Behav. Brain Res.*, vol. 317, pp. 292–300, 2017.
- [6] G. Deco, V. K. Jirsa, and A. R. McIntosh, "Resting brains never rest: Computational insights into potential cognitive architectures," *Trends Neurosci.*, vol. 36, no. 5, pp. 268–274, 2013.
- [7] L. Waschke, N. A. Kloosterman, J. Obleser, and D. D. Garrett, "Behavior needs neural variability," *Neuron*, vol. 109, no. 5, pp. 751–766, 2021.
- [8] P. Núñez *et al.*, "Abnormal meta-state activation of dynamic brain networks across the Alzheimer spectrum," *Neuroimage*, vol. 232, no. October 2020, p. 117898, 2021.
- [9] R. S. Desikan *et al.*, "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," *Neuroimage*, vol. 31, no. 3, pp. 968–980, 2006.
- [10] P. Tewarie *et al.*, "Tracking dynamic brain networks using high temporal resolution MEG measures of functional connectivity," *Neuroimage*, vol. 200, no. May, pp. 38–50, 2019.

Optimización de hidrogeles inyectables y liberación de factores de crecimiento hepático

S. Mares-Bou¹, J. Rodríguez-Fernández¹, S. Trujillo-Muñoz¹, G. Gallego-Ferrer^{1,2}

¹ Centre for Biomaterials and Tissue Engineering (CBIT), Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, España

² Centro de Investigación Biomédica en Red en Biomateriales, Bioingeniería y Nanomedicina (CIBER-BBN), España

Resumen

El daño hepático inducido por fármacos (DILI) es un evento clínico con consecuencias potencialmente mortales que supone un coste sanitario elevado, lo que genera la necesidad del desarrollo de plataformas de cultivo celular capaces de predecirlo. Los hidrogeles inyectables biomiméticos que tratan de imitar la matriz extracelular (ECM) del hígado, como mezclas de gelatina (Gel) y ácido hialurónico (HA), son candidatos ideales como modelos *in vitro* para predecir eventos de hepatotoxicidad. El objetivo es encontrar una estrategia de obtención de hidrogeles Gel-HA en medio de cultivo y con propiedades mecánicas adecuadas para mantener la integridad en cultivo, y una síntesis de HA con mayor rendimiento en cada lote de fabricación. Además, se ha puesto a punto un método de cuantificación de la liberación de factor de crecimiento hepático (HGF) a bajas dosis y desde los hidrogeles inyectables, a partir de un marcaje fluorescente.

1. Introducción

El hígado es un órgano vital en el que se realiza el metabolismo y detoxificación de medicamentos y, por tanto, susceptible de sufrir daño hepático inducido por fármacos (DILI). El DILI puede derivar en insuficiencia hepática aguda e incluso crónica. Los modelos actuales de desarrollo de fármacos, cultivos *in vitro* y animales de experimentación, predicen pobremente la hepatotoxicidad, siendo el DILI un problema importante de salud y la principal causa de retirada de fármacos del mercado. En consecuencia, el desafío actual para la industria farmacéutica y agencias reguladoras está en el desarrollo de modelos celulares altamente predictivos y reproducibles basados en ingeniería tisular que además permitan reducir la experimentación animal.

Los biomateriales de ingeniería tisular, a diferencia de las terapias celulares en 2D y sin matriz, reproducen las condiciones fisiológicas más relevantes para las células, la organización espacial 3D, la interacción célula-célula y la interacción célula-matriz. Los hidrogeles inyectables basados en componentes de la matriz extracelular (ECM) del tejido son candidatos ideales para este propósito, puesto que permiten la encapsulación de células homogéneamente en el material y por reacciones no citotóxicas, manteniéndolas en cultivos 3D que mimetizan el tejido nativo. La mayoría de los hidrogeles diseñados para simular el microambiente hepático se basan en una combinación proteína-polisacárido, similar al del presente estudio (Fig. 1), compuesto por gelatina (Gel), para mimetizar el colágeno de la ECM del hígado, y ácido hialurónico (HA), para conferir hidratación y propiedades

mecánicas similares a las del hígado [1]. El entrecruzamiento de ambos polímeros se basa en una reacción entre moléculas de tiramina (Tyr) previamente injertadas de manera no citotóxica [1].

Una función importante de la ECM es la presentación de factores de crecimiento (GFs), capaces de controlar procesos de proliferación y diferenciación celular y morfogénesis del tejido [2]. Algunos de estos sitios de unión a GFs se conservan en diferentes proteínas y polisacáridos propios de la ECM. Los hidrogeles inyectables permiten incorporar GFs en el momento de encapsulación celular sin dañarlos y ser retenidos y presentados desde fase sólida (Fig. 1), haciendo los cultivos más eficientes para el mantenimiento del fenotipo y la funcionalidad celular. Se ha demostrado que existe una interacción entre HA y el receptor de membrana CD44, hecho relevante para nuestro estudio, ya que es uno de los receptores que participa en la regulación de la señalización de GFs [3]. Además, es sabido que los polisacáridos, en especial los sulfatados, presentan afinidad por los GFs [2], lo que da pie a hipotetizar que hidrogeles combinados de Gel/HA pueden retener en alguna medida GFs hepáticos.

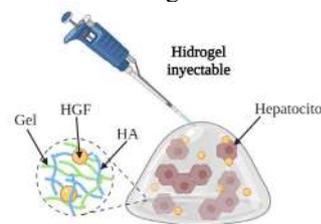


Figura 1. Ilustración del concepto de hidrogel inyectable del presente estudio.

La novedad de este trabajo consiste en regular la composición y rigidez del hidrogel a la vez que se regula la capacidad de presentar HGFs desde la matriz del hidrogel (Fig. 1), como factores determinantes para el mantenimiento del fenotipo y la funcionalidad hepática en modelos de predicción *in vitro* de hepatotoxicidad optimizados para prevenir el DILI.

2. Materiales y métodos

2.1. Síntesis de polímeros con injerto de tiramina: gelatina (Gel-Tyr) y ácido hialurónico (HA-Tyr)

La gelatina (Gel) (tipo A, Bloom 300, 80 mM[-COOH]/100 g) y el ácido hialurónico (HA) (sal de sodio), ambos de Sigma-Aldrich (Alemania), se modificaron por el método de injerto de tiramina (Tyr), siguiendo el trabajo previo de Poveda *et al.* [1]. Se incorporaron dos

modificaciones. Para tratar de aumentar la rigidez de los hidrogeles de Gel, se consideró aumentar el ratio Tyr:COOH y EDC:COOH a 5:1. Y para mejorar el rendimiento de fabricación de HA-Tyr se decidieron dos pruebas: duplicar la concentración del polímero en la mezcla reactiva inicial y duplicar el volumen de la disolución inicial en la reacción. Los reactivos y procedimientos fueron los mismos que en [1].

2.2. Grado de injerto de tiramina

El grado de sustitución de grupos tiramina (Tyr) en las cadenas poliméricas se midió mediante el espectrofotómetro Cary 60 UV-Vis (Agilent Technologies, Estados Unidos). Primero, se determinó la absorbancia de las disoluciones (0.1% p/v) de Gel-Tyr y HA-Tyr en agua mQ a la longitud de onda (λ) característica de Tyr (275 nm). Se extrapoló la concentración de Tyr mediante una recta de calibrado conformada por 5 disoluciones a concentraciones conocidas (de 0.125 mg/mL a 0.0125 mg/mL). El grado de injerto se determinó mediante la estimación de los grupos carboxilo sustituidos por Tyr (estimados con la recta de calibrado) y los grupos carboxilo iniciales de los polímeros sin injertar.

2.3. Gelificación de hidrogeles inyectables

Se fabricaron hidrogeles puros de Gel (4% p/v) y HA (2% p/v) y una mezcla de ambos Gel/HA al 50/50 en masa (4%/2% p/v). El proceso de entrecruzamiento se realizó mediante una reacción catalizada por peroxidasa de rábano (HRP) en presencia de peróxido de hidrógeno (H_2O_2) (Sigma-Aldrich, Alemania), empleando un 80% v/v de la disolución precursora del hidrogel (Gel-Tyr y/o HA-Tyr), un 10% v/v de HRP a 12.5 U/mL y un 10% v/v de H_2O_2 a 20 mM. Los medios de gelificación empleados fueron el medio de cultivo de la mezcla de nutrientes F-12 de Ham (F12, con L-glutamina y rojo fenol), específico de hepatocitos, y el medio esencial mínimo de Dulbecco (DMEM, bajo en glucosa y sin rojo fenol), ambos proveídos por Gibco (Estados Unidos).

2.4. Propiedades mecánicas

Las propiedades viscoelásticas de los hidrogeles puros de Gel y HA se caracterizaron con el reómetro Discovery HR-2 Hybrid (TA Instruments, Estados Unidos). Las medidas se obtuvieron a partir de muestras líquidas que entrecruzaron *in situ* entre los platos de acero del reómetro (\varnothing 20 mm). Se realizaron ensayos dinámico-mecánicos a cizalla sobre 3 réplicas de 3 muestras de Gel y 3 muestras de HA en F12. Las muestras de Gel fueron dos del conjugado Gel-Tyr habitual, una al 2% p/v y otra al 4% p/v, y la prueba de conjugado Gel-Tyr a mayor ratio molar al 2% p/v. Las muestras de HA fueron las de la síntesis habitual, la síntesis a doble concentración y a doble volumen, todas al 2% p/v. Se realizaron pruebas de barrido de amplitud de la deformación para calcular el módulo complejo (G^*) y así determinar la región viscoelástica lineal (LVR). Se realizaron las pruebas de barrido de tiempo durante la gelificación (45 min a 1 Hz y 1% de deformación). Luego, con el hidrogel completamente gelificado, se realizó un barrido de frecuencia (de 0.1 a 10 Hz) a la deformación constante dentro de la LVR (1%).

2.5. Contenido de agua de equilibrio

El contenido de agua de equilibrio de los hidrogeles se midió sobre 3 réplicas de 3 muestras de Gel/HA (4%/2% p/v) (100/0, 50/50 y 0/100) en DMEM. Los hidrogeles fueron fabricados en un molde de silicona y de 250 μ L. Tras gelificar, se tomó el peso inicial (m_i). Luego, se dejaron sumergidos en tampón fosfato salino sin calcio ni magnesio (PBS(-/-)) 1X durante 24 h a 37 °C y humedad relativa 100%. Transcurrido ese tiempo, se tomó el peso del hidrogel hinchado (m_h). Por último, se liofilizaron y se tomó el peso en seco (m_s). Con estos datos (Fig. 2) se calcularon las magnitudes *Water Content relaxed* (WCr) y *Equilibrium Water Content* (EWC) referidos al grado de hinchado de agua justo tras la gelificación y en equilibrio tras un día sumergidos en PBS, respectivamente.

$$WCr = \frac{m_i - m_s}{m_s} ; \quad EWC = \frac{m_h - m_s}{m_s}$$



Figura 2. Fotografías de los hidrogeles puros de HA, la barra de dimensiones corresponde a 4 mm.

2.6. Etiquetado fluorescente del HGF

Para estudiar la liberación del factor de crecimiento hepático (HGF) (PeproTech, Reino Unido) encapsulado en el hidrogel, éste se etiquetó con un marcador fluorescente amino reactivo (DyLight 488 NHS Ester, Thermo Fisher, Estados Unidos) siguiendo las instrucciones del fabricante y el protocolo de Trujillo *et al.* [4]. Previamente se hizo una prueba de concepto con albúmina de suero bovino (BSA, Sigma-Aldrich, Alemania). Se caracterizó el nivel de etiquetado fluorescente con un espectrofotómetro de gota (micro-UV-Vis NanoDrop One, Thermo Fisher, Estados Unidos) para asegurarse de la tinción.

2.7. Ensayo de liberación de HGF

Una vez obtenido el HGF con marcaje fluorescente (HGF-488), el ensayo de liberación consistió en fabricar 3 réplicas de 3 composiciones diferentes de Gel/HA (4%/2% p/v) (100/0, 50/50 y 0/100) con DMEM como solvente. El hidrogel fue de 50 μ L en total incorporando un 20% v/v de HGF-488, fijando así una concentración de 10 μ g/mL en cada hidrogel. Tras gelificar, se sumergieron en 350 μ L PBS y se tomó todo el volumen de sobrenadante de cada muestra al cabo de 4 h, 24 h y 48 h, conservándolos a -80 °C, y rellenando de nuevo con 350 μ L PBS para continuar con el proceso de liberación. El ensayo se produjo a 37 °C y con una humedad relativa al 100%.

La liberación de HGF se calculó midiendo el sobrenadante en el tiempo de ensayo previamente fijado en un lector de placas de fluorescencia (Victor X3, Perkin-Elmer, Estados Unidos). Luego, se extrapoló la concentración de HGF-488 en cada sobrenadante gracias a la recta de calibrado conformada por 7 disoluciones de HGF-488 de concentraciones conocidas (de 2.5 μ g/mL a 0.04 μ g/mL) y se calculó el volumen total de sobrenadante según cada muestra y a cada tiempo. Así pues, se estimó la liberación de HGF al medio como el cociente entre la cantidad liberada medida y la cantidad teórica inicial en cada

hidrogel (0.5 μg). En la puesta a punto del modelo de liberación se empleó BSA-488 encapsulado en hidrogeles en F12 y DMEM, resultados de los cuales se decidió hacer el ensayo de HGF en DMEM y se comprobó que el volumen de sobrenadante incorporado no repercutía en el hinchado de agua en equilibrio ni la dilución de partículas en la falta de muestra y sensibilidad del equipo.

2.8. Análisis estadístico

Se realizó una prueba de normalidad K-S y una prueba t en el grado de hinchado y el ensayo de reometría, con factor de corrección de varianza Welch. Para el grado de hinchado se aplicó un ANOVA sobre los resultados de las 3 combinaciones de hidrogel. Todas las pruebas estadísticas con un intervalo de confianza del 95%.

3. Resultados

3.1. Optimización del biomaterial

Grado de injerto de tiramina

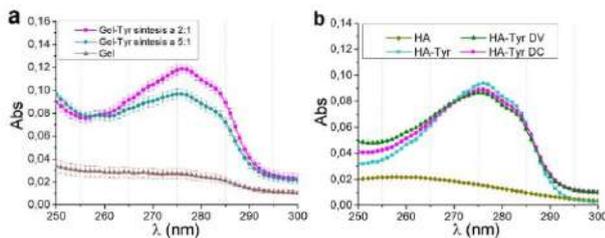


Figura 3. Espectros de las muestras de Gel (a) y HA (b).

En el caso de la Gel control se obtuvo un grado de injerto de control del 26.17% (obtenido a partir de la Fig. 3a). El objetivo de la prueba de optimización del conjugado Gel-Tyr a mayor ratio molar consistía en aumentar el grado de injerto de Tyr y así disponer de más puntos de entrecruzamiento entre las cadenas de Gel. Sin embargo, de esta muestra se tuvo un grado de injerto menor, en concreto de 21.52% (Fig. 3a). Por lo tanto, no se logró el objetivo propuesto, ya que no se consigue mayor grado de injerto ante un aumento de la proporción en ratio molar de los reactivos EDC y Tyr debido al límite de la capacidad de activación de grupos carboxilo en la reacción. En consecuencia, para conseguir hidrogeles de Gel más consistentes en cultivo, se decidió duplicar la concentración de Gel en la disolución precursora del hidrogel con respecto a la habitual para obtener mayores propiedades mecánicas (Gel 4% p/v).

En las muestras de HA se esperaba obtener un grado de injerto entorno al habitual, entre un 6-7%, ya que los ratios molares de los reactivos se mantuvieron y escalaron teniendo en cuenta la masa de HA. En consecuencia, el grado de injerto del HA resultó de 6.63% en el habitual y en las pruebas de 6.33% y 6.20%. Se deduce que las síntesis a DC y a DV han resultado con un grado de injerto equivalente al habitual (Fig. 3b).

Propiedades mecánicas

Los ensayos de reometría de los hidrogeles de Gel (Fig. 4a) revelaron un cambio no significativo entre la muestra habitual y la de prueba gelificadas a baja concentración en el hidrogel (2% p/v) y un cambio muy significativo al doble

de concentración (4% p/v). Tras el análisis del grado de injerto, se observó que no era posible aumentar éste incrementando los ratios molares en la síntesis de Gel-Tyr. Todo ello evidenciado de nuevo tras los ensayos de reología, donde no hubo diferencia significativa en las propiedades mecánicas, la optimización pasó por elevar la concentración de Gel en el hidrogel con tal de conseguir un efecto similar con mayor cantidad de entrecruzamientos. La concentración de la Gel pasó de 2% p/v a ser 4% p/v. En consecuencia, se obtuvo un hidrogel más reticulado y con mayores propiedades mecánicas debido al menor volumen ocupado de agua. Obteniéndose un módulo de almacenamiento (G') de 167 ± 64 Pa al 4% p/v, frente a 6 ± 2 Pa de la misma muestra al 2% p/v y 8 ± 1.5 Pa de la muestra de mayor ratio molar de EDC al 2% p/v.

En los ensayos reológicos de los hidrogeles de HA (Fig. 4b), la muestra habitual presentó un valor de G' ya elevado, de 775 ± 66 Pa. El objetivo del ensayo fue comprobar que los hidrogeles de HA con diferente síntesis eran equivalentes, como en el grado de injerto. En concreto, HA a doble volumen no mostró diferencias significativas con respecto al G' habitual, siendo este de 781 ± 90 Pa, mientras que a doble concentración presentó un incremento significativo, llegando hasta los 2127 ± 487 Pa (no fácilmente explicable). En definitiva, el método de fabricación de HA-Tyr a doble volumen realmente fabricó mayor cantidad de materia prima y con las mismas propiedades que el método tradicional.

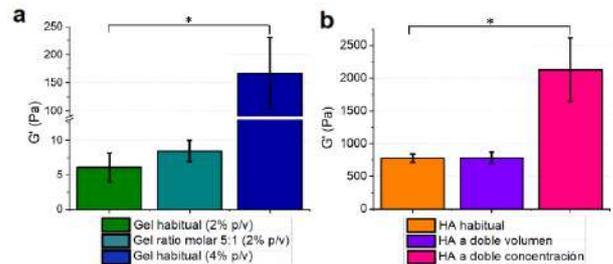


Figura 4. Gráficos con los valores de G' a 1 Hz de frecuencia y 1% de deformación fijas en las muestras de Gel (a) y HA (b) con $p < 0.05$ (*).

3.2. Ensayo de liberación de HGF desde hidrogeles

Grado de hinchado de agua de los hidrogeles

Por un lado, el WCr, sirvió para contrastar el contenido de agua justo tras gelificar cada muestra. En los hidrogeles puros de Gel fue de 36.70 ± 4 , en los de HA de 40.10 ± 1.13 y en la mezcla de 35.79 ± 0.32 , sin diferencias estadísticas entre muestras (Fig. 5). Por otro lado, el EWC de los hidrogeles está relacionado con la fracción volúmica de polímero en el hidrogel, la densidad de entrecruzamiento (en este caso relacionada con el grado de injerto de Tyr) y la naturaleza química del hidrogel en cuanto a su afinidad por el agua recogida en el parámetro de interacción de Flory-Huggins. El promedio de EWC en el caso de los hidrogeles puros de Gel fue de 27.8 ± 1.32 , en los hidrogeles puros de HA fue de 60.85 ± 4.19 , y el de la mezcla 50/50 presentó un valor intermedio de 46.05 ± 2.38 (Fig. 5). Como los hidrogeles de HA presentaron menor grado de injerto (6-7%) a menor concentración (2% p/v) que la gelatina, fueron capaces de retener una mayor

cantidad de agua en equilibrio. Estos resultados fueron comparables con el estudio de Poveda-Reyes *et al.* [1].

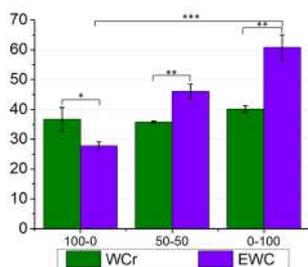


Figura 5. Resultados de WCr y EWC de los hidrogeles Gel-HA con $p < 0.05$ (*), $p < 0.01$ (**) y $p < 0.001$ (***).

Comparando ambos valores se observó que en el caso de los hidrogeles de Gel se libera parte del agua retenida tras gelificar, mientras que en los de HA y la mezcla ocurre lo opuesto, absorben más agua (véase la Fig. 5). Este hecho se justificó por la conformación de Gel y HA en equilibrio, que en el caso de la Gel debió estar menos extendida y en el caso del HA más extendida que cuando acabó de gelificar. Esta variación del tamaño del hidrogel y del volumen del sobrenadante tuvo un efecto directo en el ensayo de liberación de proteínas encapsuladas en estos hidrogeles y debió tenerse en cuenta a la hora de decidir el volumen de sobrenadante.

Liberación acumulativa de HGF desde los hidrogeles

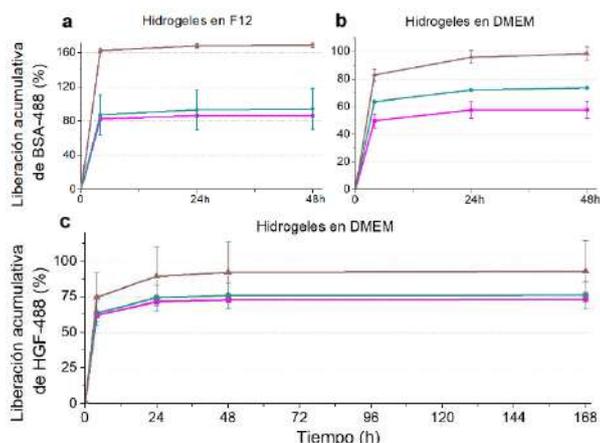


Figura 6. Gráficos con los perfiles de liberación de BSA-488 desde hidrogeles Gel/HA 100/0 (rosa), 50/50 (azul) y 0/100 (marrón) en F12 (a) y DMEM (b); y de HGF-488 desde hidrogeles Gel/HA en DMEM (c).

En la prueba con BSA-488 (Fig. 6a-b) resultó una liberación en los hidrogeles en F12 que se estabilizó en los de gelatina pura entorno al $86\% \pm 2\%$, los hidrogeles de HA al $169\% \pm 3\%$ y el mixto a $94\% \pm 24\%$. En los hidrogeles en DMEM, al $58\% \pm 6\%$ en hidrogeles de Gel pura, al $98\% \pm 5\%$ en hidrogeles de HA puro y al $74\% \pm 1\%$ en hidrogeles 50/50. En consecuencia, los ensayos de liberación en hidrogeles de F12 y DMEM mostraron una liberación de toda la proteína, liberándose la mayor cantidad tras 4 h y estabilizándose a 48 h. Se decidió realizar el ensayo con HGF en hidrogeles de DMEM, por el mejor perfil de liberación en la prueba de concepto.

En el ensayo definitivo con HGF (Fig. 6c) se observó un perfil similar al BSA y en el que se dedujo que se liberó

toda la cantidad incorporada. En este ensayo, de carácter excepcional, se calculó tras una semana para esclarecer la fase de estabilización de la liberación. En los hidrogeles de Gel se alcanzó un $73\% \pm 1\%$ de liberación de HGF, en los mixtos un $76\% \pm 9\%$ y en los de HA un $104\% \pm 13\%$.

Todos los resultados mostraron una mayor difusión en los hidrogeles con mayor contenido de HA. Esto se debió en parte a que estos hidrogeles fueron capaces de retener una mayor cantidad de agua en equilibrio, lo que favoreció la difusión de partículas del interior del hidrogel al sobrenadante. Pero en el ensayo se demostró que los hidrogeles no fueron capaces de retener HGF, puesto que se liberó toda la cantidad previamente encapsulada, es decir, que ni la Gel ni el HA demostraron tener afinidad por HGF. De cara a líneas futuras, este ensayo sirve de control negativo para analizar la liberación de HGF con el mismo método y con hidrogeles decorados con péptidos capaces de retener el HGF en la matriz de hidrogel.

En síntesis, la caracterización de los hidrogeles puros de Gel y HA pudo discutirse por la densidad de puntos de entrecruzamiento a raíz del grado de injerto de Tyr, ya que uno resultó más tupido y otro más extendido. Este hecho afectó directamente a la salida o entrada de agua en el equilibrio de cada tipo de hidrogel y, en consecuencia, a sus propiedades mecánicas y de difusión.

4. Conclusión

En este trabajo se ha conseguido mejorar las propiedades de los hidrogeles de gelatina (Gel) aumentando su concentración al 4% p/v y gracias a la síntesis al doble volumen se obtiene más materia prima de ácido hialurónico (HA) y con las mismas propiedades fisicoquímicas que el habitual. Como novedad, se ha puesto a punto un método de cuantificación de la liberación de HGF y se tomó el control negativo de HGF encapsulado en hidrogeles de Gel/HA para futuros ensayos.

Agradecimientos

Los autores agradecen la financiación a la Agencia Estatal de Investigación a través del proyecto “HEPATOPRINT” PID2019-106000RB-C21/AEI/10.13039/501100011033.

Referencias

- [1] Poveda-Reyes, S., Moulisova, V., Sanmartín-Masiá, E., Quintanilla-Sierra, L., Salmerón-Sánchez, M., y Gallego, G. (2016). Gelatin—Hyaluronic Acid Hydrogels with Tuned Stiffness to Counterbalance Cellular Forces and Promote Cell Differentiation. *Macromol Biosci*, 16, 1311–24.
- [2] Martino, M. M., y Hubbell, J. A. (2010). The 12th–14th type III repeats of fibronectin function as a highly promiscuous growth factor-binding domain. *FASEB J*, 24(12), 4711–21.
- [3] Kwon, M. Y., Wang, C., Galarraga, J. H., Puré, E., Han, L., y Burdick, J. A. (2019). Influence of hyaluronic acid modification on CD44 binding towards the design of hydrogel biomaterials. *Biomaterials*, 222, 119451.
- [4] Trujillo, S., Gonzalez-Garcia, C., Rico, P., Reid, A., Windmill, J., Dalby, M. J., y Salmeron-Sanchez, M. (2020). Engineered 3D hydrogels with full-length fibronectin that sequester and present growth factors. *Biomaterials*, 252, 120104.

DISCOVERING THE DRIVERS OF ABERRANT SPLICING IN CANCER

A. Añorbe de Frutos¹, A. San Martín Retegi¹ and J.A. Ferrer-Bonsoms¹

¹ Department of Biomedical Engineering and Science, TECNUN (University of Navarra), San Sebastián, Spain, aanorbe@alumni.unav.es

Abstract

Alternative splicing (AS) is a posttranscriptional process of mRNA that is known to be related with cancer. RNA Binding Proteins (RBPs), the conductors of AS, are related with cancer prognosis and are also targets of novel cancer treatments. The specific relationship between RBPs and AS is still unknown. In this research, we developed a statistical methodology to infer which RBPs are related with the AS of different cancers. We fine-tuned the hyperparameters of the statistical method using experiments where specific RBPs were knocked-down. Using these hyperparameters, we applied the algorithm to The Cancer Genome Atlas (TCGA), a database with a large amount of data on cancer. We found relationships between RBPs and cancers. For instance, SF3A3 and SF3B4 alterations in adenocarcinomas of the colon, lung, prostate, rectum, and stomach, as well as liver cancer. Some of these relationships have been validated in the literature but other ones are novel. Finally, a Shiny app was created to show the results interactively and simple way.

1. Introduction

Alternative splicing (AS) is a posttranscriptional process that removes introns and joins the remaining exons to generate mature mRNA from pre-mRNA in a tissue-specific way. These sequences give place to different mRNA species, called isoforms or transcripts that, in turn, lead to different proteins. As a result, these proteins may have different functions [1]. The complete set of mRNA isoforms in an organism is called transcriptome [2], [3]. All the hallmarks in cancer (angiogenesis, immortality, immune response avoidance, etc.) can be originated by alterations in splicing [1].

AS is regulated by the spliceosome which is composed of small nuclear RNAs, proteins, and polypeptides. Some proteins –including the ones in the spliceosome– interact with RNA in its various forms. They are called RBPs and they are conductors of AS. In the human species we can identify more than 1500 RBPs. The effect of these proteins on each specific gene is unclear. Some of them regulate essential activities and have a great impact on RNA metabolism. RBPs that regulate AS are known as splicing factors (SF).

In addition, alterations of RBPs can appear in cancers. Variations in both their expression and their location can influence the expression levels of oncogenes. For this

reason, RBPs can be targeted when developing cancer treatments [4].

Cross-linking and immunoprecipitation sequencing (CLIP-seq) experiments are the method of choice to measure which splicing events are being regulated by the selected protein. However, these experiments require to know in advance which is the protein to be studied. Given the splicing events, it would be very useful to provide an educated guess on which is the RBP leading to this behavior [5].

In a previous work [6], an algorithm that integrates different databases of CLIP experiments was developed to infer the potential RBPs conducting the changes in splicing. The underlying reasoning of this algorithm is to find enriched RBPs that bound to the alternative splicing events detected in a particular study. This algorithm has the advantage that only requires RNA-seq information to predict the RBPs involved in the splicing changes. Performing an enrichment analysis requires to make several assumptions –the enrichment statistical method, the detection of AS events, etc.– and the tune of different hyperparameters.

The aim of this study is twofold. Firstly, to fine tune the algorithm to improve the RBP predictions in order to study which parameters are the best predictors of AS. To do so, we run the algorithms on the data from several experiments where some RBPs were knocked down. Secondly, using these parameters, the enrichment was performed on 16 different types of cancers on the TCGA. We also developed a Shiny application to ease the analysis of the data and extracting conclusions from the scientific community. Interestingly, several novel RBPs seem to be directly related with different cancer types. Additional wet lab validations and the relationship of these RBPs with cancer prognosis will reinforce the evidences of this work.

2. Materials and methods

Experimental data. Two datasets were used in this study to fine-tune the parameters of the algorithm. In both cases, selected RBPs were knockdown in different conditions. Thus, in these datasets, the RBP which is driving the

splicing is known. In the first dataset PTBP1, ESRP2 and MNBL1 were knocked-down in samples of gastric cancer [7]. In the second data set FUS, TAF15, and TARDBP were knocked-down [6].

TCGA. We applied our RBPs prediction method to TCGA [8] which is a public database with clinical and genomic information of almost 11,000 patients of 16 types of cancer.

Transcript quantification. For the experimental data, the pseudo-alignment process were processed using Kallisto [9]. For the TCGA data, transcript abundances were downloaded from [10] where Kallisto were applied. Next, Event Pointer (EP) [11] was applied to the experimental and TCGA data to identify differentially spliced events.

Alternative Splicing Analysis. EP was applied to detect all possible splicing events in the Genecode24 reference transcriptome. EP quantifies the events using the Percent Spliced-In (PSI or Ψ) value. It is defined as the relative value of the expression of the transcripts that contain a certain exon over the absolute expression of all transcripts [6]. Moreover, EP allows to apply an expression filter to the data. This consists of removing the events that are barely expressed. This would give us more precise results. We applied EP statistics to know which events are statistically differentially spliced among the conditions under study. The one based on bootstrap was applied.

Events and RBPs relationship. A matrix relating which RBP is regulating each event is used. This matrix is called ExS (Events x Splicing Factors) [6]. It equals 1 when they are related and 0 when they are not. The data was obtained from 937 CLIP [12] experiments from 4 databases.

RBPs prediction. Once the results were obtained, we applied the proposed prediction function in which the main objective is to know which RBP is causing the splicing. This function requires the p-values corresponding to the EP statistical analysis step, and the ExS matrix. This is an enrichment process i.e. we analyze whether the number of events regulated by an RBP is enriched within the differentially spliced events.

Therefore, it is necessary to set a parameter to define which events are differentially spliced. It could be the n most differentially spliced events or a th threshold. The former, which by default is set to 1000, returns the event names in order of their p-value. The latter selects the events in which the p-value is lower than the given value. Moreover, it is also necessary to determine the statistical method to apply, which in our case is limited to Fisher Test and Poisson binomial distribution. A schema of this process is represented in Figure 1.

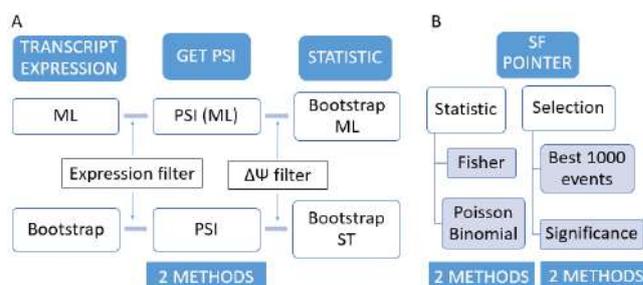


Figure 1. Schema of the procedure followed.

Finally, the prediction model returns a table with the RBPs with the corresponding p-values.

Moreover, GeneCards [13] and Ensembl [14] had been used to obtain additional biological information about RBPs. This information was recorded and an Excel document was created with the main function of each of the 195 RBPs.

Software. All this has been done using RStudio and the R programming language. Shiny has also been used, which is an R package that facilitates the creation of interactive web applications in a simple way.

3. Results

3.1. TCGA results

As already mentioned, TCGA results depend on four variables that correspond to the statistics used and the type of selection. However, in some cases, regardless of the statistics and selection used, we can observe that:

- SF3B4/SF3A3 have a relationship with adenocarcinomas of the colon, lung, prostate, rectum, and stomach, as well as liver cancer.
- SUB1 regulates oesophageal carcinoma (ESCA), as well as bladder urothelial carcinoma (BLCA).
- In the case of head and neck squamous cell carcinoma (HNSC), EFTUD2 and SF3B4 also regulate the AS.
- RNA Binding family may be involved in kidney chromophobe (KICH) as well as in rectum adenocarcinoma (READ).

In our study the threshold used was set at 0.01, which means that only RBPs with a p-value less than this value were taken into account.

Furthermore, heatmaps have been designed (Figure 2). In them, we can see in a visual way how the RBPs are affecting the different cancer types relying on the obtained prediction.

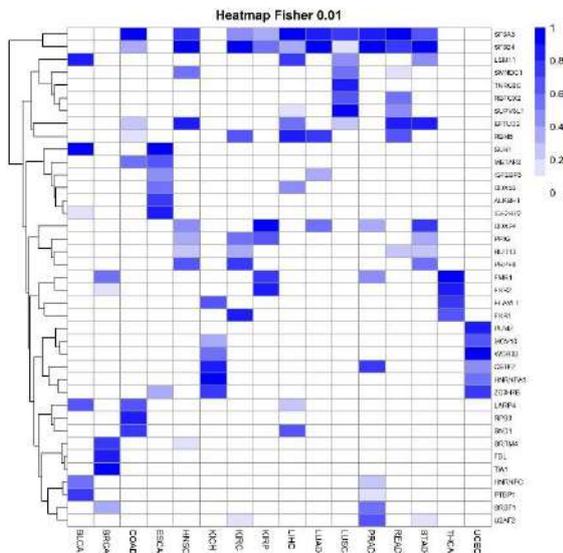


Figure 2. Heatmap applying Fisher and a threshold of 0.01 to TCGA. BLCA: Bladder urothelial carcinoma. BRCA: Breast invasive carcinoma. COAD: Colon adenocarcinoma ESCA: Esophageal carcinoma. HNSC: Head and neck squamous cell carcinoma. KICH: Kidney chromophobe. KIRC: Kidney renal clear cell carcinoma. KIRP: Kidney renal papillary cell carcinoma. LIHC: Liver hepatocellular carcinoma. LUAD: Lung adenocarcinoma. LUSC: Lung squamous cell carcinoma. PRAD: Prostate adenocarcinoma. READ: Rectum adenocarcinoma. STAD: Stomach adenocarcinoma. THCA: Thyroid carcinoma. UCEC: Uterine corpus endometrial carcinoma.

3.2. Shiny application

Once all the data is obtained, an application has been created with Shiny to display –in three different tabs– the results clearly and interactively. In this app, one of the 16 cancers and 2 of the 4 parameters can be chosen.

The first one represents the 10 most significant RBPs related to the selected cancer with their main function, p-value, and other statistical data (Figure 3). The second tab shows the differentially spliced events. Thus, the expression of events can be plotted to compare control and tumor samples. Lastly, the third tab corresponds to a survival study that will be developed in a future line.

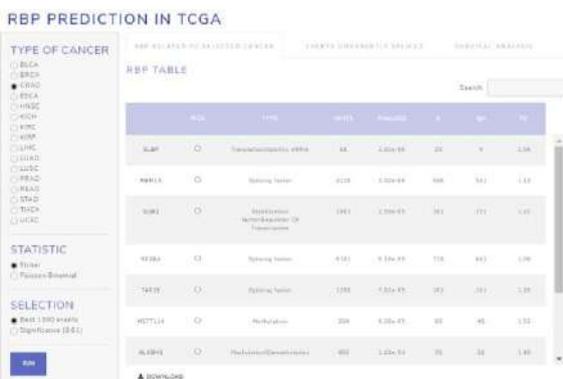


Figure 3. The first tab shows the table of RBPs related to the selected cancer ordered according to their p-value.

Moreover, all outputs can be downloaded. As there is a lot of data and samples, this app represents them dynamically.

4. Discussion

The results of our analysis show a clear relationship between RBPs and some cancers. It should be noted that variations in SF3A3 and SF3B4 have been detected in adenocarcinomas of the colon, lung, prostate, rectum, and stomach and liver cancer. There is also a link between SUB1 and ESCA and BLCA cancers. Other RBPs also have a slight relationship, such as RBFOX with READ and KICH as well as EFTUD2 and SF3B4 and HNSC.

When reviewing the existing bibliography, some articles link SF3A3 with lung and liver cancer [15]. Moreover, it has been suggested that modifying the concentration of this gene may be a useful therapeutic target [16].

Likewise, splicing factor SF3B4 has also been modified in cancer cells [17]. Specifically, in hepatocellular carcinoma, it has been seen that the concentration of it increases dramatically in tumor cells compared to non-tumor cells and that this increase is associated with the appearance of metastases and poor prognosis as well [18]. Knocking down SF3B4 inhibits the growth and metastasis in liver cancer, which suggests his role in cell proliferation [19].

These studies support our results regarding the demonstrated relationship between such RBPs and lung and liver cancer. However, our work shows a strong link between such genes and adenocarcinomas so more studies will be needed to investigate this possible association.

In the case of SUB1, our analysis shows its relationship with the ESCA. Other authors have noted that this RBP is frequently overexpressed in many cases. In addition, it has been shown that SUB1 upregulation can lead to a decrease in survival of the patient in some types of cancer [20].

Another RBP that shows an association with READ and KICH is RBFOX2. This family is enriched in some tumors. Furthermore, a clear relationship has been seen between RBFOX2 and splicing alterations in cancer [21].

In the case of HNSC cancer, our results establish a link between the RBPs SF3B4, and EFTUD2. Regarding the first, no bibliography supports this with this specific cancer, but a relationship has been seen with other cancers as already mentioned. Regarding EFTUD2, a clear relationship between this RBP and other diseases such as mandibulofacial dysostosis with microcephaly has been seen, but more studies are needed regarding cancer [22].

Other RBPs do not show clear enrichment, nor we have found any literature that supports these slight relationships.

Among the problems faced by this type of research, one is the existence of mutations of RBPs that, together with their overexpression and subexpression, can modify the activity

and concentration of other RBPs. Another issue is the large number of RBPs and the different types of cancer analyzed, due to the scarcity of the bibliography that relates them.

Likewise, dividing the different types of cancers according to the affected organ can lead to errors since, although they are of the same histological type, there may be multiple subgroups with different origins and be caused by different causes. By only studying the genetic aspect, other variables that influence the origin of cancer, such as exposure to an oncogenic factor (such as tobacco), as well as gender, age, or diseases that degenerate into malignant processes, are ignored. So, the results can be misleading [23].

Another limitation is the ExS matrix since it is not a complete matrix because the number of RBPs is very large and there is no information on all of them. However, due to their important roles, in the last years, they are being investigated as therapeutic targets for cancers [24].

The development of the app allows users to easily analyze the results. The aim of this app is to make the results public and available to all researchers.

This work can be enhanced by adding other databases such as the tumor alterations relevant for genomics-driven therapy (TARGET) that includes cancer types not present in TCGA. Further, another line of research is to relate these RBPs to survival, because the main objective of predicting AS is to improve survival.

5. Conclusions

In conclusion, a clear relationship has been seen between various RBPs and cancers. An application has been developed that allows the results to be displayed and downloaded. The objective is to facilitate the reading of these data. Also, more databases will be added in the future for more robust results.

6. References

- [1] Y. Wang, J. Liu, B. Huang, and Y.-M. Xu, "Mechanism of alternative splicing and its regulation," *Biomed. Reports*, vol. 3, no. 2, pp. 152–158, Mar. 2015, doi: 10.3892/br.2014.407.
- [2] F. Carazo, J. P. Romero, and A. Rubio, "Upstream analysis of alternative splicing: A review of computational approaches to predict context-dependent splicing factors," *Brief. Bioinform.*, vol. 20, no. 4, pp. 1358–1375, Mar. 2018, doi: 10.1093/bib/bby005.
- [3] D. Li, C. S. McIntosh, F. L. Mastaglia, S. D. Wilton, and M. T. Aung-Htut, "Neurodegenerative diseases: a hotbed for splicing defects and the potential therapies," *Transl. Neurodegener.*, vol. 10, no. 1, Dec. 2021, doi: 10.1186/s40035-021-00240-7.
- [4] D. Kang, Y. Lee, and J. S. Lee, "Rna-binding proteins in cancer: Functional and therapeutic perspectives," *Cancers (Basel)*, vol. 12, no. 9, pp. 1–33, Sep. 2020, doi: 10.3390/cancers12092699.
- [5] A. Busch and K. J. Hertel, "Splicing predictions reliably classify different types of alternative splicing," *RNA*, vol. 21, no. 5, pp. 813–823, May 2015, doi: 10.1261/rna.048769.114.
- [6] F. Carazo, M. Gimeno, J. A. Ferrer-Bonsoms, and A. Rubio, "Integration of CLIP experiments of RNA-binding proteins: A novel approach to predict context-dependent splicing factors from transcriptomic data," *BMC Genomics*, vol. 20, no. 1, Jun. 2019, doi: 10.1186/s12864-019-5900-1.
- [7] S. Cheng *et al.*, "A functional network of gastric-cancer-associated splicing events controlled by dysregulated splicing factors," *NAR Genomics Bioinform.*, vol. 2, no. 2, pp. 1–12, 2020, doi: 10.1093/nargab/lqaa013.
- [8] "The Cancer Genome Atlas." <https://www.cancer.gov/tcga> (accessed Oct. 18, 2021).
- [9] N. L. Bray, H. Pimentel, P. Melsted, and L. Pachter, "Near-optimal probabilistic RNA-seq quantification," *Nat. Biotechnol.*, vol. 34, no. 5, pp. 525–527, 2016, doi: 10.1038/nbt.3519.
- [10] P. J. Tatlow and S. R. Piccolo, "A cloud-based workflow to quantify transcript-expression levels in public cancer compendia," *Sci. Rep.*, vol. 6, no. November, p. 39259, 2016, doi: 10.1038/srep39259.
- [11] J. P. Romero *et al.*, "EventPointer: An effective identification of alternative splicing events using junction arrays," *BMC Genomics*, vol. 17, no. 1, Jun. 2016, doi: 10.1186/s12864-016-2816-x.
- [12] M. Ramanathan, D. F. Porter, and P. A. Khavari, "Methods to study RNA–protein interactions," *Nat. Methods*, vol. 16, no. 3, pp. 225–234, 2019, doi: 10.1038/s41592-019-0330-1.
- [13] Stelzer G *et al.*, "GeneCards – the human gene database," *The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analysis*, 2016. <https://www.genecards.org/> (accessed Aug. 20, 2021).
- [14] "Ensembl genome browser 104." <http://www.ensembl.org/index.html> (accessed Aug. 20, 2021).
- [15] N. Li, L. Zhao, C. Guo, C. Liu, and Y. Liu, "Identification of a novel DNA repair-related prognostic signature predicting survival of patients with hepatocellular carcinoma," *Cancer Manag. Res.*, vol. 11, pp. 7473–7484, 2019, doi: 10.2147/CMAR.S204864.
- [16] E. Siebring-van Olst, M. Blijlevens, R. X. de Menezes, I. H. van der Meulen-Muileman, E. F. Smit, and V. W. van Beusechem, "A genome-wide siRNA screen for regulators of tumor suppressor p53 activity in human non-small cell lung cancer cells identifies components of the RNA splicing machinery as targets for anticancer treatment," *Mol. Oncol.*, vol. 11, no. 5, pp. 534–551, 2017, doi: 10.1002/1878-0261.12052.
- [17] F. Xiong and S. Li, "SF3b4: A Versatile Player in Eukaryotic Cells," *Front. Cell Dev. Biol.*, vol. 8, no. January, pp. 2016–2021, 2020, doi: 10.3389/fcell.2020.00014.
- [18] T. Iguchi *et al.*, "Increased Copy Number of the Gene Encoding SF3B4 Indicates Poor Prognosis in Hepatocellular Carcinoma," *Anticancer Res*, vol. 36(5), pp. 2139–2144, 2016.
- [19] Q. Shen *et al.*, "Barrier to autointegration factor 1, procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3, and splicing factor 3b subunit 4 as early-stage cancer decision markers and drivers of hepatocellular carcinoma," *Hepatology*, vol. 67, no. 4, pp. 1360–1377, 2018, doi: 10.1002/hep.29606.
- [20] J. Zhang *et al.*, "An integrative ENCODE resource for cancer genomics," *Nat. Commun.*, vol. 11, no. 1, 2020, doi: 10.1038/s41467-020-14743-w.
- [21] E. Sebestyén *et al.*, "Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks," *Genome Res.*, vol. 26, no. 6, pp. 732–744, 2016, doi: 10.1101/gr.199935.115.
- [22] S. Y. Kim, D. H. Lee, J. H. Han, and B. Y. Choi, "Novel splice site pathogenic variant of EFTUD2 is associated with mandibulofacial dysostosis with microcephaly and extracranial symptoms in Korea," *Diagnostics*, vol. 10, no. 5, pp. 2–13, 2020, doi: 10.3390/diagnostics10050296.
- [23] V. M. L. De Sousa and L. Carvalho, "Heterogeneity in Lung Cancer," *Pathobiology*, vol. 85, no. 1–2, pp. 96–107, 2018, doi: 10.1159/000487440.
- [24] S. Mohibi, X. Chen, and J. Zhang, "Cancer the 'RBP' eutics–RNA-binding proteins as therapeutic targets for cancer," *Pharmacol. Ther.*, vol. 203, 2019, doi: 10.1016/j.pharmthera.2019.07.001.

Análisis de texturas como ayuda al diagnóstico neurooncológico en imágenes de resonancia magnética

Marta Romera^{1,2*}, Cristian Galán², Pere Ràfols^{1,3,4}

¹Department of Electronic Engineering, Rovira i Virgili University, Tarragona, Spain. ²ASCIREs, Barcelona, Spain. ³Spanish Biomedical Research Centre in Diabetes and Associated Metabolic Disorders (CIBERDEM), Madrid, Spain. ⁴Institut d'Investigació Sanitària Pere Virgili, Tarragona, Spain

*contacto a: marta.romera@estudiants.urv.cat

Resumen

Debido a la creciente preocupación del cáncer es necesario el desarrollo de sistemas de ayuda al médico que permitan diagnosticar y seleccionar el tratamiento más apropiado.

A priori, a partir de la simple observación de imágenes médicas de resonancia magnética no es posible conocer en detalle la tipología y extensión de los tumores. Pero se ha intentado explotar esta información mediante el análisis de texturas el que proporciona una amplia gama de características que permiten cuantificar la heterogeneidad distintiva de diferentes tejidos.

El objetivo de este trabajo se basa en observar el gran potencial de estas características que evitan emplear procedimientos invasivos como la biopsia, para poder discernir entre tumores cerebrales primarios y metástasis cerebrales en RM y diferenciar entre tumores con necrosis o sin.

El estudio retrospectivo realizado incluye 30 pacientes, 13 con metástasis cerebral, 9 con tumores cerebrales primarios necróticos y 8 sin necrosis.

Palabras clave: tumor cerebral primario; glioma; glioblastoma multiforme; metástasis cerebral; necrosis; análisis de texturas imágenes de resonancia magnética; Machine Learning; Matlab; Radiomics.

1. Motivación

El cáncer es la segunda causa principal de muerte a nivel mundial. Concretamente, el cáncer cerebral es un tipo de cáncer muy agresivo y equivale aproximadamente al 3% de los casos de cáncer de todo el mundo, siendo más frecuentes en hombres que en mujeres [1].

La tasa de supervivencia a los 5 años es baja, particularmente en edades más avanzadas, debido a que en la mayoría de casos no se presenta sintomatología asociada en fases tempranas.

Con este enfoque, se pretende proporcionar a los profesionales sanitarios una herramienta complementaria de apoyo a la decisión para dar una respuesta rápida a algunos desafíos médicos relacionados con las patologías cerebrales, ya que estos normalmente se resuelven demasiado tarde y además, empleando procedimientos invasivos, neuroimagen exhaustiva o pruebas neuropsicológicas.

2. Introducción

2.1. Tumores cerebrales

Un tumor cerebral o neoplasia cerebral equivale al crecimiento de células anormales en el tejido del cerebro. Los tumores pueden ser benignos, si permanecen en su ubicación original y no son cancerosos, o bien, malignos, cuando son capaces de invadir el tejido normal que lo rodea y propagarse por el cuerpo y son cancerosos [2].

Al mismo tiempo, según el origen del tumor, se pueden clasificar en primarios y secundarios.

Los tumores primarios causan aproximadamente el 2% de las muertes por cáncer [3] y según la OMS se puede clasificar en grados del I al IV, los grados I y II se refieren a los benignos o de bajo grado, mientras que los de grado III y IV tienen carácter maligno. Concretamente, el glioblastoma multiforme pertenece al grado IV [4] y hace referencia al glioma más agresivo y con mayor propagación. La supervivencia media de este es de 15-18 meses [5] y sólo un 5% de los pacientes pueden llegar a sobrevivir más de 5 años, debido a su propia naturaleza, con regiones tanto de necrosis como de hemorragias.

Los tumores cerebrales secundarios o metástasis cerebral son los tumores más frecuentes del SNC [6] y equivalen a la diseminación de células cancerosas a través del torrente sanguíneo, sistema linfático o nervios [7] desde el tumor original (primario) hasta otro órgano del cuerpo que provienen normalmente de cáncer de pulmón, mama o carcinoma de células renales, melanoma o cáncer colorrectal.

2.2. Analisis de texturas

El análisis de texturas equivale a un área emergente de la radiómica [2] que se basa en la aplicación de métodos matemáticos que permiten extraer parámetros de las imágenes como la distribución y relación de los niveles de gris entre los píxeles de una determinada región de interés (ROI) [8] [9]. Y posteriormente, a través del análisis y la interpretación de estos parámetros de textura se puede cuantificar la heterogeneidad intrínseca del tumor que suele ser imperceptible al ojo humano.

Las aplicaciones de análisis de textura en resonancia magnética implican un proceso general que consta de diferentes pasos que se muestran en la Figura 1.

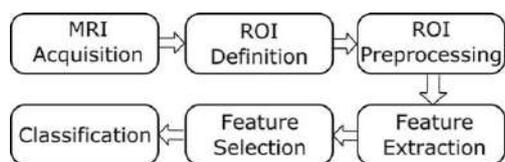


Figura 1. Proceso de análisis de texturas [9].

La resonancia magnética equivale al método imagenológico de elección para la evaluación y diagnóstico de tumor cerebrales debido que es una técnica no invasiva y segura que permite la caracterización y diferenciación de los tejidos blandos de múltiples áreas del cuerpo.

Además, las imágenes RM que se adquieren tienen una alta resolución y muy buen contraste. En función del contraste aplicado varias imágenes de RM.

Por un lado, las imágenes potenciadas en T1 aportan información anatómica completa, ya que muestran con más claridad los límites entre los diferentes tejidos. En este caso, los líquidos como la sustancia gris, se ve con baja intensidad de señal y aparecen imágenes oscuras (es decir, hipointensas), en cambio, la grasa como la sustancia blanca del cerebro se muestra con alta intensidad de señal [8].

Por otro lado, en las imágenes potenciadas en T2 los líquidos aparecen más brillantes (es decir, hiperintensos).

Y finalmente, la secuencia FLAIR es similar a una imagen potenciada en T2, pero reduciendo el brillo de los líquidos.

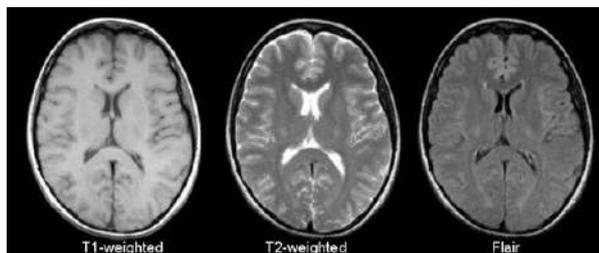


Figura 2. Comparación entre vistas axiales de imágenes potenciadas T1, potenciadas T2 y FLAIR (de izquierda a derecha). Estas permiten ver la diferencia de contraste entre secuencias de resonancia magnética de diferentes tejidos cerebrales [10].

3. Materiales

Las imágenes RM empleadas en este trabajo han sido proporcionadas por el PACS de la empresa ASCIRES.

En el presente trabajo, ha habido una participación de 51 sujetos, de los cuales 21 se han descartado debido que no cumplían con las características deseadas para la obtención de resultados.

Los criterios de inclusión que se han considerado han sido diversos. En primer lugar, el tumor cerebral debía haber sido confirmado patológicamente por el médico y neurólogo, en segundo lugar, no tenía que haber sufrido ningún tratamiento previo como la biopsia. Posteriormente, todas las imágenes debían contener las secuencias potenciadas en T1, T2 y FLAIR y finalmente, el tumor tenía que tener un tamaño suficientemente visible.

En cuanto a los casos que se han analizado, 20 han sido hombres y 10 mujeres, con un rango de edad entre los 39 y los 81 años y la edad media de estos equivale a 60.1667 ± 12.2258 . Concretamente, de los 30 pacientes totales, 13 tienen metástasis y respecto a los 17 restantes 9 han sufrido necrosis en el tumor que presentaban. Los correspondientes estudios han sido realizados entre el 11 de Marzo de 2016 y el 10 de Diciembre de 2020.

4. Metodología

La metodología empleada en el presente trabajo se puede dividir en varios pasos que satisfacen el procedimiento del análisis de texturas, necesario para conseguir los objetivos iniciales planteados.

En primer lugar, se han anonimizado las imágenes con Medexprim Suite™ y se realizó el correregistro de las imágenes RM con la herramienta Viewer 3D del programa AW Server de General Electrics donde es necesario combinar el sistema de coordenadas de la imagen en movimiento I^M con la imagen fija (o de referencia) I^F [11].

En segundo lugar, se lleva a cabo la segmentación manual del tumor cerebral y se vuelve a realizar el correregistro con la ROI extraída.

En tercer lugar, tiene lugar el preprocesado de la ROI donde se utiliza la normalización de datos para evitar picos de intensidad y la cuantización de niveles de grises divididos en 16, 32, 64, 128 y 256 niveles de gris que se utilizan para el cálculo de las características de la textura [8].

Posteriormente, se extraerán las características de textura mediante la toolbox Radiomics 2D de Matlab, implementada por Vallières [12] donde se obtienen un total de 129 texturas, correspondientes a 43 texturas a partir de las 3 secuencias RM y los 5 métodos estadísticos diferentes: Global, GLCM, GLRLM, GLSZM y NGTDM

El siguiente paso es la selección de aquellas características más relevantes con el algoritmo ReliefF con un rango de pesos entre -1 y 1, indicando poder clasificatorio bajo y elevado, respectivamente.

Finalmente, se ha entrenado con 3 modelos de clasificación: máquinas de soporte de vectores (SVM), K-vecinos más cercanos (KNN) y árboles de decisión (DT).

Y por otro lado, se ha validado el rendimiento de cada clasificador con cross-validación, del tipo K-fold, ayudando a que no se produzca sobreestimación, dividiendo la muestra en $k = 5$ partes el grupo entrenamiento y el grupo test y calculando la precisión en cada una.

5. Resultados

La obtención de resultados se llevará a cabo por cada modelo donde observará y se comparará diversas métricas como son los vectores de AUC y la matriz de confusión junto con los valores correspondientes de exactitud, sensibilidad y especificidad.

Este proceso se llevará a cabo por cada uno de los objetivos planteados que aparecen en las secciones siguientes.

5.1. Diferenciación entre tumores cerebrales primarios y secundarios (metástasis)

En primer lugar, obtendremos los valores del área bajo una curva ROC (AUC) ya que representa la probabilidad de que un dato se clasifique correctamente.

En la Tabla 1 podemos observar como los valores máximos de AUC se obtienen empleando el modelo SVM. Y el mejor resultado, corresponde al 0.977 donde se utilizan 60 características y un nivel de gris de 64. Al mismo tiempo, podemos observar cómo por este valor óptimo de AUC se ha obtenido el porcentaje de exactitud correspondiente al 86.67%, lo que indica una correcta clasificación.

CLASSIFICADOR	SVM					KNN					DT				
	16	32	64	128	256	16	32	64	128	256	16	32	64	128	256
Número de características	52	6	60	75	96	8	11	83	106	71	96	9	6	6	36
AUC	0.74	0.88	0.977	0.923	0.905	0.79	0.75	0.825	0.883	0.906	0.629	0.852	0.759	0.817	0.818
Sensibilidad (%)	53.85	69.23	92.31	92.31	76.02	59.23	76.92	84.62	92.31	1	61.54	84.62	76.92	84.62	76.92
Especificidad (%)	70.59	76.47	82.35	81.24	78.58	88.24	70.59	76.47	82.35	82.35	70.59	88.24	76.47	82.35	71.43
Exactitud (%)	63.33	73.33	86.67	90	73.33	80	73.33	80	86.67	90	67.67	86.67	76.67	80	80

Tabla 1. Valores de AUC, número de características, porcentajes de sensibilidad, especificidad y exactitud para diferenciar tumores primarios y secundarios.

Posteriormente, con el fin de saber de las características más relevantes obtenidas con el filtro ReliefF a que secuencia RM pertenecen, se buscan los porcentajes correspondientes a cada una. De todas las secuencias (T1, T2 y FLAIR) resulta de gran importancia la FLAIR.

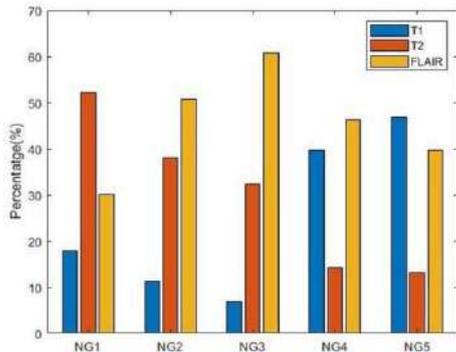
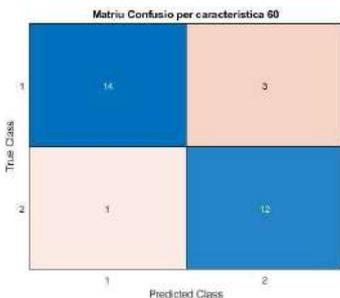


Figura 3. Comparación de los valores de los porcentajes entre las secuencias por los 5 niveles de gris.

Por último, se obtiene la matriz de confusión (MC) que nos da la información de fiabilidad del modelo. Esto, se observa en la Figura 4, en esta se predicen correctamente 12 casos con metástasis (TP) y 14 casos que no tienen (TN). Sin embargo, también podemos observar como 1 paciente ha sido predicho incorrectamente pero realmente tenía metástasis (FN) y los 3 restantes (FP) habían sido predichos como positivos y no tenían metástasis.



PACIENT	TEST	PREDIT
GL_001	NO	NO
GL_002	NO	NO
GL_003	NO	NO
GL_005	NO	NO
GL_006	NO	SI
GL_007	NO	NO
GL_012	NO	SI
GL_013	NO	NO
GL_014	NO	NO
GL_015	NO	NO
GL_017	NO	NO
GL_019	NO	NO
GL_020	NO	NO
GL_021	NO	NO
GL_022	NO	SI
GL_024	NO	NO
GL_026	NO	NO
ML_003	SI	SI
ML_006	SI	SI
ML_012	SI	SI
ML_013	SI	SI
ML_014	SI	NO
ML_016	SI	SI
ML_018	SI	SI
ML_019	SI	SI
ML_020	SI	SI
ML_022	SI	SI
ML_023	SI	SI
ML_024	SI	SI
ML_025	SI	SI

Figura 4. Matriz de confusión obtenida para diferenciar tumores primarios y secundarios, junto con los pacientes correspondientes.

5.2. Distinción entre tumores cerebrales primarios con necrosis y sin

Por otra parte, para distinguir entre tumores necróticos y no necróticos tenemos que el modelo SVM es el que obtiene mejores resultados de AUC de nuevo. Específicamente, por Ng 16 y empleando 30 características de texturas se encontraría el valor máximo de AUC correspondiente a un valor de 0.9. Además, podemos observar cómo por este valor se obtiene un porcentaje de exactitud, de 70.59%, es decir, equivale a una clasificación bastante correcta (Tabla 2).

CLASSIFICADOR	SVM					KNN					DT				
	16	32	64	128	256	16	32	64	128	256	16	32	64	128	256
Número de características	30	1	1	1	13	1	9	105	2	52	2	1	1	1	8
AUC	0.90	0.85	0.85	0.80	0.85	0.65	0.85	0.70	0.65	0.70	0.85	0.70	0.70	0.45	0.77
Sensibilidad (%)	66.67	77.78	88.89	77.78	90	55.56	55.56	66.67	66.67	90	88.89	66.67	88.89	55.56	50
Especificidad (%)	75	62.5	50	12.5	71.43	62.52	87.5	75	62.5	42.86	75	87.5	50	37.5	71.43
Exactitud (%)	70.59	70.59	70.59	47.06	82.25	58.82	70.59	70.59	64.71	70.59	82.25	76.47	70.59	47.06	58.82

Tabla 2. Valores de AUC, número de características, porcentajes de sensibilidad, especificidad y exactitud obtenidos por cada modelo (SV, KNN y DT) y nivel de gris para la diferenciación del grado de malignidad del tumor primario (con necrosis y sin).

A continuación, se seleccionan las características con peso positivo y se muestra en la Figura 5 como la secuencia más determinante ha sido la potenciada T1.

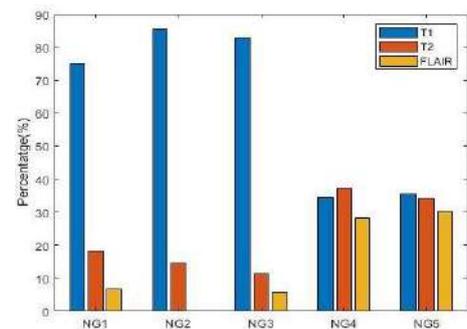
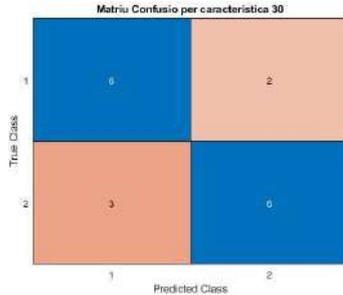


Figura 5. Comparación de los valores de los porcentajes entre las secuencias por los 5 niveles de gris para la diferenciación entre necrosis y sin esta.

Finalmente, en este caso, la MC por el valor máximo de AUC (Figura 6) representa como se predicen correctamente como verdaderos positivos (TP) 6 casos y como verdaderos negativos (TN) otros 6 casos.

Mientras que falsos negativos (FN) equivale a 3 casos, es decir, tenían necrosis y se han clasificado como negativos y finalmente, como falsos positivos (FP) se detectan 2 casos, es decir, eran casos que no tenían necrosis y se han identificado como si tuvieran.



PACIENT	TEST	PREDIT
GL_001	NO	NO
GL_002	SI	SI
GL_003	NO	NO
GL_005	SI	NO
GL_006	NO	NO
GL_007	SI	NO
GL_012	SI	SI
GL_013	SI	SI
GL_014	NO	NO
GL_015	SI	SI
GL_017	NO	NO
GL_019	SI	NO
GL_020	SI	SI
GL_021	SI	SI
GL_022	NO	SI
GL_024	NO	SI
GL_026	NO	NO

Figura 6. Matriz de confusión obtenida para diferenciar tumores entre necrosis y sin, junto con los pacientes correspondientes.

6. Conclusiones

A lo largo del trabajo se han obtenido varias conclusiones que satisfacen los objetivos iniciales presentados. En primer lugar, se ha podido comprobar la importancia que tiene el uso del análisis de texturas de imágenes de resonancia magnética en pacientes con tumores cerebrales, ya que a partir de este se logra obtener una gran cantidad de información de la patología sin tener que hacer una biopsia.

Posteriormente, se ha empleado una buena metodología que permite la obtención de datos estadísticos, las cuales con su posterior análisis, ayudan a clasificar los tumores cerebrales primarios en función del grado al que pertenecen, gracias a la observación de la existencia de tejido necrótico al tumor. O bien, la distinción del origen del tumor, en función de si son primarios o secundarios.

Los resultados preliminares obtenidos en el trabajo se pueden clasificar con un buen nivel de precisión ya que se han obtenido unas AUC s máximas iguales o por encima del 90%.

Como conclusión, a la vista de estos resultados prometedores, resulta de elevada importancia continuar con el desarrollo de instrumentos de evaluación de imágenes y datos que permitan la combinación de información clínica, radiológica y textural como primer diagnóstico a partir de RM, ya que sin duda las técnicas de aprendizaje automático desempeñan un papel indispensable en este ámbito.

Agradecimientos

Me gustaría expresar mi gratitud sincera a mi tutor Pere Ràfols por su supervisión. A mis cotutores de la empresa Jose Manuel y Eva, por darme la posibilidad de poder realizar mi trabajo dentro de la empresa ASCIRES. Y en especial, a mi tutor de empresa Cristian, por aconsejarme, y escucharme. Y finalmente, también, a mi familia y amigos, por la confianza que me han transmitido desde el primer día.

Referencias

- [1] Sylvia C Eisele and Jörg Dietrich. Necrosis cerebral por radiación: desafío diagnóstico y tratamiento clínico. Rev. neurol. (Ed. impr.), pages 225–232, 2015.
- [2] Cooper GM. The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates; 2000. The Development and Causes of Cancer.
- [3] MOLINA, F., et al. Factores pronósticos en los tumores cerebrales. En Anales del Sistema Sanitario de Navarra. 2001. p. 63-72.
- [4] D. N. Louis, A. Perry, G. Reifenberger, A. von Deimling, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, O. D. Wiestler, et al., "The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary," Acta Neuropathol., vol. 131, no. 6, pp. 803–820, Jun. 2016.
- [5] VENUR, Vyshak Alva; PEEREBOOM, David M.; AHLUWALIA, Manmeet S. Current medical treatment of glioblastoma. Current Understanding and Treatment of Gliomas, 2015, p. 103-115.
- [6] GAVRILOVIC, Igor T.; POSNER, Jerome B. Brain metastases: epidemiology and pathophysiology. Journal of neuro-oncology, 2005, vol. 75, no 1, p. 5-14
- [7] WELCH, Danny R.; HURST, Douglas R. Defining the hallmarks of metastasis. Cancer research, 2019, vol. 79, no 12, p. 3011-3027.
- [8] Larroza, Andrés, Vicente Bodí, and David Moratal. "Texture analysis in magnetic resonance imaging: review and considerations for future applications." Assessment of cellular and organ function and dysfunction using direct and derived MRI methodologies (2016): 75-106.
- [9] Castellano, Gabriella, et al. "Texture analysis of medical images." Clinical radiology 59.12 (2004): 1061-1069.
- [10] Rovira, A., Tintoré, M., Álvarez-Cermeño, J. C., Izquierdo, G., & Prieto, J. M. (2010). Recomendaciones para la utilización e interpretación de los estudios de resonancia magnética en la esclerosis múltiple. Neurología, 25(4), 248-265.
- [11] Gehrung, M., Tomaszewski, M., McIntyre, D., Disselhorst, J., & Bohndiek, S. (2020). Co- registration of optoacoustic tomography and magnetic resonance imaging data from murine tumour models. Photoacoustics, 18, 100147.
- [12] Mvallieres. "Mvallieres/Radiomics." GitHub, github.com/mvallieres/radiomics.

Análisis de las variables biomecánicas para el estudio del índice de ruptura de la Triple A

E. Navarro Ruiz¹, M.A. Pardo Picazo,²

¹ Estudiante Master en Inteligencia Artificial, Universidad Internacional de Valencia, enriquenavaruz26@gmail.com

² Departamento de Ingeniería Civil, Universidad de Alicante, Alicante, España mpardo@ua.es

Resumen

Los métodos computacionales representan una fenomenal ayuda para mejorar el diagnóstico de las Aneurismas de Aorta Abdominal. El uso de estas herramientas de análisis de imágenes clínicas facilita al especialista información real del aneurisma gracias a la representación tridimensional exacta de la estructura anatómica. De cada imagen se puede extraer variables relativas a su morfología y relacionarlas entre sí para encontrar un índice de ruptura que prevea cuando la aorta se pueda romper. La dinámica de fluidos computacional resuelve las ecuaciones de continuidad y de Navier-Stokes y permite la simulación del comportamiento del fluido sanguíneo en el lumen del aneurisma. Se obtienen como resultados la evolución en régimen transitorio de las distintas velocidades, presiones y tensiones tangenciales de cada aneurisma.

1. Introducción

Las Aneurismas de Aorta Abdominal (AAA o Triple A) es una patología de base arteriosclerótica y se define como la dilatación permanente, localizada e irreversible de la aorta abdominal ubicada desde la zona proximal de las arterias renales hasta la bifurcación ilíaca. Las AAA son unas de las principales enfermedades cardiovasculares donde presenta mundialmente el 1.3% de las muertes en personas con una edad superior a 55 años. (Sakalihan et al., 2018)

Los criterios clínicos convencionales para la detección de la AAA se centran en la medición del diámetro transversal máximo y su tasa de crecimiento. Una aorta en condiciones normales mide aproximadamente 1.9 y 2.3 centímetros de diámetro máximo para mujeres y hombres respectivamente. Se recomienda tratamiento cuando esta magnitud alcanza 5 cm (mujeres) o 5.5cm (varones) o si su tasa de crecimiento es superior a 0.5cm/año (Ortiz-Monzón et al., 2004). El tratamiento convencional consiste en una reparación quirúrgica abierta o una reparación endovascular.

La dilatación de la aorta genera un cambio en las paredes arteriales, modificando su composición y sus propiedades mecánicas. Estos cambios pueden originar roturas con gravísimas consecuencias con una alta probabilidad de óbito. Y es que más del 30% de los casos la evolución de las AAA se produce de manera asintomática (Sakalihan et al., 2018). Cuando los síntomas se manifiestan, suelen aparecer como dolores abdominales y lumbares, pulso en el abdomen y/o fatiga (Vilalta et al., 2012).

En este trabajo, se pretende analizar el comportamiento fluidomecánico de la sangre (fluido no newtoniano) mediante la simulación de su flujo en régimen estacionario a través de las Aneurismas Aórticos Abdominal. Para ello, se relacionan las variables morfométricas de las AAA, los factores biomecánicos, el comportamiento del fluido sanguíneo y la interacción fluido-estructura para la determinación y/o cálculo del riesgo de rotura. Se establecen tres objetivos principales:

- Representación tridimensional de las AAA tras el procesamiento de las pruebas diagnósticas y cuantificación de los parámetros afines a cada caso.
- Empleo de la técnica de dinámica de fluidos computacional (CFD) para la simulación del comportamiento de la sangre en el lumen de la aorta abdominal.

Para materializar este trabajo, se procesarán las imágenes diagnósticas de AAA (apartado 2.1) en formato DICOM para generar un modelo tridimensional y exportarlo en formato STL (apartado 2.2) y STEP (apartado 3.3). Se extraerán las variables que describen la geometría de la AAA para calcular el índice de ruptura de cada caso y por último se aplicarán las ecuaciones que rigen la física de los fluidos para realizar una simulación de la sangre en el dominio de la AAA.

2. Metodología

2.1. Procesamiento de imágenes clínicas para generación de modelos 3D de AAA

Esta primera etapa engloba el tratamiento de las imágenes clínicas Tomografía Axial Computarizada (TAC). Considera todos los trabajos desde la adquisición de la imagen hasta su modelación tridimensional (3D). En este proceso los datos de la imagen, recogidas en formato DICOM, se procesan para conseguir una representación volumétrica de los AAA y obtener datos que describan la morfología del modelo. Las imágenes clínicas se tratan con el software *Mimics Innovation Suite*, que dispone de numerosas herramientas de procesamiento de datos aplicado a la anatomía. Una vez importadas las imágenes, el software devuelve tres vistas de la aorta (axial o transversal, sagital y coronal) con las que ubicar el aneurisma.

El proceso continúa con la delimitación de una estructura anatómica en la prueba diagnóstica, comúnmente designado como segmentación. Para que la segmentación sea exitosa se emplea la escala de Hounsfield (TD and J, 2019). Esta escala muestra una unidad de absorción (HU) y relaciona la energía de los rayos incidentes del TAC sobre cada estructura anatómica que atraviesa. Se proporciona un valor de 0 HU al agua, -1000 HU al aire y +1000 HU al hueso compacto.

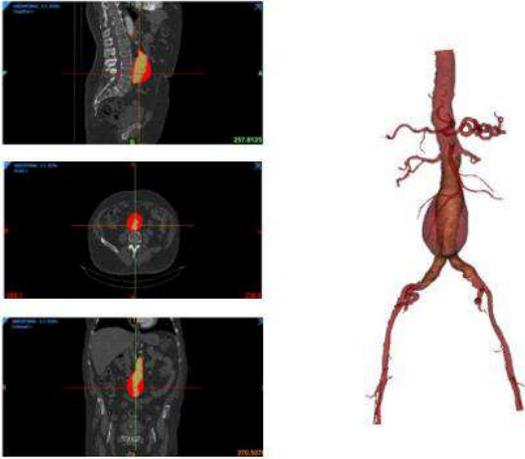


Figura 1. Segmentación AAA en pruebas diagnósticas y representación tridimensional.

El procesado de todas las imágenes del TAC en el que se han delimitado las estructuras de estudio (Aorta y lumen), continúa con la reconstrucción volumétrica del 3D de la AAA como se muestra en la *Figura 1*. Para un mejor acabado y visualización del modelo 3D, se emplea el módulo *3Matic* del software *Mimics Innovation Suite*. Se suele aplicar un suavizado (débil) que optimice el archivo (imagen), mostrando toda la información sin trabajar con un fichero muy pesado.

2.2. Biodeterminantes geométricos

En la siguiente etapa se parte de los modelos tridimensionales en formato STereoLithography (STL) para buscar e identificar los índices que ayudan en la detección de la AAA a partir de las estructuras segmentadas en formato STL, se extraen las siguientes variables geométricas (véase *Figura 2*):

- Diámetro máximo transversal, D : diámetro máximo del aneurisma: Se define el punto O como el punto de intersección entre los extremos de la AAA y el plano que define el diámetro máximo.
- Diámetro de la luz, DL : Diámetro efectivo del vaso a través del cual fluye la sangre.
- Diámetro infrarrenal, d : diámetro de la aorta abdominal localizada justo en el inferior de las aortas renales y antes de la aneurisma.
- Espesor de la pared, t : Promedio del grosor de la pared aneurismática.
- Longitud, L : Distancia entre la AAA proximal y distal.
- Longitud Anterior, LA : Es la distancia entre el punto O y la pared de la aneurisma anterior en el plano de diámetro máximo.

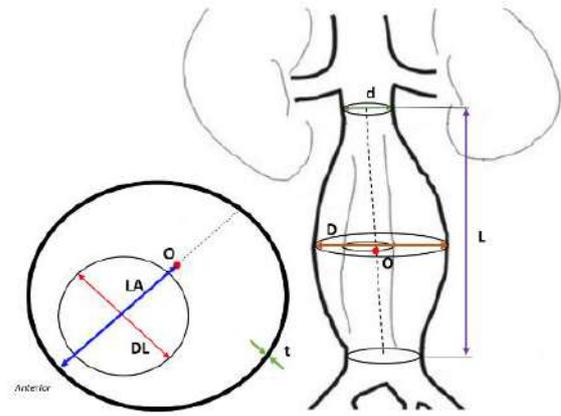


Figura 2. Medidas geométricas AAA

Las relaciones de los parámetros anteriores (Navarro Ruiz, 2021), permiten la definición de varios factores biomecánicos geométricos (FBG) para calcular el índice de ruptura de los casos de AAA, mostrando así, las sugerencias o acciones a realizar.

2.3. Simulación fluido sanguíneo en AAA mediante CFD

La simulación se realiza con el software *ANSYS* en su versión 2020R2 (*Fluent*) que utiliza la dinámica de fluidos computacional.

2.3.1 Ecuaciones de conservación de la masa y de la energía

La sangre se comporta como flujo incompresible y no newtoniano, donde la viscosidad varía en función de la tensión tangencial aplicada. La ecuación de continuidad para fluidos es un principio fundamental de la conservación de la masa y se define en su forma diferencial como:

$$\partial t + \nabla(\rho v) = 0 \quad (1)$$

Donde ρ es la densidad del fluido (sangre) en kg/m^3 , t el tiempo medido en segundos y v la velocidad del fluido medido en m/s .

Y, por otro lado, las ecuaciones en derivadas parciales de Navier-Stokes describen el movimiento de un fluido viscoso:

$$\rho \left(\frac{\partial v}{\partial t} + v \cdot \nabla v \right) = -\nabla p + \mu \nabla^2 v + f \quad (2)$$

donde, p es la presión medida en Pa (N/m^2), μ es la viscosidad dinámica medida $\text{N}\cdot\text{s/m}^2$ y f son las fuerzas externas.

Asimismo, en la ecuación de Navier-Stokes se ha de considerar que la viscosidad (μ) no es una constante (flujo no newtoniano). La relación entre la velocidad y la tensión tangencial siempre depende de la velocidad tangencial al lumen, pero en este caso, el fluido (la sangre) se vuelve menos viscosa cuando aumenta la velocidad de cizallamiento. Para trasladar estas características a las ecuaciones del movimiento sanguíneo se emplea el modelo de Carreau (YI and KR, 1991), ya que es el modelo que más se ajusta al comportamiento de la sangre ya que presenta varias condiciones reales como:

- Decrecimiento de la viscosidad cuando incrementa la tensión tangencial
- Régimen del fluido laminar con un número de Reynolds inferior a 2300 para evitar la hemólisis.
- Densidad de la sangre 1056 kg/m³
- Condiciones de flujo pulsátil o continuo.

La expresión que define la viscosidad en este modelo es:

$$\mu_{efe}(\dot{\gamma}) = \mu_{inf} + (\mu_0 - \mu_{inf})(1 + (\lambda\dot{\gamma})^2)^{\frac{n-1}{2}} \quad (3)$$

Donde se tiene que μ_{efe} es la viscosidad efectiva, μ_0 , μ_{inf} , λ y n son coeficientes de material. Para el caso de la sangre, los coeficientes son $\mu_0 = 0.056$ (kg/(m · s)), $\mu_{inf} = 0.0035$ (kg/(m · s)), $\lambda = 3.313$ (s) y $n = 0.3568$

Además, el flujo sanguíneo de una aorta abdominal se puede aproximar al flujo de Poiseuille en un conducto cilíndrico (de la Portilla de Juan and López, 2016), por lo que se puede especificar tanto las velocidades (u_r) y las tensiones tangenciales (Wall Shear Stress ó WSS).

$$u_r = \frac{\Delta p}{4\mu}(R^2 - r^2) \quad (4)$$

$$WSS = 4\mu \frac{u_r}{R} \quad (5)$$

Donde Δp es la variación de presiones entre dos puntos, y μ es la viscosidad dinámica. R corresponde con el radio de la luz y r la distancia a la que se requiere calcular la velocidad u_r .

2.3.2 Condiciones de contorno

Desde la óptica física, las condiciones de la pared del lumen dictan que la velocidad en la pared es nula. Para conocer la velocidad del fluido sanguíneo en la sección de entrada, se debe de conocer cuál es la evolución temporal del caudal en la AAA (flujo sanguíneo pulsátil). Dada la bibliografía (Leung et al., 2006) se recogieron datos de la velocidad del flujo sanguíneo en el tramo abdominal y se realizó un ajuste como una serie de Fourier de orden 3 con un 95% de confianza para obtener una función de la velocidad de entrada en un ciclo cardiaco. De la misma manera, se define también cuales son los datos de salidas, es decir, las presiones en las secciones de la bifurcación iliaca. (Figura 3)

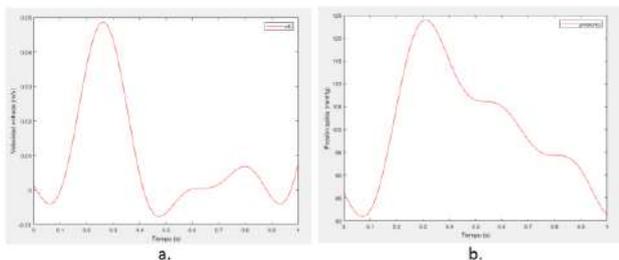


Figura 3. a) $V_{in}(t)$ y b) $P_{out}(t)$

2.3.3 Opciones de cálculo

En el presente trabajo se analiza el fenómeno en régimen transitorio que incluye un ciclo cardiaco de duración 1 s.

Primeramente, se importa el modelo en formato STEP y se discretiza el volumen geométrico, mediante triangulación,

para obtener relaciones entre los puntos que componen la geometría tridimensional. Este proceso (“mallado”) está limitado por la longitud de separación entre los nudos de la triangulación, que se ha supuesto inferior a 2 mm. La Figura 4 representa el proceso anteriormente descrito y muestra en su parte derecha las variables calculadas: velocidad, presión y tensión tangencial. El usuario debe definir la frecuencia de recogida de datos, siendo ésta una decisión compleja, ya que muchos datos suponen ficheros de trabajo muy pesados y difíciles de trabajar, mientras que pocos datos pueden suponer la pérdida de información. En el presente trabajo se decide recoger 100 datos en un tiempo total de 1 segundo con una frecuencia de 0.01 segundos.

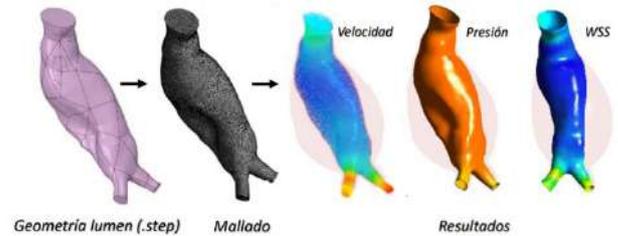


Figura 4. Simulación flujo sanguíneo en AAA

3. Caso de estudio

Se estudian distintas aortas abdominales que han sido segmentados y representados tridimensionalmente. Estos casos se presentan en la Tabla 1, y se muestran diversos diagnósticos, ya que los CASOS 01, 02 y 03 presentan una aneurisma de aorta abdominal, mientras que el CASO 04 presenta una intervención quirúrgica mediante “stent” y el CASO 05 es una aorta abdominal sana.

3.1. Resultados

Caso	3D	Resultados
		AAA
01		V_{max} : 1.025 m/s ² en bifurcación iliaca Presión (Max-Min): 17000-10790 Pa WSS _{max} : 13.87 Pa en bifurcación iliaca
		AAA
02		V_{max} : 1.968 m/s ² en bifurcación iliaca Presión (Max-Min): 17970-19060 Pa WSS _{max} : 26.98 Pa en bifurcación iliaca
		AAA
03		V_{max} : 0.783 m/s ² Infrarrenal y bifurcación Presión (Max-Min): 16670-10810 Pa WSS _{max} : 11.38 Pa Infrarrenal y bifurcación
		AAA con reparación por ‘stent’
04		V_{max} : 9.240 m/s ² en bifurcación iliaca Presión (Max-Min): 16710-10760 Pa WSS _{max} : 27.17 Pa. En la reparación
		Sano
05		V_{max} 3.399 m/s ² en bifurcación iliaca Presión (Max-Min): 31040-8289 Pa WSS _{max} : 61.14 Pa en bifurcación iliaca

Tabla 1. Casos de estudio, representación tridimensional y resultados de simulación.

Los resultados más significativos de las simulaciones mediante CFD se muestran en la *Tabla 1*. Las variables que se presentan son la velocidad máxima y su ubicación, las presiones máximas y mínimas en el lumen de la AAA y las tensiones tangenciales de la pared (WSS) y su localización.

4. Discusión y conclusiones

4.1. Discusión

Las velocidades de entrada y presiones de salida para cada uno de los casos se ajustan a los valores propuestos (Leung et al., 2006). Del mismo modo, los casos con AAA presentan resultados similares obteniendo mayores resultados de velocidad y WSS en las bifurcaciones ilíacas y en la zona infrarrenal.

Para el Caso 04 que presenta una reparación endovascular de aneurismas los resultados son los esperados. Las velocidades y tensiones tangenciales son mayores en las zonas donde se ha producido la reparación generando una mayor tensión en el 'stent' con el fin de no transmitírselo a la pared aneurisma como demostró (Li and Kleinstreuer, 2005). Por lo que el estudio clínico indica que la intervención quirúrgica fue satisfactoria desde la óptima de la dinámica de fluidos y mecánica de biosólidos.

Para el Caso 05 que presenta una aorta abdominal sana, los resultados son altos en comparación con los que sufre una aneurisma abdominal. Las presiones varían entre los 200 y 60 mmHg y WSS son 3-4 veces mayor que en una AAA. En el estudio (Alishahi et al., 2011) las simulaciones en aortas sanas presentan resultados de WSS y velocidades mayores. Esto podría deberse a que una aorta sana presenta unas características fisiológicas distintas y mejores a las que posee una AAA, como se demuestra en los estudios.

Cabe destacar que los máximos valores de tensiones tangenciales y velocidades surgen en el momento de la sístole, por lo tanto, se podría acotar el tiempo del ciclo cardiaco al tiempo de la fase sistólica. (Truijers et al., 2007)

4.2. Conclusiones

Los objetivos propuestos en el presente estudio se han cumplido consiguiendo realizar una representación tridimensional exacta de cada AAA obteniendo casos con características diferentes para tener un estudio con variabilidad. Las variables morfométricas de cada caso han sido extraídas para cuantificarlas, relacionarlas y calcular un índice de ruptura.

Para la realización de la simulación del flujo sanguíneo en el lumen (CFD) se ha caracterizado el modelo sanguíneo, la velocidad de entrada y la presión de salida. Los resultados obtenidos muestran velocidades máximas en los tramos más estrechos de la AAA (algo totalmente lógico y esperado) y, por tanto, mayores tensiones tangenciales en dichas ubicaciones (casi siempre en la bifurcación ilíaca). Las presiones son mayores en la fase sistólica y coincide en los tramos donde las velocidades y tensiones tangenciales son máximas. La simulación proporciona información relativa del flujo sanguíneo, interacción fluido-pared y enfatiza los puntos críticos o los puntos candidatos a producirse una ruptura de la pared.

Referencias

- Alishahi, M., Alishahi, M.M., Emdad, H., 2011. Numerical simulation of blood flow in a flexible stenosed abdominal real aorta. *Sci. Iran.* 18, 1297–1305. <https://doi.org/10.1016/j.scient.2011.11.021>
- de la Portilla de Juan, F., López, L.B.M., 2016. Apuntes sobre biomecánica de fluidos para ingenieros biomédicos y de la salud. FOINCO.
- Leung, J.H., Wright, A.R., Cheshire, N., Crane, J., Thom, S.A., Hughes, A.D., Xu, Y., 2006. Fluid structure interaction of patient specific abdominal aortic aneurysms: A comparison with solid stress models. *Biomed. Eng. Online* 5, 1–15. <https://doi.org/10.1186/1475-925X-5-33>
- Li, Z., Kleinstreuer, C., 2005. A New Wall Stress Equation for Aneurysm-Rupture Prediction. *Ann. Biomed. Eng.* 33, 209–213. <https://doi.org/10.1007/s10439-005-8979-2>
- Navarro Ruiz, E., 2021. Análisis de las variables biomecánicas para el estudio del índice de ruptura de la Triple A.
- Ortiz-Monzón, E., Gómez-Palónés, F., Plaza-Martínez, A., Zaragoza-García, J.M., Blanes-Mompó, J.I., Martínez-Meléndez, S., Martínez-Perelló, I., Crespo-Moreno, I., Briones-Estébanez, J.L., 2004. Endovascular treatment of ruptured abdominal aortic aneurysms. *Angiologia* 56, 169–181. [https://doi.org/10.1016/s0003-3170\(04\)74860-9](https://doi.org/10.1016/s0003-3170(04)74860-9)
- Sakalihasan, N., Michel, J.B., Katsargyris, A., Kuivaniemi, H., Defraigne, J.O., Nchimi, A., Powell, J.T., Yoshimura, K., Hultgren, R., 2018. Abdominal aortic aneurysms. *Nat. Rev. Dis. Prim.* <https://doi.org/10.1038/s41572-018-0030-7>
- TD, D., J. S., 2019. Hounsfield Unit. Definitions.
- Truijers, M., Pol, J.A., SchultzeKool, L.J., van Sterkenburg, S.M., Fillinger, M.F., Blankensteijn, J.D., 2007. Wall Stress Analysis in Small Asymptomatic, Symptomatic and Ruptured Abdominal Aortic Aneurysms. *Eur. J. Vasc. Endovasc. Surg.* 33, 401–407. <https://doi.org/10.1016/j.ejvs.2006.10.009>
- Vilalta, G., Nieto, F., Lipsa, L., Vilalta Alonso, J., Vaquero, C., Pérez, M., 2012. PREDICCIÓN DEL RIESGO DE RUPTURA DE ANEURISMAS DE AORTA ABDOMINAL. MÉTODO BASADO EN LOS BIODETERMINANTES GEOMÉTRICOS. *DYNA Ing. E Ind.* 87, 66–73. <https://doi.org/10.6036/4145>
- YI, C., KR, K., 1991. Effects of the non-Newtonian viscosity of blood on flows in a diseased arterial vessel. Part 1: Steady flows. *Biorheology* 28, 241–262. <https://doi.org/10.3233/BIR-1991-283-415>

SleepBCI: a platform for memory enhancement during sleep based on automatic scoring

M. Esparza-Iaizzo^{1,2}, I. Álvarez³, J. G. Klinzing², L. Montesano², J. Minguez² and E. López-Larraz²

¹ Universitat Pompeu Fabra, Barcelona, Spain, martineladio.esparza01@estudiant.upf.edu

² Bitbrain, Zaragoza, Spain, {jens.klinzing, luis.montesano, javier.minguez, eduardo.lopez}@bitbrain.es

³ Parc de Salut Mar, Multidisciplinar Sleep Unit, Barcelona, Spain, ialvarez@psmar.cat

Summary

Sleep is a complex physiological process responsible for several developmental, restorative and cognitive processes. Sleep-related disorders diminish life quality, and their diagnosis relies on classifying 8-hour polysomnographic (PSG) recordings. Manual scoring is the gold standard for analysing sleep and is a time-consuming task that can only be performed by trained experts. We propose SleepBCI, a platform to enable closed-loop memory enhancement during sleep using automatic scoring as its cornerstone. In order to achieve this, a convolutional neural network is validated offline on single channel information from 4 different datasets with expert-level accuracy. Given these results, we implemented a brain-computer interface based on monitoring EEG activity in real time, intended to enhance previously learned memories during deep sleep. We validated it in a controlled study on healthy participants, who performed a word-sound association task before and after sleep. Recall performance after waking was significantly greater for reactivated cues compared with non-reactivated cues. The platform and results presented in this work serve as a proof-of-concept for future sleep technology. It lays the foundation for a take-home device capable of alleviating the workload of physicians, preventing at-risk populations (elderly or obese patients) from unnecessary PSG tests and finally, delivering therapy to slow cognitive impairment or dementia.

1. Introduction

Roughly one third of our lifetime is spent sleeping. Considering this, proper sleep is crucial to perform daily functions and acts as the backbone of high-level brain functionality such as cognition and emotion [1]. Given the large amount of time we spend asleep, there is a wide spectrum of pathological conditions that can alter our rest and negatively impact life quality [2]. The International Classification of Sleep Disorders recognizes over 80 sleep-specific pathologies affecting approximately 50 to 70 million people in the United States of America [3]. Furthermore, it is believed that 20% of individuals suffering from sleep disorders are misdiagnosed, which reveals the daily challenges faced in current clinical practice [4]. The diagnosis of all sleep-related pathology is bottlenecked by manual classification of 8-hour multi-channel sleep recordings. Regarded as the gold standard, this process is tedious, time consuming, and can only be performed by well-trained professionals with experience in the field [5]. This highlights the pressing need to develop efficient and reliable scoring systems capable of achieving expert level scores in a fully automated manner. Moreover, prior research suggests that real-time stimulation during

manually-detected deep sleep can become a potential intervention for target conditions such as Alzheimer's and Mild Cognitive Impairment [6]. Considering these current clinical needs and opportunities, we developed and validated SleepBCI, a Brain-Computer Interface (BCI) for automatic deep sleep stimulation that employs real-time automatic scoring.

The concept of sleep scoring originates from brain wave architecture and other physiological parameters experiencing large changes during the night. This allows classification of sleep into well-established stages [7]. The two main phases, aside from wakefulness, are non-rapid eye movement (NREM) and rapid eye movement (REM). Both perform cyclic alternations, where NREM dominates the first half of the night and REM dominates the second [8]. According to the American Association of Sleep Medicine (AASM), NREM sleep can be broken down into sub-stages (N1-N3) depending on the depth of sleep [9]. To do so, experts record the subject's electrophysiological activity during the night and inspect extracts of 30 seconds, known as epochs. The successive representation of labelled epochs is known as a hypnogram (see Fig. 1).

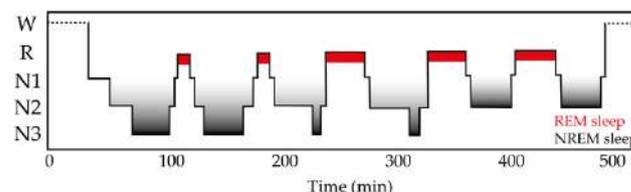


Figure 1. Standard hypnogram

The first attempts to perform automatic sleep scoring with results close to that of a human scorer date back to the 70s. Human-level scoring is taken from inter-rater reliability, considered at or above 80%. This result was obtained in a study by several professionals scoring the same set of PSG recordings and calculating the consistency across experts [10]. Recent literature has shown promising results through the use of different deep learning (DL) techniques such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs) [11][12]. Nevertheless, these studies utilize end-to-end approaches and multiple channels. Real-time single-channel scoring would enable the automatization of stimulation protocols during deep sleep. Eventually, the clinical acceptability of these algorithms could lead to the development of comfortable devices with significant benefits on dementias and similar diseases.

2. Methods

Firstly, we established a deep CNN for real-time automatic sleep staging on four different datasets. Then, we performed a closed-loop auditory memory reactivation study by employing the network's real-time feature. Through the automatic detection of deep stages of sleep using single-channel information, we delivered acoustic cues associated to previously learned declarative memories.

2.1. Offline staging of automatic staging

The polysomnographic data that served as training and testing for our automatic sleep staging pipeline was obtained from four open access sleep datasets. They consisted of SleepEDF¹ (healthy subjects, $n = 77$), ISRUC² (patients, $n = 100$), and DREEM³ (healthy subjects, DOD-H, $n = 25$; obstructive sleep apnea patients, DOD-O, $n = 50$).

Regarding the CNN's architecture, we expanded the framework proposed by Yildirim et al. [12] to achieve real-time sleep staging of 30 second epochs (100 Hz) of EEG data. In detail, the network's structure contains 2 submodels (see Fig. 2) of repeated sequences of 1D convolutions, max pooling, and dropout layers. First, we carried out feature extraction of every epoch in a temporally independent manner (Fig. 2A). Afterwards, we rearranged all epochs simultaneously into 5-unit vectors that correspond to the probability of selecting each class (Fig. 2B). Finally, the maximum value of this vector corresponds to the sleep label.

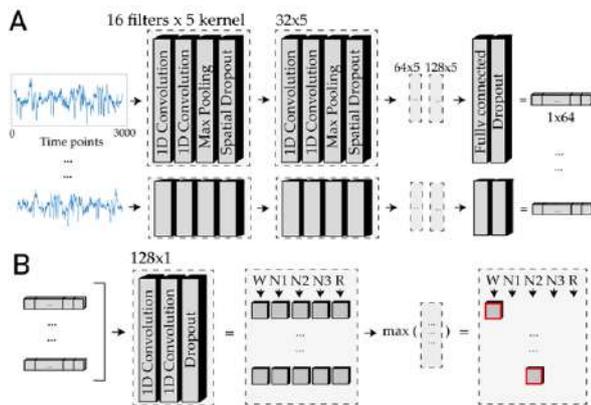


Figure 2. CNN architecture.

Raw PSG data from single-channel frontal electrodes was rearranged into 3D matrices (time points, epochs and channels) and then split into train + validation and testing with 80% and 20% ratios, respectively. The number of channels was limited to 1 and taken from frontal positions. The training set was partitioned once again (training and validation) with the same ratios as aforementioned. Before passing the data onto the model it was z-scored according to the training set's mean and standard deviation. The input size is a sliding window of 5 epochs (2.5 minutes)

which updates every 30 seconds for new epochs ensuring real-time labelling of the PSG data. Accuracy, precision, recall and f1-score were selected as metrics to evaluate the architecture.

2.2. Closed-loop memory reactivation study

Real-time automatic staging offers little advantage over end-to-end approaches in terms of PSG scoring performance. It does, however, show potential to conduct automated closed-loop memory enhancement protocols.

Ten volunteers (six males, four females, mean \pm SEM age: 22.9 ± 0.567 years) were recruited for this study with no prior history of sleep-related disorders or cognitive impairment. They arrived at the Centre Fòrum - Parc de Salut Mar at 22.30 pm. After obtaining informed consent and PSG setup, they performed a memory task (word-sound associations) learning and recall phases. This task, which was adapted from Fuentemilla et al. [13], consisted of presenting words and recognizable sounds (e.g., ambulance siren) with no intricate relationship during a learning phase. Afterwards, the recall phase presented subjects with the sound and asked them to recall the associated word. Once the night began and when reaching N3 or N2, half of the auditory cues that were memorized prior to sleep were played to the dormant subjects (see Fig. 3). Participants slept for 7 hours and were woken at 6.45 of the next morning to undergo the recall phase of the memory task once again.

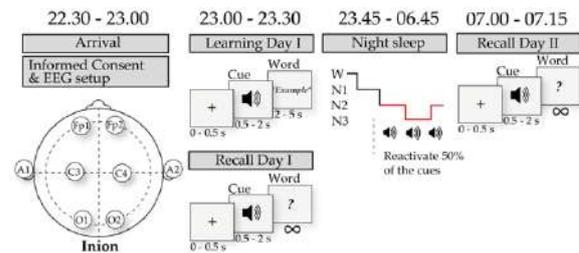


Figure 3. Experimental pipeline

EEG and EOG information were obtained continuously with a Xltek® Brain Monitor EEG Amplifier and processed with Natus SleepWorks (1998-2008 Xltek by Natus. Oakville, Canada). A minimalistic setup of 6 EEG channels (Fp1/M2, Fp2/M1, C3/M2, C4/M1, O1/M2 and O2/M1) and 2 EOG channels was employed, and this study was approved by the Ethics Committee of Parc de Salut Mar (CEIm) reference number 2021/9597.

The detailed schematic of the real-time functioning of SleepBCI is shown in Fig. 4. EEG activity is acquired from the Fp2-M1 channel, exported every 30 seconds (for every new epoch), and automatically preprocessed as aforementioned. Labels are produced using the CNN, and the weights employed for online staging were obtained from a database of 20 PSGs recorded by the hospital personnel. Depending of the outputted labels, either white noise (W, N1, or R) or white noise + auditory cues is

¹ <https://www.physionet.org/content/sleep-edfx/1.0.0/>

² <https://sleeptight.isr.uc.pt>

³ <https://dreem-dod-h.s3.eu-west-3.amazonaws.com/index.html>

delivered (N2 or N3). During deep sleep, the neuronal ensembles that codified word-sound associations are reactivated and long-term consolidation takes place.

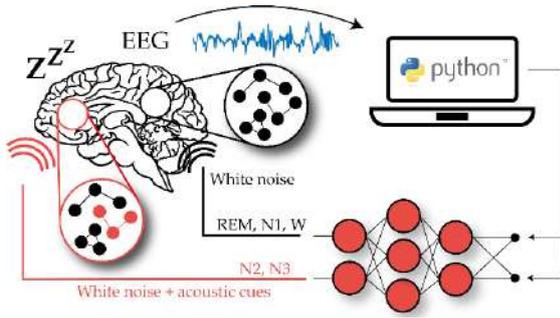


Figure 4. Schematic of SleepBCI function.

3. Results

3.1. Offline validation of automatic staging

Performance of the model is analysed through the aforementioned metrics on SleepEDF, ISRUC, DOD-H, DOD-O and visualised in Table 1. The only datasets that achieve an accuracy over inter-rater reliability (0.80) are SleepEDF and DOD-H.

	Accuracy	Precision	Recall	F1-score
SleepEDF	0.837	0.851	0.837	0.837
ISRUC	0.714	0.705	0.705	0.705
DOD-H	0.824	0.832	0.824	0.824
DOD-O	0.719	0.734	0.721	0.721

Table 1. CNN performance on each dataset.

The accuracy distribution amongst classes can be seen in Fig. 5A with trends across datasets. It is clearly visible that the main confusion is in N1 and more discreetly in REM phase. Combination of N2 and N3 shows very good values in terms of specificity, where true labels that are not N2 or N3 are rarely predicted as so. Fig. 5B represents the development of a single night (hypnogram), where once again the mentioned confusion (REM/N2) and (W/N1) are observed.

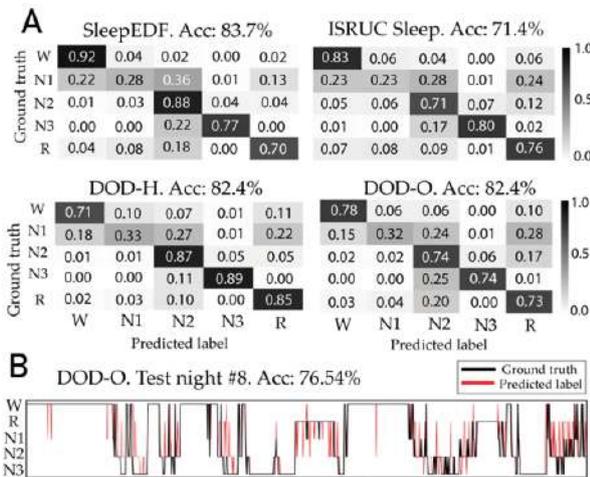


Figure 5. Accuracy distribution of each dataset (A) and sample hypnogram (B).

3.2. Closed-loop memory reactivation study.

Making use of SleepBCI’s capability of real-time stimulation during deep sleep, we aimed to prove that memory enhancement is possible. Day 2 recall of word-sound associations is therefore the most relevant parameter. Group behaviour shows a significant increase in this metric (see Fig. 6A, one-tailed permutation test, $p < 0.005$) for the words associated to reactivated cues (R+, in black) over those associated to non-reactivated cues (R-, in gray). Bar charts show average values and error bars represent standard error of the mean. Individual differences between reactivated and non-reactivated cues can also be seen Fig. 5B. Another aspect to be analysed was the relationship between stimulation and individual recall performance. The latter was computed through the difference of reactivated and non-reactivated cues (R+ - R-). As can be observed in Fig. 5D, there is no significant correlation ($r < 0.295$, $p > 0.05$) between the number of acoustic stimulations and recall performance, although it shows a positive slope.

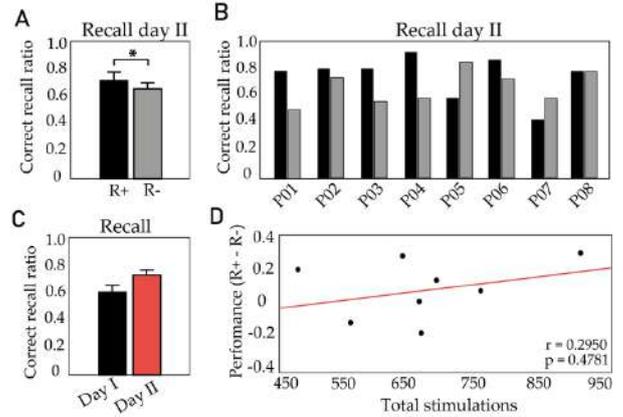


Figure 6. Recall comparison between reactivated and non-reactivated cues at group (A), and individual level (B) and between days (C). Panel D represents the correlation between stimulations and individual recall performance.

4. Discussion

4.1. Offline validation of automatic staging

There are major differences in performance throughout the four datasets that were employed in this study. At first glance, only SleepEDF and DOD-H achieve a score in performance over 80%. As previously mentioned, this accuracy value is extremely relevant since it refers to the reliability between human scorers and is considered the gold standard. These datasets (SEDF and DOD-H) are the only ones formed in their entirety by a large enough cohort of healthy subjects which is a requisite to train DL algorithms. In direct opposition, abnormal patterns of brain activity during the night that may stem from sleep-related pathology could hinder training and lead to a decrease in classification performance. ISRUC Sleep and DOD-O, which are formed by patients with sleep disorders, are perfect examples of this.

Another aspect to be discussed are the commonalities in confusion matrices between datasets. Agreement between predicted and expert labels for the N1 stage is significantly decreased in each dataset, similar to several deep learning

studies [11][12]. This happens mainly because of the low number of events available for training considering that N1 only makes up 5% of the night. Furthermore, N1 is regarded as a transition phase between Wake and N2, and it is normally labelled keeping in mind the information of future epochs and EOG information [9]. Since we perform real-time staging with no future context or EOG channels, we observe a pronounced fall in accuracy during the N1 stage.

4.2. Closed-loop memory reactivation study.

The results shown in Fig. 5 highlight the fact that SleepBCI is capable of performing an enhancement of sleep in a completely automated manner. We replicated the experimental protocol of Fuentemilla et al. [13] employing automatic staging rather than manual scoring in a cohort of 8 subjects. In detail, the hippocampus is known to perform a binding action on different components of an episode [8]. In our case, these components were acoustic cues and their associated words. During a standard night of sleep, the hippocampal neuronal ensembles responsible for encoding these recently acquired memories are activated once again. This triggers the consolidation process of memories by their integration into neocortical areas allowing long-term retrieval. By delivering the acoustic cues of the associated word-sound pairs during the night, we are favouring this reactivation even further and generating a difference in the consolidation process between cues (reactivated vs. non-reactivated) [8]. Nevertheless, the limited number of participants may not allow strong correlations to emerge between performance and stimulations.

Being able to modulate sleep automatically and achieve significant enhancement of memory overnight has strong clinical implications related mainly to mild cognitive impairment and dementia. Several studies have reported an increase in memory by using different stimulation protocols [14][15]. Furthermore, recent literature has put forward the relationship between deep sleep EEG hallmarks and accumulation of tau and amyloid- β , proteins responsible for neural damage. This reveals the potential of deep sleep stimulation as an interventional target of neurodegeneration [6]. Furthermore, if we are able of doing this through single-channel EEG data (as in SleepBCI), we could potentially create a take-home device capable of delivering treatment. Even though this is a promising research line, it is still unclear whether accumulations of these proteins are directly responsible for the entirety of Alzheimer's pathology. Before claiming that stimulation during deep sleep can open a back door for Alzheimer's disease, these questions still need to be answered.

The work we present here has successfully proved that closed-loop sleep stimulation through automatic classification is possible. Physicians overload and cognitive decline due to deep sleep alterations can be eventually tackled by the work carried out in this work, tethering it to real-world biomedical needs and highlighting its inherent practical application.

Acknowledgments

We would like to thank the personnel at Centre Fórum - Hospital del Mar for giving their support and allowing us to make safe use of the clinical setting employed in this project. We would also like to extend a special acknowledgement to Dr. Jordi García-Ojalvo and Dra. Gemma Piella for their guidance and useful input to this work.

References

- [1] Walker MP. The role of sleep in cognition and emotion. *Annals of the New York Academy of Sciences*, vol. 1156, no 1, 2009, p. 168-197.
- [2] Chokroverty S. Overview of sleep & sleep disorders. *Indian J Med Res*, vol. 131, no 2, 2010, p. 126-140.
- [3] Colten HR. Extent and health consequences of chronic sleep loss and sleep disorders. *Sleep disorders and sleep deprivation: an unmet public health problem*, 2006, p. 55-135.
- [4] Ohayon MM. Prevalence and comorbidity of sleep disorders in general population. *La Revue du praticien*, 2007, vol. 57, no 14, p. 1521-1528.
- [5] Biswal S. Expert-level sleep scoring with deep neural networks. *Journal of the American Medical Informatics Association*, 2018, vol. 25, no 12, p. 1643-1650.
- [6] Lee YF. Slow wave sleep is a promising intervention target for Alzheimer's disease. *Frontiers in neuroscience*, 2020, vol. 14, p. 705.
- [7] Nir Y. Sleep and consciousness. En *Neuroimaging of consciousness*. Springer, Berlin, 2013. p. 133-182.
- [8] Klinzing JG, Niethard N, Born J. Mechanisms of systems memory consolidation during sleep. *Nature neuroscience*, 2019, vol. 22, no 10, p. 1598-1610.
- [9] Iber C. The AASM manual for the scoring of sleep and associated events: Rules. *Terminology and Technical Specification*, 2007.
- [10] Danker-Hopfe H. Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard. *Journal of sleep research*, 2009, vol. 18, p. 74-84.
- [11] Phan, Huy, et al. SeqSleepNet: end-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2019, vol. 27, no 3, p. 400-410.
- [12] Yildirim O, Baloglu UB, Acharaya UR. A deep learning model for automated sleep stages classification using PSG signals. *International journal of environmental research and public health*, 2019, vol. 16, no 4, p. 599.
- [13] Fuentemilla L. Hippocampus-dependent strengthening of targeted memories via reactivation during sleep in humans. *Current Biology*, 2013, vol. 23, p. 1769-1775.
- [14] Ladenbauer J. Promoting sleep oscillations and their functional coupling by transcranial stimulation enhances memory consolidation in mild cognitive impairment. *Journal of Neuroscience*, 2017, vol. 37, no 30, p. 7111-7124.
- [15] Ngo, Hong-Viet V. Auditory closed-loop stimulation of the sleep slow oscillation enhances memory. *Neuron*, 2013, vol. 78, no 3, p. 545-553.

Telerehabilitation of post-stroke motor impairment

A. García¹, A. Iriarte¹, C. Margelí¹, B. Mayans¹, C. Molas¹, J. Torner¹, G. Serrancolí¹

¹ Simulation and Movement Analysis Lab, Multimedia Applications Lab, Universitat Politècnica de Catalunya, Barcelona
quedaren@gmail.com, gil.serrancoli@upc.edu

Abstract

The single most successful predictor for improvement and retention of motor outcomes post-stroke is the number of hours devoted to rehabilitation. In this work, the use of information technologies to provide rehabilitation therapy to individuals in their homes (telerehabilitation) is proposed as a means for making rehabilitation more accessible and encourage chronic stroke survivors to continue training their mobility. A computer application was developed with the ultimate objective of implementing said proposal. Then, a feasibility trial was conducted to test the applicability of this approach. The trial was successful in highlighting the potential of the framework, as well as aspects where it was lacking or could be improved. In the end, this study allowed developers to trace routes for the future of the project.

1. Motivation

Stroke is a cerebrovascular incident, wherein sudden interruption of blood flow to parts of the brain can result in devastating consequences. In Spain, stroke stands as the second highest cause of death [1]. However, even in those who survive, damage to nerve tissue can result in significant sensorimotor and cognitive complications, so much so that the condition ranks as the single highest cause of acquired disability and the second highest cause of dementia [1]. Nevertheless, the problem of stroke does not end there. In 2019, the SEN (Sociedad Española de Neurología) showed that stroke prevalence also carries concerning social and economic connotations [2].

On the social side, their analysis highlights a well-known correlation between stroke and lower levels of perceived health, quality of life, and mental health, even when compared to other sufferers of chronic conditions. This level of negative self-perception leads to an increased risk for social isolation. This is because lack of confidence in mobility and poor emotional state often hinder the individual's motivation to be active and engage with their community.

On the economic side, they identified a stroke-associated economic burden on public spending even after removing the costs linked to acute care. Stroke survivors display a great likelihood of using urgency and hospitalization services, even when compared to subjects of other chronic conditions. Stroke also decreases the working population, resulting in implicit losses in production due to the decrease in total work hours. This is because individuals left with significant impairment may become too dependent to be productive on their own. Of course, this also translates into an economic burden to the individuals themselves, as even those with only mild impairment may face difficulties finding employment.

It should be clear, then, that stroke-provoked disability presents challenges not only to those who suffer from it, but also to society as a whole. The most obvious way to attack these challenges is to facilitate the rehabilitation of patients. The public health system already includes some amount of rehabilitation training as part of the treatment for cerebral infarction. However, it only covers the first six months after the lesion. This is, generally, not enough time to warrant a full recovery [3-4]. It should also be noted that, although recovery might not be significant after the acute phase, keeping a rehabilitation treatment is crucial to avoid further physical and cognitive deterioration. While many patients will continue to pursue rehabilitation in outpatient centers, access to these venues may be limited by waiting lists, transportation issues or costs [5].

This problem with availability is a clear indication that the demand for rehabilitation well exceeds what the health sector can currently provide and this should come as no surprise. According to the 2012 Iberictus study, Spain sees somewhere between 110,000 and 120,000 new stroke cases every year [2]. Additionally, the prevalence of stroke has been growing due to lower mortality and aging populations. Projections predict that this trend will continue [6-7]. Thus, the number of people living with post-stroke residual impairment will only grow, making the problem even worse.

In this work, a natural solution is explored. If the healthcare system cannot provide enough rehabilitation hours, patients should supplement by training on their own. What we propose is the creation of computer applications that can support and encourage the rehabilitation efforts of subjects in their own homes - an endeavor known as telerehabilitation.

2. Introduction

2.1. Telerehabilitation

Telerehabilitation is not too dissimilar in principle to the modalities of remote-work or remote-learning that we have all become too familiar with due to the COVID19 pandemic. Perhaps one of the things that this global crisis has taught us is the power of using virtual platforms to supplement activities in many sectors.

Information technologies can be used to facilitate patient rehabilitation at home in two key ways. First, Virtual Reality (VR) can be used to enhance rehabilitation. Second, patient performance data can be collected and stored so physiotherapists can engage in asynchronous monitoring of their training.

With respect to the first, several motor learning mechanisms have been cited in the literature as potential theoretical advantages of VR approaches over conventional rehabilitation therapy. For instance, real time feedback, or contextualizing training tasks in environments meaningful to the patient [8]. Given this, one might expect that VR rehabilitation should perform better than conventional methods. However, clinical trials do not seem to point to such a strong conclusion. Instead, the results show modest success. That is to say, virtual-environment rehabilitation therapy is at least as effective as conventional methods [9-10].

This is, however, all that is needed to justify an interest in telerehabilitation. After all, the single best predictor for improvement in rehabilitation outcomes is the amount of time dedicated to training. As long as this approach is not worse than conventional rehabilitation, patients training more will always be a net gain. Easier access to rehabilitation (right in their homes with plug-and-play devices) and the implementation of gamification ideas (to keep the exercises engaging) promise to significantly improve adherence to training. This is particularly important for individuals who are otherwise not doing any kind of training, perhaps due to a lack of motivation or inability to access rehabilitation centers.

As for the second means of facilitation - keeping physiotherapists involved, even when the training is done at home - it promises to keep the endeavor safe and appropriate for each patient. One might, however, imagine that this will subject the approach to the same kind of availability problem as conventional therapy due to a shortage of personnel. This is not the case. After all, asynchronous supervision requires much less involvement. Hypothetically, a transition towards telerehabilitation would free up physiotherapists to treat more patients, spending less time on each one, only calling them in for a session when something seems to be going wrong or for a periodic check-up. While this approach would, perhaps, not solve the availability problem, it does promise to significantly reduce it.

2.2. Objectives

Recognizing all the considerations so far presented, the LAM (Laboratori d'Aplicacions Multimèdia), a lab within the Universitat Politècnica de Catalunya (UPC), and ADFO (Associació Diversitat Funcional d'Osona), a non-profit devoted to support people with motor disabilities, came together under one goal - to create a robust framework to facilitate the telerehabilitation of individuals living with post-stroke motor impairment.

The objective of this project is a little more concrete. A non-immersive VR application, henceforth referred to as the Telerehabilitation System, has been developed to act as this framework. Early tests with the Telerehabilitation System have resulted in very positive feedback from participating volunteers and physiotherapists. In light of these findings, the project has reached a new stage - less exploratory and now concerned with refining the proposed framework and assessing how robust it can actually be in practice. To this end, a feasibility trial is being carried out with the goals of

putting the system to the test in a real use-case scenario and gathering insights to guide future lines of development.

3. Telerehabilitation System

To ensure maximum affordability and accessibility, the Telerehabilitation System uses a minimalist set-up. A low-cost depth camera (Intel RealSense D415, Intel, Santa Clara, CA, USA) functions as the main means of motion tracking. Kinematics of the patient's motion are inferred from depth information using algorithms from NuiTrack's SDK [11] - resulting in real time skeleton tracking without markers. The patient's movements are then used to animate an avatar, through which they can perform various tasks in a virtual environment. The application is easy to install and can run on just about any personal computer.

The system has been designed with two target users in mind - injured subjects, who will use the application to support their rehabilitation training, and physiotherapists, who will interact with the application to oversee the progress of their patients and adjust the parameters of their training accordingly.

The patient side consists of a series of gamified activities which can be further split up into exercises and games. Exercises are simple virtual environments where the patient is prompted to execute a particular exercise guided by a character (See Figure 1) - who is constantly showing the appropriate motion. Games attempt to emulate daily activities which prompt the patient to execute certain target motions in a meaningful context - for instance movement of the arm interpreted as the cleaning of a mirror (See Figure 2).



Figure 1. Vertical abduction exercise



Figure 2. Example of a game consisting of cleaning a mirror.

The current version of the Telerehabilitation System features 6 exercises and 7 games. For a list of these, see Figures 3 and 4 which show the selection panels associated with each activity type. ADFO provided the expertise needed to ensure all games and exercises would be relevant to the patients' training and that the applications would be accessible to patients with different degrees of impairment. In the exercises, a virtual avatar guides the movement of a particular human joint and the user has to follow her (shoulder flexion, shoulder vertical abduction, shoulder horizontal abduction, elbow flexion and medio-lateral translation of the pelvis). The games are designed so that the user performs movements typical from daily life activities (to keep the balance avoiding objects, flex the shoulder to reach certain goals, cook, or clean a bathroom).

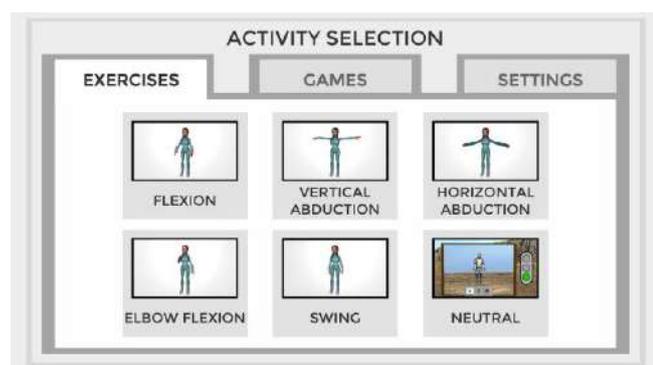


Figure 3. Exercise selection panel.



Figure 4. Game selection panel.

The application allows for some level of asynchronous monitoring by means of the database functionality. Data is collected automatically as patient's use the application. The data exported include biomechanical data (joint angles and joint positions), performance in games, and time played. Physiotherapists can log in with special credentials to see the various exercises their patients have been doing and how well they have been performing on them.

4. Methodology

A small sample ($n = 16$) of individuals who had suffered stroke (with more than six months after the injury) was recruited for the pilot study. Only those who met the three following criteria were accepted as participants: Participants must be able to keep themselves on both feet without help, they must have acceptable visual acuity and they must be younger than 75 years-old.

Of the initial sample, only ten people took part in the actual trial (five men and five women, age: 52.9 ± 8.0 years old). ADFO personnel (occupational therapists and physiotherapists) were the ones responsible for communicating with potential candidates and procuring their informed consent.

Participants were randomly split into two groups and both groups underwent a training program in two phases (crossover study). During the first phase of the program, group 1 used the application to follow their exercise plan, while group 2 functioned as a sort of "control" group, performing exercises without the application. For the second phase of the program, groups 1 and two exchanged roles. Each phase consisted of 8 weeks of daily training. Six participants out of ten completed the study.

Patients were evaluated in-person by ADFO personnel four times - before the beginning of the trial, after the end of the first phase, before the beginning of the second phase and at the end of the second phase. A "washing-out period" of two weeks was introduced between the first and second phases as a means of differentiating both treatment periods. The application was used to collect kinematic data of subjects during these evaluations, regardless of group. Additionally, functional tests were performed and questionnaires (FIM - Functional Independence Measure, Berg scale - evaluation of balance, SF-36 - health survey) were filled to collect further quantitative and qualitative information.

The application also collected a variety of data during the training itself, both in the form of kinematic information and performance metrics of the games (such as score or accuracy). Both these and the data from the evaluation are in the process of being analyzed.

5. Results

Figure 5 shows an example of data extracted from questionnaires during the in-person evaluations. Here, changes in FIM suggests higher improvement or lower deterioration for four out of six subjects when comparing the period with the application to the period without. Of the two remaining, Subject 4 demonstrates very small changes in FIM increment, whereas Subject 6 exhibited a more significant discrepancy to the main trend. Subject 6, however, received medication just before the evaluation after the period using the application. The medication interfering with the results could explain this subject's large drop in this phase.

Figure 6 shows an example of data collected throughout the trial. The graph displays the maximum medio-lateral translation values of the pelvis for subject 1. These are plotted at each evaluation and during treatment with the application. An upwards trend in maximal range of motion manifests as the trial progresses.

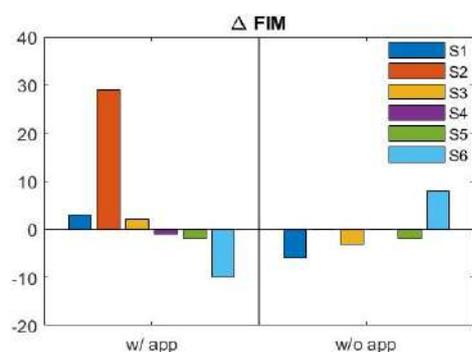


Figure 5. Increment of FIM (Functional Independence Measure) after the period using the application (w/ app) and after the period without using the app (w/o app). SX in the legend indicates the subject number.

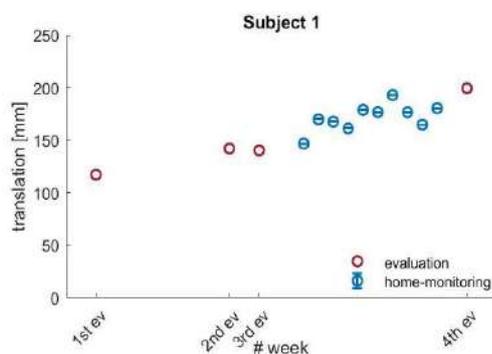


Figure 6. Maximum medio-lateral translation of the pelvis during the four evaluations and the mean value of each week of the treatment.

6. Discussion

The example graphs shown in the prior section are in line with the hypothesis that the Telerehabilitation System can facilitate rehabilitation. Future work in the short term will be oriented towards comparing and summarizing all recorded data.

Feedback from both patients and physiotherapists during this study proved invaluable in discerning new ways to improve the application. Challenges such as patients lacking internet allowed for the framework to become more accessible to accommodate them. Similarly, in answering the needs of the physiotherapists, the Telerehabilitation System seeks to become a platform for professionals to plan out rehabilitation programs easily.

The trial had relative optimal adherence, with six patients finishing the testing period all the way through. For this reason, a second stage of development is planned to expand the gamification aspects of the application.

To make the process of assessing patient performance simpler, there are also plans to allow for plots much like Figure 6 to be directly generated by the application. The one presented here was constructed using MATLAB.

Thus, the feasibility trial has served its purpose, giving deeper insights as to the strengths and weaknesses of our

approach, and painting a picture of where the project should go moving forward.

Acknowledgements

The authors acknowledge the feedback received from Ariadna Pamplona and Elisenda Currius from Associació Diversitat Funcional d'Osona (Osona) and the support received from this institution. All authors also acknowledge the funding received from: DKV Impacta 2020 (DKV Seguros), Tecnologías Accesibles 2020 (Indra and Fundación Universia), and Grants for Cooperation Projects 2020 and 2021 (UPC's Center for Development Cooperation). G. Serrancolí acknowledges the support from Serra Hünter Program.

Bibliography

- [1] A. Menéndez Pérez, "Día mundial del ictus," pp. 1–4, 2017.
- [2] "El Atlas del ictus," 2019. https://www.sen.es/images/2020/atlas/Informes_comunidad/Informe ICTUS_Cataluna.pdf. (accessed 01-10-2021).
- [3] T. H. Murphy and D. Corbett, "Plasticity during stroke recovery: from synapse to behaviour," *Nat. Rev. Neurosci.*, vol. 10, no. 12, pp. 861–872, Dec. 2009, doi: 10.1038/nrn2735.
- [4] A.-S. Wahl and M. E. Schwab, "Finding an optimal rehabilitation paradigm after stroke: enhancing fiber growth and training of the brain at the right moment," *Front. Hum. Neurosci.*, vol. 7, 2014, doi: 10.3389/fnhum.2013.00911.
- [5] L. Sheehy *et al.*, "Home-based virtual reality training after discharge from hospital-based stroke rehabilitation: a parallel randomized feasibility trial," *Trials*, vol. 20, no. 1, p. 333, Dec. 2019, doi: 10.1186/s13063-019-3438-9.
- [6] F. A. Batchelor, S. F. Mackintosh, C. M. Said, and K. D. Hill, "Falls after Stroke," *Int. J. Stroke*, vol. 7, no. 6, pp. 482–490, Aug. 2012, doi: 10.1111/j.1747-4949.2012.00796.x.
- [7] J. Díaz-Guzmán, J.-A. Egido, R. Gabriel-Sánchez, G. Barberá-Comes, B. Fuentes-Gimeno, and C. Fernández-Pérez, "Stroke and Transient Ischemic Attack Incidence Rate in Spain: The IBERICTUS Study," *Cerebrovasc. Dis.*, vol. 34, no. 4, pp. 272–281, 2012, doi: 10.1159/000342652.
- [8] L. Piron *et al.*, "Exercises for paretic upper limb after stroke: A combined virtual-reality and telemedicine approach," *J. Rehabil. Med.*, vol. 41, no. 12, pp. 1016–102, 2009, doi: 10.2340/16501977-0459.
- [9] M. Yates, A. Kelemen, and C. Sik Lanyi, "Virtual reality gaming in the rehabilitation of the upper extremities post-stroke," *Brain Inj.*, vol. 30, no. 7, pp. 855–863, Jun. 2016, doi: 10.3109/02699052.2016.1144146.
- [10] J. Chen, W. Jin, X.-X. Zhang, W. Xu, X.-N. Liu, and C.-C. Ren, "Telerehabilitation Approaches for Stroke Patients: Systematic Review and Meta-analysis of Randomized Controlled Trials," *J. Stroke Cerebrovasc. Dis.*, vol. 24, no. 12, pp. 2660–2668, Dec. 2015, doi: 10.1016/j.jstrokecerebrovasdis.2015.09.014.
- [11] "What is NuiTrack™SDK?," [Online]. Available: http://download.3divi.com/Nuitrack/doc/Overview_page.htm. (accessed 01-10-2021)

Premios José María Ferrero Corral

Jueves 25 de Noviembre de 2021, 15:00

Prediction of short-term survival in glioblastoma patients with total tumor resection from multimodal pre-surgical MRI images using machine learning

R Navarro González¹, S. Cepeda Chafra², C. Alberola-López, Rodrigo de Luis García, Santiago Aja-Fernández

¹ Laboratorio de Procesado de Imagen (LPI), Universidad de Valladolid, Valladolid, España, rafael.navarro@uva.es

² Servicio de Neurocirugía, Hospital Universitario Río Hortega, Valladolid, España

Summary

Glioblastoma is a gruesome disease with poor prognosis. Its treatment has not changed significantly in the last thirty years. We present a machine learning model fed with radiomics features, obtained from multimodal pre-surgical MRI images, which is capable of classifying glioblastoma patients in two groups, according to their predicted overall survival (over 6 or below 6 months). Our model obtained an accuracy of 81 % and an AUC of 0.86. Similar models might be used for stratifying patients depending on risk, and determining the best possible action for each patient, improving medical care and outcomes.

1. Introduction

Glioblastoma is the most frequent primary brain tumor in the adult population. Its treatment is based on a combination of radiotherapy, chemotherapy and surgical resection. The median survival from diagnosis of a glioblastoma patient is 14 months and most die due to tumor recurrence [1]. One of the main characteristics of glioblastoma is its genetic and morphological heterogeneity [2]. However, it has been shown that it is possible to obtain predictions about patient survival from their pre-surgical MRI images [3]. From the MRI images, a massive amount of quantitative features of the tumor are extracted. From this set, a subset is drawn out through feature selection algorithms. Selected features are then fed to machine learning models. This procedure is known as radiomics, and has been shown to be superior to expert capability in many tasks [4]. Accurate prediction of survival would allow stratification of patients into different risk groups, which would lead to better patient management by offering possible treatments or directing them to palliative care, ultimately improving medical care.

There is a large number of publications that apply this methodology to glioblastoma for survival prediction. During the last years, at the MICCAI conference, together with the BraTS challenge [5], a competition has been held in which different methods were presented for classifying glioblastoma patients into 3 different groups depending on survival, (<10 months, >10 months < 15 months, >15 months) and other contributions have tried to predict survival with other cut-off points as in [6, 7], where 36 months and 12 months were chosen as thresholds. However, few publications have attempted to perform short-term prediction. Short-

term prediction could be a marker that determines the surgical decision by offering a data-driven value to the practitioner. Other publications include diffusion or perfusion MRI data [7]. However, we only include structural MRI which could simplify the translation to the clinic.

Therefore, in this paper we present a fully automatic method that allows the classification of glioblastoma patients into two groups, survival of less than 6 months and survival of more than 6 months, based on post treatment structural MRI images.

2. Methods

2.1. Data

Our dataset is composed of 207 with a large majority belonging to the public BraTS 2020 competition database (n=119). 39 cases come from three different sources grouped in The Cancer Imaging Archive (TCIA) (Ivy-Gap n=24, TCGA n=3, and CPTAC-GBM n=14). Finally, 21 and 28 cases belong to the datasets generated at Institution 1 and Institution 2, respectively during 2019. For each case, T1, T2, FLAIR and T1C (T1 after contrast) were acquired.

These 207 cases were separated into a training and a testing group in a 70/30 ratio, with 144 and 63 cases respectively. For training, a 10x repeated 5-folds cross validation strategy is followed. Once the most promising models are identified, they are trained with the 144 total cases and tested on the testing group.

Dataset	n	Age (SD)	OS (IQR)	Censored	OS<6M (%)
BraTS	119	62±12	448 (367)	0	21 %
TCIA	39	58±13	444 (522)	8 %	15 %
Institution 1	21	66±10	413 (304)	19 %	14 %
Institution 2	28	58±13	466 (217)	32 %	0 %
Whole dataset	207	61±12	412 (217)	7 %	17 %
Split 30/70					
Training	144	61±12	420 (347)	14 % (14)	17 %
Testing	63	61±12	401 (366)	3 % (2)	17 %

Table 1. Dataset details. Overall survival was defined as the time in days from diagnosis to death (censored = 0) or the last date that the patient was known to be alive (censored = 1).

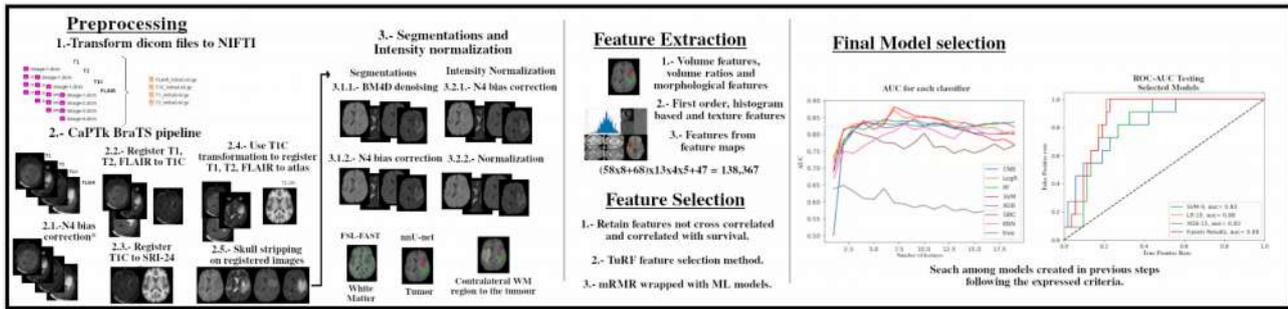


Figure 1. Flowchart showing the radiomics process. Preprocessing is performed. Features are extracted using MatLab and a combination of feature selection methods are used to obtain a subset of key features. Different classification methods are then tested to obtain a final model.

2.2. Preprocessing and tumour segmentation

All images were first subjected to BraTS preprocessing pipeline, implemented in CaPTk [8]. Two different preprocessing methods were applied to the output images:

1. The aim of the first preprocessing is to obtain a high quality white matter segmentation (FSL-FAST [9]) and tumor segmentation, required for the second preprocessing and feature extraction. BM4D denoising method [10] and N4 inhomogeneities correction were applied [11]. BM4D could reduce the predictive power of the texture features, but improve segmentations.
2. The second preprocessing aims at harmonizing the intensity between images coming from different scanners, acquisition methods, and institutions. Inhomogeneities are corrected with N4 and each image was scaled using the mean value of the white matter intensity of the contralateral area of the tumor.

Tumor segmentation of institutional and TCIA cases was obtained from the BraTS 2020 winning nnUnet [12] convolutional network. This network is able to achieve good segmentation results, and does not require any additional external input.

2.3. Feature extraction

A total of five regions of interest (ROIs) were defined: 3 of them are obtained by the tumor segmentation (Enhancing tumor ET, non-enhancing tumor NET, edema ED) and 2 more are constructed from them: tumor-core (TC), which is the sum of ET and NET, and whole-tumor (WT), which is the sum of the three initial regions. Moreover, twelve different filters are applied to the images. Wavelet filtering is applied, yielding 8 decompositions per level. these are all possible combinations of applying either a High or a Low pass filter in each of the three dimensions. Also, four Laplacian of Gaussian filters, are applied emphasizing

areas of gray level change, where sigma (sigma values: 2, 3, 4, 5) defines how coarse the emphasized texture should

be. These types of features have shown improvements in the final prediction result [13]. From these ROIs and images, a total of 138,367 features were extracted. These features are defined as follow:

- Volume features (6): Volume of the whole tumor (WT), the tumor core (TC), each region (ET, NET, ED) and the volume of the brain.
- Volume ratios (7): volume ratios; ratio of WT and the brain, ratio of ET and WT, ratio of NET and WT, ratio of ED and WT, ratio of ET and ED, and the ratio of NET and ED.
- Morphological features (34): the percentage of the tumor inside the cerebellum, brain stem, basal ganglia, and the parietal, occipital, frontal and temporal lobes. These measurements are obtained by segmenting the SRI-24 atlas with Freesurfer [14], and fusing the Freesurfer regions to create the aforementioned anatomical areas. The volumes are obtained by measuring the presence of the tumor in those regions. These measurements are also calculated for the left and right hemispheres. Furthermore, the centroid of the tumor is determined and measurements of compactness, sphericity, the ratio volume to surface of the tumor, the TC surface area, and another sphericity measurement defined in [15] are computed.
- First order, histogram based and texture features (68x4x5x13): 68 features are calculated over the whole ROIs, described in table 2.
- Features from feature maps (58x8x4x5x13): Feature maps are defined calculating each feature from table 2, on a block of 3x3x3 pixels in each ROI. The eight first-order features are calculated on these maps. For the features calculated with maps, the histogram bins are not obtained, but the frequency and probability of pixel intensities are calculated for a 5 bin histogram.

The total number of features can be calculated with a simple sum:

$$(58 \times 8 + 68) \times 4 \times 5 \times 13 + 47 = 138,367$$

Feature type	Feature	N°
First order	Kurtosis, skewness, variance, standard deviation, median, mean, min and max	8
Histogram based	each bin of a 20 bins histogram	20
NGTDM	Coarseness, Contrast, Busyness, Complexity, Strength	5
GLRLM	GLN, GLV, HGRE, LGRE, LRE, LRLGE, LRLGE, RLN, RLV, RP, SRE, SRHGE, SRLGE	13
GLCM	Energy, Contrast, Entropy, Homogeneity, Correlation, Variance, SumAverage, Auto Correlation, Dissimilarity	9
GLSZM	SZE, LZE, GLN, ZSN, ZP, LGZE, HGZE, SZLGE, SZHGE, LZLGE, LZHGE, GLV, ZLV	13
Total:		68

Table 2. General features

MatLab was used to program and extract these features. Particularly, for texture features, the package presented in [16] which is IBSI compliant [17], was used.

2.4. Feature selection

After extraction, a 3-step feature selection process was performed:

1. First, the Spearman correlation between each feature and survival in days is calculated, and features that present a statistically significant correlation (p -value < 0.05) are retained. In addition, the Spearman correlation matrix between the statistically significant features is calculated, and the correlation values are used to eliminate those that are correlated with each other (correlation factor > 0.66), keeping the feature with the lowest p -value. This reduce the volume of features from 138367 features to 12940 by removing features that are not statistically significant, and from 12940 to 1648 by removing those that are correlated with each other.

2. Second, the TuRF method was used [17]. The number of features is reduced to the top 100 with the best TuRF weights. In addition, 20 features highly reproducible as age, volume ratios between the different regions of the tumor, or morphological features such as sphericity, are reincorporated at this point.

3. An information measure, maximum relevance-minimum redundancy (mRMR), was applied to this subset. This metric generates a ranking of feature groups by iterating through the different features. For each iteration the subsets that mRMR generates, are evaluated over 8 different machine learning methods.

2.5. Machine learning models

A total of 8 different models present in the scikit-learn Python library and XGBoost, were evaluated. There is a significant class imbalance between long survival (>6 months) and short survival (< 6 months) cases with a case-to-case ratio close to 6:1. To account for this the weights of each model have been adapted. Moreover, instead of evaluating the accuracy of the models, their

AUC was analyzed. The models employed are: complement naive Bayes (CNB), logistic regression (LR), random forest (RF), support vector machine (SVM), extreme gradient boosting (XGB), gradient boosting classifier (GBC), K-nearest neighbors (KNN) and decision tree C4.5 (tree).

3. Results

To obtain the final results, the ranking generated by mRMR is iterated over. For each number of features, 50 subsets are built and evaluated with the 8 models presented. Training is performed in a 10x 5-fold cross validation scheme. The maximum AUC value obtained by the best feature set allows to make a better targeted search towards the optimal classifier and feature set. They can be consulted in the figure 1.

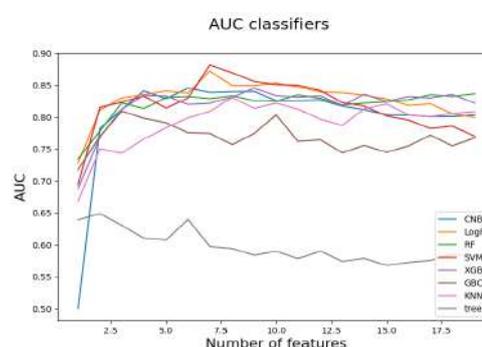


Figure 2. Results AUC for each model and n° of features

This search is performed over the models with best performance, starting from a number of features of 5. Although, a model with a high number of features tends to perform better usually simpler models generalize better and are less prone to overfitting. The final model is selected following the next criteria:

- Among the best presented models (CNB, LR, SVM, RF, XGB)
- Between 5 and 15 features
- AUC, precision and F1 over 80 %
- Balanced results among classes

Its testing results are summarized in the next table.

	Precision	Recall	F1-score	N° of cases
Logistic Regression – 10 features				
OS > 6 months	0.93	0.83	0.88	52
OS < 6 months	0.47	0.73	0.57	11
Weighted avg	0.85	0.81	0.82	63
Accuracy			0.81	63
AUC			0.88	63

Table 3. Results of the final model

The model has an accuracy of 81 % for the testing dataset, showing a good performance for the majority class and a fair performance for the minority class.

4. Discussion

A model able to classify glioblastoma patients according to their overall survival (over six months vs below six months) has been built. This model uses

structural MRI commonly acquired for brain tumors and is focused only in patients whom underwent total resection during surgery.

The difference in performance between the majority class (>6 months) and the minority class (< 6 months) might be a sign of overfitting. The low number of cases for < 6 months in the dataset makes this task challenging. A bigger dataset would be very helpful in order to confirm these results. Furthermore, preprocessing was identified as a critical step in the pipeline, since it harmonizes images coming from different machines, acquisition protocols and institutions. It is important to note that although the package used to calculate texture features are IBSI compliant [16, 17], meaning that their reproducibility has been tested, the local features extracted from feature maps in this work, are not. Although, these features has been demonstrated moderately reproducible across vendors and image settings in breast X-rays in [18] we consider important to determine the reproducibility of these features in MRI since its apparent that they are good at capturing the heterogeneity of the tumor. Moreover, model selection for binary classification is still a challenging problem. Even though, AUC can be used as a general metric for model performance, it can be problematic if the threshold in use is not calibrated. Other metrics, like accuracy, have been demonstrated inferior to AUC for model selection [19].

Short term prediction of survival in glioblastoma can become an important element of its management. Surgery is still the main therapeutic element, and patients and surgeons face a complicated decision which can be better supported after the validation of a similar classifier as the one presented in this work.

References

- [1] Batash, R., Asna, N., Schaffer, P., Francis, N., & Schaffer, M. (2017). Glioblastoma multiforme, diagnosis and treatment; recent literature review. *Current medicinal chemistry*, 24(27), 3002-3009.
- [2] Tirosh, I., & Suvà, M. L. (2020). Tackling the many facets of glioblastoma heterogeneity. *Cell Stem Cell*, 26(3), 303-304.
- [3] Bakas, S., Shukla, G., Akbari, H., Erus, G., Sotiras, A., Rathore, S., ... & Davatzikos, C. (2020). Overall survival prediction in glioblastoma patients using structural magnetic resonance imaging (MRI): advanced radiomic features may compensate for lack of advanced MRI modalities. *Journal of Medical Imaging*, 7(3), 031505
- [4] Kumar, V., Gu, Y., Basu, S., Berglund, A., Eschrich, S. A., Schabath, M. B., ... & Gillies, R. J. (2012). Radiomics: the process and the challenges. *Magnetic resonance imaging*, 30(9), 1234-1248.
- [5] Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., ... & Eaton-Rosen, Z. (2018). Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. *arXiv preprint arXiv:1811.02629*.
- [6] Fu, X., Chen, C., & Li, D. (2021). Survival prediction of patients suffering from glioblastoma based on two-branch DenseNet using multi-channel features. *International Journal of Computer Assisted Radiology and Surgery*, 16(2), 207-217.
- [7] Ai, Z., Li, L., Fu, R., Lu, J. M., He, J. D., & Li, S. (2017). Integrated Cox's model for predicting survival time of glioblastoma multiforme. *Tumor Biology*, 39(4), 1010428317694574.
- [8] Rathore, S., Bakas, S., Pati, S., Akbari, H., Kalarot, R., Sridharan, P., ... & Davatzikos, C. (2017, September). Brain cancer imaging phenomics toolkit (brain-CaPTk): an interactive platform for quantitative analysis of glioblastoma. In *International MICCAI Brainlesion Workshop* (pp. 133-145). Springer, Cham.
- [9] Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). *Fsl. Neuroimage*, 62(2), 782-790.
- [10] Maggioni, M., Katkovnik, V., Egiazarian, K., & Foi, A. (2012). Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE transactions on image processing*, 22(1), 119-133.
- [11] Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: improved N3 bias correction. *IEEE transactions on medical imaging*, 29(6), 1310-1320.
- [12] Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P. F., Kohl, S., ... & Maier-Hein, K. H. (2018). nnu-net: Self-adapting framework for u-net-based medical image segmentation. *arXiv preprint arXiv:1809.10486*.
- [13] Chaddad, A., Daniel, P., & Niazi, T. (2018). Radiomics evaluation of histological heterogeneity using multiscale textures derived from 3D wavelet transformation of multispectral images. *Frontiers in oncology*, 8, 96.
- [14] Fischl, B. (2012). FreeSurfer. *Neuroimage*, 62(2), 774-781.
- [15] Pérez-Beteta, J., Molina-García, D., Ortiz-Alhambra, J. A., Fernández-Romero, A., Luque, B., Arregui, E., ... & Pérez-García, V. M. (2018). Tumor surface regularity at MR imaging predicts survival and response to surgery in patients with glioblastoma. *Radiology*, 288(1), 218-225.
- [16] Dancheva, Z., Bochev, P., Chaushev, B., Yordanova, T., & Klisarova, A. (2016). Dual-time point 18FDG-PET/CT imaging may be useful in assessing local recurrent disease in high grade bone and soft tissue sarcoma. *Nuclear Medicine Review*, 19(1), 22-27.
- [17] Hatt, M., Vallieres, M., Visvikis, D., & Zwanenburg, A. (2018). IBSI: an international community radiomics standardization initiative.
- [18] Gastouniotti, A., Oustimov, A., Keller, B. M., Pantalone, L., Hsieh, M. K., Conant, E. F., & Kontos, D. (2016). Breast parenchymal patterns in processed versus raw digital mammograms: A large population study toward assessing differences in quantitative measures across image representations. *Medical physics*, 43(11), 5862-5877.
- [19] Rosset, S. (2004, July). Model selection via the AUC. In *Proceedings of the twenty-first international conference on Machine learning* (p. 89).

Evaluación de la robustez de la red neuronal funcional en la enfermedad de Alzheimer

M. Revilla-Vallejo ¹, C. Gómez ^{1,2}, J. Gomez-Pilar ^{1,2}, R. Hornero ^{1,2,5}, M. A. Tola-Arribas ^{2,3}, M. Cano ⁴, J. Poza ^{1,2,5}

¹ Grupo de Ingeniería Biomédica, Universidad de Valladolid, Valladolid, España, marcos.revilla@gib.tel.uva.es.

² Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina, (CIBER-BBN), Valladolid, España.

³ Departamento de Neurología, Hospital Universitario Río Hortega, Valladolid, España.

⁴ Departamento de Neurofisiología Clínica, Hospital Universitario Río Hortega, Valladolid, España.

⁵ IMUVA, Instituto de Investigación en Matemáticas, Universidad de Valladolid, Valladolid, España.

Resumen

La enfermedad de Alzheimer (EA) es una enfermedad degenerativa con alta prevalencia en la sociedad actual. El diagnóstico de la misma es complejo y costoso. Por ello, se investiga en nuevas formas de caracterización de las alteraciones que esta enfermedad provoca en la red neuronal. En el presente trabajo se pretende evaluar la robustez de la red neuronal funcional obtenida a partir de registros de electroencefalograma. Se emplearon dos bases de datos con 195 y 202 sujetos, respectivamente. Ambas, formadas por sujetos cognitivamente sanos, pacientes con deterioro cognitivo leve y pacientes con demencia por EA. La red neuronal funcional construida a nivel de fuente con la medida de conectividad amplitud envelope correlation (AEC) fue atacada (i.e., reducción del peso de los enlaces) a lo largo de 100 iteraciones. Tras cada ataque, se calculó la correlación de Spearman entre la red atacada y la red original. La variación de correlación, como la medida de robustez propuesta, permitió replicar los resultados en ambas bases de datos. Las tendencias entre grupos mostraron una dicotomía entre bandas de frecuencia bajas y altas, con una disminución de robustez en las frecuencias bajas a medida que la EA progresa y un incremento en las frecuencias altas. Se observa, en definitiva, una dependencia de la de robustez de la red con la frecuencia.

1. Introducción

La enfermedad de Alzheimer (EA) es un trastorno neurodegenerativo que provoca cambios cerebrales estructurales y funcionales [1]. La severidad del deterioro en la función cognitiva de los pacientes con EA varía a medida que esta avanza. Las primeras manifestaciones llevan asociadas alteraciones subjetivas, que no se manifiestan en los test cognitivos convencionales [2]. El deterioro cognitivo leve (DCL) por EA, estado prodrómico de la enfermedad, se caracteriza fundamentalmente por una pérdida de memoria, mientras que las capacidades funcionales se conservan prácticamente en su totalidad [1, 2]. La EA en sus fases posteriores destaca por un deterioro cognitivo suficiente para alterar la independencia y afectar claramente a la vida cotidiana de los pacientes [1]. En la actualidad, el diagnóstico de la EA es complejo, siendo necesarios protocolos que incluyen diferentes tipos de evaluaciones por parte de neurólogos [2]. El marco de referencia diagnóstico empleado es el establecido por el *National Institute on Aging and Alzheimer's Association* (NIA-AA) [3].

La actividad electromagnética cerebral muestra patrones alterados como consecuencia de la EA [2]. La existencia de técnicas para el registro de esta actividad, como el electroencefalograma (EEG), abren la puerta a su uso diagnóstico, con el objetivo de determinar potenciales biomarcadores de la enfermedad que permitan su detección de forma precoz, precisa y con un bajo coste [1, 2]. A partir de la actividad registrada, es posible obtener patrones de conectividad funcional entre las diferentes regiones cerebrales. Estudios previos han mostrado cómo estos patrones de conectividad son alterados por la EA [4]. La teoría de grafos ha jugado un papel fundamental en la caracterización de las alteraciones de la conectividad cerebral [5]. Sin embargo, la inherente complejidad de la EA, con los mecanismos neurodegenerativos asociados, hace necesarias nuevas formas de caracterización que contribuyan a la determinación de nuevos biomarcadores partiendo de la actividad eléctrica registrada [1, 2]. Por esto, surge la necesidad de emplear en las redes neuronales el concepto de robustez previamente utilizado en redes como Internet [6, 7]. Se entiende robustez como la habilidad de una red para permanecer inalterada ante perturbaciones [7]. La robustez se ha evaluado tradicionalmente a través de la alteración de parámetros que caracterizan la red tras ser perturbada (i.e., eficiencia) [6, 7]. De esta forma, se obtiene una robustez desagregada, dependiente de los aspectos que evalúa cada parámetro.

En este trabajo se plantea la hipótesis de que la EA provoca alteraciones de la actividad neuronal que afectan a la red; lo que a su vez debería afectar a la robustez e integridad de la misma. En consecuencia, el objetivo de este trabajo es, simulando alteraciones en la red neuronal, presentar una metodología de evaluación de la robustez independiente de parámetros de red individuales y analizar cómo esa robustez evoluciona con la severidad de la EA.

2. Materiales

2.1. Sujetos

En el presente trabajo se han empleado dos bases de datos. La primera (BD1), registrada en el Hospital Universitario Río Hortega de Valladolid, está formada por 195 sujetos: 45 sujetos de control cognitivamente sanos, 69 pacientes con DCL por EA y 81 pacientes con de-

mencia debida a EA. La segunda (BD2), registrada de forma conjunta en España y Portugal, consta de 202 sujetos: 51 sujetos de control, 51 pacientes con DCL por EA y 100 pacientes con demencia debida a EA. Los diferentes grupos se determinaron siguiendo los criterios diagnósticos NIA-AA [3]. Todos los pacientes, o bien sus familiares, dieron su consentimiento informado para participar en el estudio, diseñado según el Código Ético de la Asociación Médica Mundial (Declaración de Helsinki). Los comités correspondientes aprobaron el protocolo de trabajo.

La Tabla 1 resume los datos sociodemográficos de las dos bases de datos. Los grupos de la BD1 no muestran diferencias estadísticamente significativas en la distribución de sexos (p -valor > 0.05 , test Chi-cuadrado), pero sí los de la BD2 (p -valor < 0.05 , test Chi-cuadrado). Ambas bases de datos muestran diferencias estadísticamente significativas entre grupos en la edad y en el *Mini-Mental State Examination* (MMSE) (p -valor < 0.05 , test de Kruskal-Wallis). La edad del grupo de enfermos de Alzheimer es significativamente mayor en la BD1, mientras que en la BD2, la edad del grupo de sujetos con DCL es mayor que la de los otros dos grupos. Por su parte los valores de MMSE muestran diferencias entre los tres grupos en las dos bases de datos, dando cuenta del deterioro cognitivo consecuencia de la progresión de la EA.

2.2. Registros electroencefalográficos

La actividad eléctrica cerebral se registró en ambas bases de datos con un sistema de 19 canales, con los electrodos posicionados según el sistema internacional 10-20. En el caso de la BD1 se empleó un electroencefalógrafo XLTEK[®], *Natus Medical*, con frecuencia de muestreo de 250 Hz. Por otro lado, en la BD2 se hizo uso de un Nihon Kohden Neurofax JE-921A con frecuencia de muestreo de 500 Hz. En ambos casos, se solicitó a los sujetos que permanecieran sentados, en reposo y con los ojos cerrados, obteniendo un total de 5 minutos de registro. Las señales EEG obtenidas en las dos bases de datos fueron preprocesadas de la siguiente forma [9]: (i) filtrado digital con un filtro FIR (*Finite Impulse Response*) en la banda 1-70 Hz; (ii) filtrado de ranura para eliminar la interferencia de 50 Hz; (iii) análisis de componentes independientes (ICA) para la detección y eliminación de artefactos; y (iv) segmentación de épocas de 5 s y selección de aquellas libres de artefactos mediante inspección visual. No se encontraron diferencias estadísticamente significativas entre grupos en el número de épocas libres de artefactos.

3. Métodos

3.1. Estimación de la conectividad

La actividad cerebral procedente de los 19 electrodos fue proyectada a nivel de fuente, obteniendo 68 señales correspondientes a las regiones corticales en las que se divide el volumen cerebral según el atlas *Desikan-Killiany*. El algoritmo de localización de fuentes empleado fue sLORETA (*Standardized Low - Resolution Brain Electromagnetic Tomography*) [9]. La red neuronal se construyó

	Controles	DCL	EA
BD1			
<i>N</i>	45	69	81
Edad (años)	76 [74, 79]	77 [72, 80]	82 [76, 84]
Sexo (H:M)	14:31	29:40	34:47
MMSE	29 [28, 30]	27 [26, 28]	21 [18, 24]
BD2			
<i>N</i>	51	51	100
Edad (años)	77 [74, 81]	86 [81, 90]	79 [75, 85]
Sexo (H:M)	26:25	15:36	28:72
MMSE	29 [28, 30]	23 [21, 26]	20 [13, 22]

Tabla 1. Datos sociodemográficos y clínicos de las bases de datos empleadas; *N*: número de sujetos; *H*: hombre; *M*: mujer; *MMSE*: *Mini-Mental State Examination* (rango: [0, 30]). Edad, *MMSE*: mediana [rango intercuartil].

calculando la conectividad funcional entre las señales en cada una de estas 68 regiones mediante la *amplitude envelope correlation* (AEC) ortogonalizada [9]. Esta medida estima la correlación entre cada par de señales usando sus amplitudes, además de aplicar una corrección para reducir los efectos derivados de la conducción de volumen [9]. Las redes neuronales funcionales se construyeron para las bandas de frecuencia convencionales obtenidas tras un filtrado paso-banda: delta (δ), 1-4 Hz; zeta (θ), 4-8 Hz; alfa (α), 8-13 Hz; beta-1 (β_1), 13-19 Hz; beta-2 (β_2), 19-30; y gamma (γ), 30-70 Hz.

3.2. Estimación de la robustez de la red

La estimación de la robustez de una red es un problema de investigación actual que afecta a diversos ámbitos, desde las redes tecnológicas como Internet a las redes neuronales funcionales. Esta propiedad puede revelar características inherentes a la red que los métodos de caracterización tradicionales pasan por alto [7]. Esto se consigue mediante la inclusión de perturbaciones o alteraciones en la red y la evaluación de cómo estas alteraciones afectan a la misma; en definitiva, cómo de resiliente a esas alteraciones es la red [6]. Esto es aún más relevante en el contexto de la EA, en la que los mecanismos neurodegenerativos subyacentes alteran la red y deterioran su funcionamiento a medida que la enfermedad progresa [1, 7].

En el presente trabajo, la red neuronal funcional obtenida consiste en una matriz de conectividad de tamaño 68×68 , que refleja las interacciones entre las 68 regiones corticales consideradas. Cada una de estas regiones constituye un nodo de la red. Para evaluar la robustez, la red fue atacada mediante una estrategia de ataque a nodos primarios. Concretamente, en cada iteración de ataque, se seleccionó un nodo con probabilidad directamente proporcional al grado de nodo (i.e., la suma de los valores de las conexiones de cada nodo). El valor de todas las conexiones del nodo seleccionado se redujo a la mitad [5, 10]. Tras cada iteración de ataque, se calculó la correlación de Spearman de la matriz de conectividad obtenida con respecto a la matriz de conectividad sin perturbar. De esta forma, es posible cuantificar la robustez en términos de la variación de correlación observada

a medida que la red es atacada. Una mayor variación de correlación dará cuenta de una menor robustez. Se escoge esta estrategia de ataque para alterar los nodos o *hubs* más relevantes para la red y como primera aproximación en relación con la deposición de proteína β -amiloide en ciertas regiones cerebrales característica de la EA. Del mismo modo, se escogen 100 iteraciones de ataque como valor de compromiso entre una alteración suficiente de la matriz y el coste computacional asociado.

3.4. Análisis estadístico

La normalidad de los valores de correlación a lo largo de las diferentes iteraciones de ataque se evaluó con el test de Lilliefors, mientras que la homoscedasticidad empleando el test de Levene. Los resultados obtenidos mostraron que los valores de correlación no cumplían las condiciones de los tests paramétricos. Por tanto, se empleó el test U de Mann-Whitney para evaluar las diferencias en las comparaciones dos a dos entre los tres grupos objeto del estudio.

4. Resultados y discusión

En la Figura 1 se observan los resultados de la variación de correlación para las seis bandas de frecuencia convencionales en las dos bases de datos consideradas. Es posible apreciar el gran paralelismo entre ambas. Este hecho apoya la replicabilidad del método de evaluación de la robustez empleado y, por tanto, la posibilidad de extender esta metodología a nuevas bases de datos, así como a otras patologías de carácter neurodegenerativo.

Las curvas de evolución parten del valor unitario, valor máximo de correlación, al no haber perturbado aún la red. A medida que transcurren las iteraciones de ataque, el valor de correlación disminuye: las redes son cada vez más diferentes de la red de partida, y por tanto, están menos correladas con esta. Es posible apreciar cómo la variación de correlación a lo largo de las iteraciones de ataque es similar en todas las bandas consideradas. Las bandas de frecuencia están asociadas a diferentes procesos cognitivos [11]. Sin embargo en el estado de reposo, en el que se realizaron los registros EEG, es posible que la configuración neuronal dominante sea la conocida como *default mode network* [12] y, en consecuencia, no se aprecien diferencias en la respuesta de las bandas.

En la Tabla 2 se observan los resultados estadísticos obtenidos tras comparar los valores de correlación de los tres grupos en la iteración de ataque final. A pesar de que la respuesta a los ataques es similar en todas las bandas, los patrones entre grupos cambian notablemente. Se observa como en las bandas δ , θ y γ es el grupo de controles el que menor variación de correlación muestra; en cambio, en α y β_1 es el grupo de pacientes con EA. En β_2 no hay diferencias entre grupos. Las diferencias estadísticamente significativas corroboran la alineación entre las dos bases de datos y permiten observar también las tendencias comunes entre δ , θ y γ , por un lado; y α y β_1 , por otro. Estudios previos han sugerido esta dicotomía entre altas y bajas frecuencias, con patrones

	C vs. DCL	DCL vs. EA	C vs. EA
BD1			
δ	0.00389	0.60807	0.00044
θ	0.00667	0.02080	0.21790
α	0.02801	0.00701	0.00002
β_1	0.01023	0.05937	0.00001
β_2	0.74546	0.75430	0.56163
γ	0.13175	0.50691	0.02025
BD2			
δ	0.00059	0.30726	0.0000003
θ	0.01255	0.13438	0.14820
α	0.17216	0.07002	0.00085
β_1	0.15790	0.22785	0.00121
β_2	0.32520	0.78754	0.15042
γ	0.05822	0.01073	0.00001

Tabla 2. p -valores resultado de aplicar el test U de Mann-Whitney a los valores de correlación en la iteración 100. Los p -valores significativos ($p < 0.05$) se resaltan en negrita.

de conectividad diferentes en las altas frecuencias [8], así como alteraciones en estas consecuencia de la progresión de la EA [13].

Por otra parte, la aproximación descrita en este trabajo, frente a la robustez basada en la alteración de parámetros de teoría de grafos planteada previamente [6, 7], evita la caracterización limitada de propiedades concretas de la red y permite incorporar toda la información que proporciona la matriz de conectividad.

La principal limitación de este trabajo se encuentra en el enfoque de evaluación de la robustez desagregado por bandas de frecuencia. La red neuronal desempeña su funcionamiento de forma integrada. Por tanto, es necesario un enfoque de robustez global, en el que se sintetice la información procedente de las diferentes bandas de frecuencia y se uniformicen las tendencias entre grupos. Otra limitación encontrada es el empleo de una única estrategia de ataque. Evaluar la robustez utilizando otras formas de perturbación constituye también un aspecto a abordar en futuros estudios. Por último, las diferencias estadísticamente significativas en edad y sexo entre los grupos de las bases de datos empleadas pueden sesgar los resultados obtenidos. Por tanto, es necesario evaluar el efecto de estas en análisis posteriores.

5. Conclusiones

La nueva metodología propuesta de evaluación de la robustez de una red es replicable, observándose una elevada alineación entre las dos bases de datos. Además, las variaciones de correlación y, por tanto, la robustez de la red neuronal, dan cuenta de las alteraciones consecuencia de la EA. La caracterización de la respuesta de la red neuronal ante perturbaciones puede ser clave de cara a un diagnóstico precoz de la EA. Por último, las diferencias entre las bandas convencionales sugieren una dependencia de la robustez con la frecuencia.

Agradecimientos

Este estudio ha sido financiado por el ‘Ministerio de

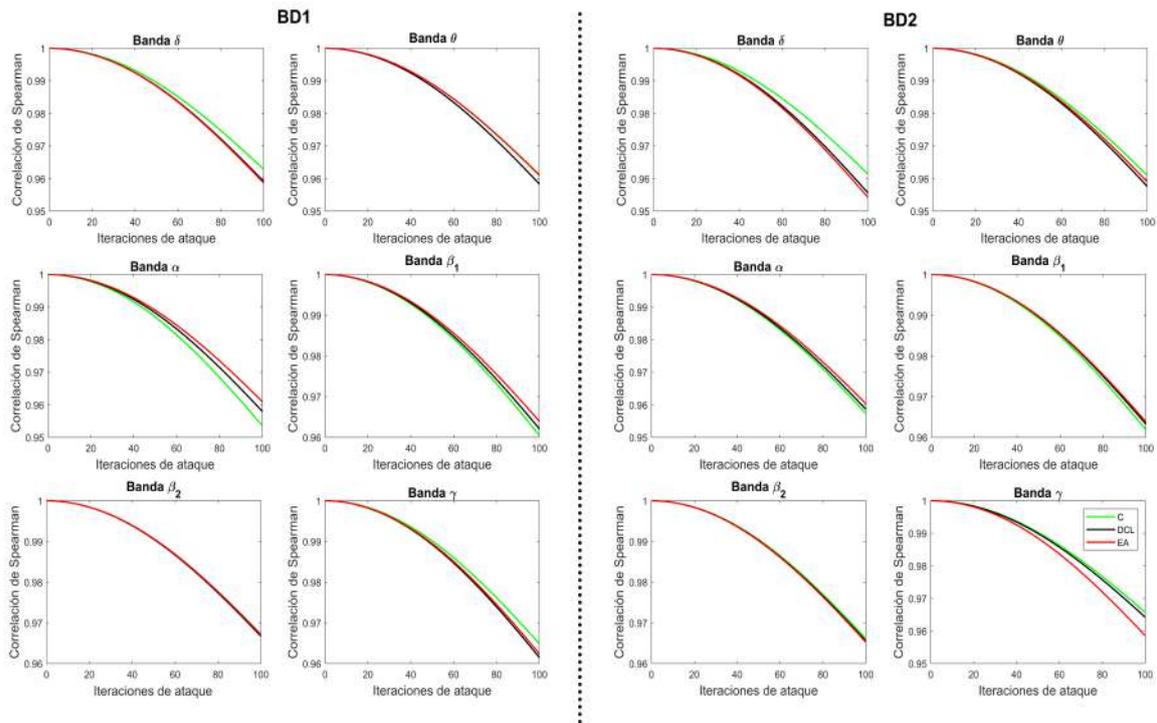


Figura 1. Evolución de la correlación en las seis bandas convencionales para la estrategia de ataque a nodos primarios en las dos bases de datos. En verde, sujetos cognitivamente sanos; en negro, pacientes con DCL por EA; en rojo, pacientes con demencia debida a EA.

Ciencia e Innovación - Agencia Estatal de Investigación' y el Fondo Europeo de Desarrollo Regional (FEDER) bajo el proyecto PGC2018-098214-A-I00, por la Comisión Europea y el FEDER bajo el proyecto 'Análisis y correlación entre la epigenética y la actividad cerebral para evaluar el riesgo de migraña crónica y episódica en mujeres' ('Programa de Cooperación Interreg V-A España-Portugal POCTEP 2014-2020'), y por el CIBER en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN) a través del Instituto de Salud Carlos III co-financiado con fondos FEDER. M. Revilla-Vallejo disfruta de un contrato predoctoral financiado por el Fondo Social Europeo, y la Junta de Castilla y León, a través de la Consejería de Educación.

Referencias

- [1] D.S. Knopman *et al.* Alzheimer disease. *Nat. Rev. Dis. Primers*, 13(7):1–21, 2021.
- [2] H. Zetterberg and B.B. Bendlin. Biomarkers for Alzheimer's disease - preparing for a new era of disease-modifying therapies. *Mol. Psychiatry*, 26(4):296–308, 2020.
- [3] J.R.Jr. Clifford *et al.* NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers Dement.*, 14(4):535–562, 2018.
- [4] C.J. Stam *et al.* Graph theoretical analysis of magnetoencephalographic functional connectivity in Alzheimer's disease. *Brain*, 132(Pt1):213–224, 2009.
- [5] M. Rubinov and O. Sporns. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3):1059–1069, 2010.
- [6] A.-L. Barabási. *Network Robustness*. In *Network Science*. 2018.
- [7] M. Bellingeri *et al.* Robustness of weighed networks. *Physica A*, 489:47–55, 2018.
- [8] P. Núñez *et al.* Characterizing the fluctuations of dynamic resting-state electrophysiological functional connectivity: Reduced neuronal coupling variability in mild cognitive impairment and dementia due to Alzheimer's disease. *J. Neural Eng.*, 16(5):1–21, 2019.
- [9] M. Revilla-Vallejo *et al.* Exploring the Alterations in the Distribution of Neural Network Weights in Dementia Due to Alzheimer's Disease. *Entropy*, 23(500):1–16, 2021.
- [10] J. Gomez-Pilar *et al.* Altered predictive capability of the brain network EEG model in schizophrenia during cognition. *Schizophr. Res.*, 201:120–129, 2018.
- [11] N. Kulkarni *et al.* *Electroencephalogram and Its Use in Clinical Neuroscience. EEG-Based Diagnosis of Alzheimer Disease*. Academic Press, 2018.
- [12] R.L. Buckner *et al.* The brain's default network: updated anatomy, physiology and evolving insights. *Nat. Rev. Neurosci.*, 20(10):593–608, 2019.
- [13] L. Koelewijn *et al.* Alzheimer's disease disrupts alpha and beta-band resting-state oscillatory network connectivity. *Clin. Neurophysiol.*, 128(11):2347–2357, 2017.

Análisis de las alteraciones en la organización jerárquica de la actividad neuronal debidas a la enfermedad de Alzheimer

V. Gutiérrez-de Pablo¹, C. Gómez^{1,2}, J. Poza^{1,2,3}, J. Gomez-Pilar^{1,2}, R. Hornero^{1,2,3}

¹ Grupo de Ingeniería Biomédica, Universidad de Valladolid, Valladolid, España, victor.gutierrez@gib.tel.uva.es

² Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), España

³ Instituto de Investigación en Matemáticas de la Universidad de Valladolid (IMUVA), Valladolid, España

Resumen

La enfermedad de Alzheimer (EA) constituye la principal causa de demencia en los países occidentales. Para su estudio se pueden utilizar diversas técnicas neurofisiológicas y de neuroimagen, entre las que destaca la electroencefalografía (EEG) por su bajo coste, portabilidad y ser no invasivo. La EEG ha revelado diferentes alteraciones que provoca la EA en la actividad neuronal: lentificación del espectro, pérdida de complejidad e irregularidad, alteraciones topológicas y dinámicas de los patrones de conectividad entre regiones y de las configuraciones de red. Las medidas empleadas para cuantificar estos cambios reflejan el comportamiento aislado de las diversas jerarquías o niveles de organización cerebral. Por ello, en este trabajo se pretende ir un paso más allá y estudiar cómo la EA afecta a las relaciones entre los diferentes niveles de esta estructura jerárquica. Para ello, se han calculado diversos parámetros a partir del EEG de 252 sujetos: 51 controles, 51 sujetos con deterioro cognitivo leve, 50 sujetos con EA leve, 50 sujetos con EA moderada y 50 sujetos con EA severa. Los resultados muestran que las medidas espectrales y no lineales tienden a agruparse, así como las medidas de conectividad, dinámica y de red. Sin embargo, entre estos dos grupos de medidas aparece una disrupción que los disocia a medida que avanza la EA, lo que sugiere que la estructura jerárquica de los niveles electrofisiológicos se ve comprometida. Los resultados obtenidos podrían ser de utilidad para obtener nuevos biomarcadores y clasificadores de cara al diagnóstico de la enfermedad.

1. Introducción

La enfermedad de Alzheimer (EA) es una patología neurodegenerativa, caracterizada por la aparición progresiva de diversas alteraciones de carácter cognitivo, funcional y conductual [1]. Es la causa de demencia más habitual, con una elevada prevalencia e impacto socioeconómico en los países industrializados [2]. Generalmente, los enfermos con EA experimentan una patología previa a la demencia, denominada deterioro cognitivo leve (DCL) [2]. Esta etapa se considera un estado prodrómico de la enfermedad [3].

A nivel fisiológico, la EA genera alteraciones cerebrales que pueden ser detectadas con el empleo de diferentes técnicas: tomografía por emisión de positrones (PET), imagen por resonancia magnética (MRI), magnetoencefalografía (MEG) o electroencefalografía (EEG) [4]. En el presente estudio se ha empleado EEG, ya que es una técnica que permite la adquisición de la actividad neuronal de forma no invasiva, tiene una alta resolución temporal, su coste es bajo y es portátil [5]. Todas estas ventajas han hecho que el EEG esté muy extendido en entornos clínicos [4], [5].

Diversos estudios han descrito una serie de alteraciones provocadas por la EA en la actividad EEG: (i) una lentificación progresiva del espectro [6]; (ii) una reducción de la complejidad e irregularidad [6]; (iii) una alteración en los patrones topológicos y dinámicos de conectividad entre regiones cerebrales [4]; y (iv) una modificación de las configuraciones de la red funcional existente entre los grupos neuronales [4]. Para analizarlas, se han desarrollado diversas metodologías. Las medidas espectrales, derivadas de la densidad espectral de potencia (DEP), han permitido explorar los patrones de sincronización local [5]. Diversas métricas que cuantifican la complejidad e irregularidad de las series temporales han permitido analizar la dinámica no lineal de la actividad cerebral [5], [6]. Los cambios en la sincronización global de los grupos neuronales han sido caracterizados mediante medidas de conectividad funcional (CF) y CF dinámica, así como las propiedades de la red neuronal con métricas derivadas de la teoría de grafos [4], [7]. A pesar de la diversidad de metodologías empleadas, algunos estudios sugieren que existen relaciones entre ellas. Estas reflejan la estructura jerárquica subyacente a la organización cerebral funcional, que puede verse alterada por la EA [8], [9]. Así, se ha observado una asociación entre la lentificación del espectro de la señal EEG y la pérdida de complejidad con el avance de la EA [5]. También se ha analizado la relación entre la reducción de complejidad e irregularidad con la alteración de los patrones de conectividad en diferentes fases de la EA [9]. Por otra parte, se ha analizado también la relación entre estas alteraciones producidas por la EA y otros factores asociados a ella, como datos sociodemográficos, información clínica y genética [10]. No obstante, se requiere un mayor estudio que permita profundizar en el conocimiento de estas asociaciones y su origen.

La hipótesis en la que se basa este trabajo es que los diferentes niveles de análisis de la actividad EEG representan la organización jerárquica neuronal y que la aparición de la EA provoca una desestructuración de esta jerarquía. Para evaluarla, se han empleado diferentes medidas derivadas de la DEP, medidas de complejidad e irregularidad, medidas de CF y CF dinámica, y medidas de organización de red. Se han generado redes de asociación para cada una de las fases de la EA, y se ha cuantificado su estructura de red mediante el cálculo de diferentes parámetros. El objetivo de este trabajo es, por tanto, estudiar las alteraciones progresivas que provoca la EA en la estructura de relaciones entre las diversas jerarquías de organización cerebral, cuantificadas mediante diversos parámetros extraídos del EEG.

2. Materiales

2.1. Sujetos

La base de datos empleada consta de 252 sujetos: 51 controles cognitivamente sanos (25 mujeres y 26 hombres; edad media de 80,1 años), 51 sujetos con DCL por EA (36 mujeres y 15 hombres; edad media de 85,5 años), 50 pacientes con demencia leve por EA (29 mujeres y 21 hombres; edad media de 80,5 años), 50 pacientes con demencia moderada por EA (43 mujeres y 7 hombres; edad media 81,3 años) y 50 pacientes con demencia severa por EA (43 mujeres y 7 hombres; edad medida de 80 años). El diagnóstico de DCL y demencia debida a EA se realizó siguiendo los criterios del *National Institute of Aging and Alzheimer's Association* (NIA-AA) [11], [12]. Los sujetos de control eran personas cognitivamente sanas, mayores de 65 años y sin patologías neurológicas previas. El deterioro cognitivo fue evaluado mediante el test *Mini-Mental State Examination* (MMSE) [10]. Los valores medios de MMSE para los distintos grupos son: 28,8 para controles, 23,2 para pacientes con DCL por EA, 22,5 para pacientes con demencia leve por EA, 13,6 para pacientes con demencia moderada por EA, y 2,4 para pacientes con demencia severa por EA. Todos los sujetos, sus cuidadores y/o familiares dieron su consentimiento informado para participar en el estudio, siendo este aprobado por el Comité Ético de la Universidad de Oporto, Portugal.

2.2. Registros electroencefalográficos

Las señales EEG se registraron mediante el empleo de un sistema de 19 canales (Nihon Kohden Neurofax JE-921A), con una frecuencia de muestreo de 500 Hz. Los electrodos se colocaron según el sistema internacional 10-20: Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fz, Cz y Pz. Se registraron 5 minutos de actividad en estado de reposo, mientras los sujetos permanecían despiertos, relajados y con los ojos cerrados en un entorno libre de ruidos. Posteriormente, se aplicó el siguiente preprocesado [10]: (i) eliminación de la línea de base; (ii) filtrado FIR (*finite impulse response*) de ranura a 50 Hz; (iii) filtrado FIR paso banda entre 0.5 y 70 Hz; (iv) análisis de componentes independientes para eliminar las componentes ruidosas; y (v) segmentación de la señal en épocas de 5 segundos y rechazo visual de aquellas artefactuadas.

3. Métodos

3.1. sLORETA

Para minimizar los efectos de conducción de volumen, las señales EEG se analizaron a nivel de fuentes cerebrales. Para la estimación de la activación de las fuentes se empleó el algoritmo *standardized Low Resolution Brain Electromagnetic Tomography* (sLORETA), que supone *a priori* que la actividad entre generadores neuronales vecinos está correlacionada [13]. En este estudio se ha trabajado con 15000 fuentes neuronales, proyectadas en 68 regiones de interés (ROI) siguiendo el atlas Desikan-Killiany [14].

3.2. Redes de asociación

Las redes de asociación reflejan las relaciones directas entre las variables consideradas, que constituyen los nodos. Las variables extraídas del EEG se calcularon por época y ROI para, posteriormente, promediarlas por épocas hasta obtener un valor por ROI en el caso de las medidas espectrales y no lineales, y entre pares de ROI para las medidas de CF y CF dinámica. Por otro lado, para las medidas de red se obtuvo un único valor. Los tipos de medidas calculadas fueron:

- 1) *Medidas espectrales*: potencia relativa (RP) en las bandas de frecuencia típicas: delta (1-4 Hz), zeta (4-8 Hz), alfa (8-13 Hz), beta 1 (13-19 Hz), beta 2 (19-30 Hz) y gamma (30-70 Hz); frecuencia mediana (MF); frecuencia alfa individual (IAF); frecuencia de transición (TF); frecuencia límite espectral del 95% (SEF95); entropía espectral (SE); entropía de Tsallis; entropía de Escort-Tsallis; y entropía de Renyi [15].
- 2) *Medidas no lineales*: complejidad de Lempel-Ziv (LZC); entropía muestral (SampEn); entropía aproximada (ApEn); entropía difusa (FuzzyEn), auto-información mutua (AMI); y medida de tendencia central (CTM) [15].
- 3) *Medidas de CF y CF dinámica*: *phase lag index* (PLI); *amplitude envelope correlation* (AEC) ortogonalizada; *overlapping weighted degree* (OWD); y participación en las bandas delta, zeta, alfa, beta 1, beta 2 y gamma [7], [16].
- 4) *Medidas de red*: coeficiente de agrupamiento (CIC); grado de red (G); longitud de camino (PL); *closeness centrality* (CC); entropía de grafo (GE); densidad global (DenGlobal); y el índice de *small world* (SW) para las bandas delta, zeta, alfa, beta 1, beta 2 y gamma, calculadas a partir del PLI y la AEC [17].

Se utilizó el test Chi-Cuadrado para obtener el grado de asociación entre las variables consideradas. Para generar las redes de asociación, los valores del estadístico Chi-Cuadrado se utilizaron para establecer los pesos de los enlaces entre las variables. Estos enlaces se umbralizaron a continuación mediante su *p*-valor corregido, con el procedimiento *false discovery rate* (FDR), para seleccionar únicamente los que eran significativos ($\alpha < 0.05$).

Para representar las redes, se empleó el algoritmo *Force Atlas 2*, que usa el grado de nodo como efecto repulsor; en concreto, los nodos con conexiones más fuertes ejercen menor repulsión hacia el resto de nodos que aquellos con conexiones más débiles [10]. Por otro lado, la anchura del enlace representa el grado de asociación entre variables, de forma que los nodos con mayor asociación tienden posicionarse más cercanos entre sí y con un enlace más grueso. Posteriormente, se calcularon diferentes parámetros sobre estas redes de asociación para cuantificar sus propiedades, todas ellas calculadas con el software *Gephi*. Concretamente, las medidas empleadas fueron: CIC, modularidad, PL, G, *eigenvector centrality*, y densidad de grafo.

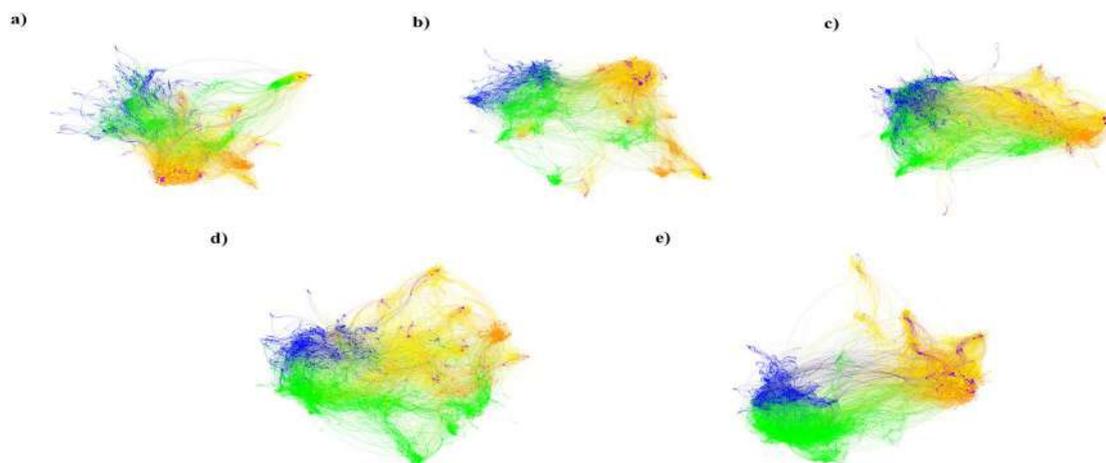


Figura 1. Redes de asociación obtenidas para cada grupo estudiado. (a) Controles; (b) Pacientes con DCL; (c) Pacientes con demencia leve por EA; (d) Pacientes con demencia moderada por EA; y (e) Pacientes con demencia severa por EA. En verde, medidas espectrales. En azul, medidas no lineales. En amarillo, medidas de CF. En morado, medidas de CF dinámica. En naranja, medidas de red.

4. Resultados y discusión

En la figura 1 se puede apreciar las redes de asociación obtenidas para cada estadio de severidad de la enfermedad, donde se muestran las asociaciones significativas entre las variables bajo análisis. En la figura 1.a, correspondiente a los sujetos de control, se aprecia como los 4 niveles de medidas aparecen conectadas entre sí, pero sin formar clústeres grandes donde se concentren los nodos. En la figura 1.b, correspondiente a los pacientes con DCL, se observa como las medidas espectrales y no lineales comienzan a separarse de las medidas de CF, CF dinámica y de red, mientras que las medidas del mismo tipo se aproximan entre sí. Estos cambios en la red se han podido cuantificar con diferentes parámetros de red, como se puede observar en la figura 2, en la cual se aprecia que los valores de CIC, modularidad y PL decrecen en este estadio, mientras que los valores de *eigenvector centrality*, densidad de grafo y G aumentan. A continuación, en la figura 1.c, que corresponde a los pacientes con demencia leve por EA, se sigue apreciando esta unión entre las variables del mismo tipo, si bien disminuye la separación entre las medidas de sincronización local (espectrales y no lineales) y las medidas de sincronización global (CF, CF dinámica y red). Los parámetros de red también muestran este cambio, ya que los valores de CIC, PL y modularidad siguen disminuyendo, mientras que aumentan los valores de *eigenvector centrality* y G; en este estadio el valor de densidad no varía. Posteriormente, en la figura 1.d, la red de los pacientes con demencia moderada por EA, las medidas de sincronización local y global comienzan de nuevo a separarse y, por otro lado, los grandes clústeres que forman las medidas espectrales y no lineales se encuentran muy próximos, lo que conlleva un valor de asociación alto, que se había observado con anterioridad [5]. Este cambio de tendencia se ha cuantificado en los parámetros de red de la figura 2, en la cual puede apreciarse como los valores de CIC, modularidad y PL pasan a incrementarse, y los de densidad y G decrecen después del incremento presentado en las fases previas de la enfermedad. Por último, en la red de la figura 1.e, que representa a los pacientes con demencia severa por EA, la asociación entre las medidas espectrales y no lineales es

más evidente, mientras que la separación entre estas medidas y las de CF, CF dinámica y red aumenta. Esto provoca la aparición de dos grandes clústeres que agrupan las medidas de sincronización local y global, aunque poseen asociaciones entre ellos. Como también puede observarse en la figura 2, los parámetros de red calculados muestran este cambio; de tal manera que los valores de G y densidad se ven incrementados, y para el resto de parámetros el valor disminuye en esta última fase de la EA.

Los resultados obtenidos muestran la organización jerárquica de la actividad neuronal y las asociaciones existentes entre los diferentes niveles que la componen. A medida que avanza la enfermedad, esta organización comienza a verse alterada, evidenciándose principalmente entre las medidas espectrales y no lineales (activación local), y las medidas de CF, CF dinámica y red (sincronización global). Las diversas regiones en las que se divide el cerebro están conectadas entre sí directamente a través de tractos anatómicos o indirectamente por asociaciones funcionales, que permiten la ejecución de los diferentes procesos [4]. Sin embargo, el procesamiento de la información necesaria para llevar a cabo estas actividades se ve afectado por la aparición de la EA, de tal manera que esta patología ha sido identificada como un “síndrome de desconexión” [4]. Estas alteraciones provocan que diferentes regiones, no necesariamente adyacentes y previamente conectadas entre sí para la ejecución de una determinada función, dejen de estar sincronizadas; esto hace que tengan que crear nuevas conexiones para llevar a cabo la misma función o que aparezca un déficit de la misma [4]. La pérdida de conectividad entre las diferentes regiones cerebrales implica que las señales neuronales generadas en cada región, y cuyas propiedades se analizan con las medidas de activación local, dejen de estar sincronizadas de forma progresiva. Esto impactaría a su vez en las relaciones jerárquicas entre las medidas de activación local (espectrales y no lineales) y las de sincronización global (CF, CF dinámica y red), lo cual se ve reflejado en las redes de asociación de la figura 1.

Finalmente, existen algunas limitaciones en el presente estudio. En primer lugar, la base de datos es limitada;

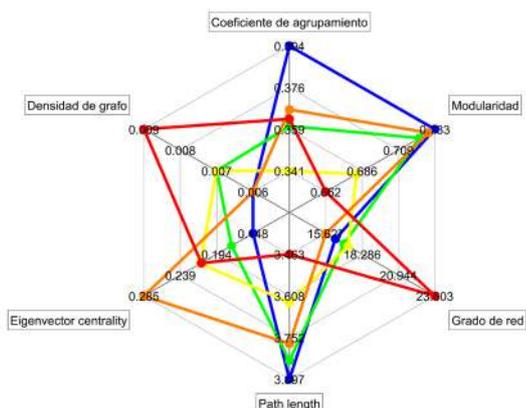


Figura 2. Valores de los parámetros de red calculados para las redes de asociación. En azul, controles. En verde, pacientes con DCL. En amarillo, pacientes con EA leve. En naranja, pacientes con EA moderada. En rojo, pacientes con EA severa.

debido a ello, sería interesante aumentar el número de sujetos para incrementar la estabilidad de las redes de asociación. En segundo lugar, el empleo de registros EEG longitudinales permitiría evaluar de manera precisa la evolución patológica de los sujetos. En tercer lugar, sería conveniente en futuros trabajos emplear métodos como *Bootstrap* para analizar si los parámetros de la red de asociación son capaces de cuantificar las posibles diferencias entre la distribución de los grupos y, por tanto, clasificar sujetos de forma individual.

5. Conclusiones

Los resultados del presente estudio apoyan la idea de que la estructura existente entre los diferentes niveles de organización jerárquica neuronal se ve afectada por las alteraciones que provoca la aparición de la EA. En concreto, se produce una disociación entre los niveles de activación local y de sincronización global, que podría deberse a una pérdida de las conexiones funcionales existentes entre las regiones cerebrales. Estos cambios pueden cuantificarse con parámetros extraídos de las redes de asociación, que muestran la potencia de esta técnica para identificar potenciales biomarcadores de esta patología.

Agradecimientos

Este estudio ha sido financiado por el proyecto ‘Análisis y correlación entre la epigenética y la actividad cerebral para evaluar el riesgo de migraña crónica y episódica en mujeres’ (Programa de Cooperación Transfronteriza INTERREG V-A España-Portugal 2014-2020) de la Comisión Europea y del Fondo Europeo de Desarrollo Regional (FEDER), por el proyecto PGC2018-098214-A-I00 del ‘Ministerio de Ciencia e Innovación – Agencia Estatal de Investigación’ y FEDER, y por el CIBER en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN) a través del Instituto de Salud Carlos III cofinanciado con fondos de FEDER.

Referencias

[1] J. L. Cummings *et al.*, *La neuropsiquiatría de la enfermedad de Alzheimer y demencias relacionadas*. Novartis, 2004.

[2] G. Livingston *et al.*, “Dementia prevention, intervention, and care: 2020 report of the Lancet Commission,” *Lancet*,

vol. 396, pp. 413–446, 2020, doi: 10.1016/S0140-6736(20)30367-6.

- [3] R. C. Petersen, “Mild Cognitive Impairment,” *Continuum (N. Y.)*, vol. 2, no. 2, pp. 404–418, 2016, doi: 10.1212/CON.0000000000000313.
- [4] M. Yu *et al.*, “The human connectome in Alzheimer disease — relationship to biomarkers and genetics,” *Nat. Rev. Neurol.*, vol. 17, pp. 545–563, 2021, doi: 10.1038/s41582-021-00529-1.
- [5] J. Dauwels *et al.*, “Slowing and Loss of Complexity in Alzheimer’s EEG: Two Sides of the Same Coin?,” *Int. J. Alzheimers. Dis.*, vol. 2011, pp. 1–10, 2011, doi: 10.4061/2011/539621.
- [6] J. Jeong, “EEG dynamics in patients with Alzheimer’s disease,” *Clin. Neurophysiol.*, vol. 115, no. 7, pp. 1490–1505, 2004, doi: 10.1016/j.clinph.2004.01.001.
- [7] P. Núñez *et al.*, “Abnormal meta-state activation of dynamic brain networks across the Alzheimer spectrum,” *Neuroimage*, vol. 232, no. January, p. 117898, 2021, doi: 10.1016/j.neuroimage.2021.117898.
- [8] J. M. Buldú *et al.*, “Frequency-based brain networks: From a multiplex framework to a full multilayer description,” *Netw. Neurosci.*, vol. 2, no. 4, pp. 418–441, Oct. 2018, doi: 10.1162/netn_a_00033.
- [9] S. Nobukawa *et al.*, “Classification Methods Based on Complexity and Synchronization of Electroencephalography Signals in Alzheimer’s Disease,” *Front. Psychiatry*, vol. 11, p. 255, 2020, doi: 10.3389/fpsy.2020.00255.
- [10] V. Gutiérrez-de Pablo *et al.*, “Redes de asociación: una nueva forma de explorar la huella fisiopatológica de la enfermedad de Alzheimer,” en *XXXVIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica*, 2020, pp. 148–162.
- [11] M. S. Albert *et al.*, “The Diagnosis of Mild Cognitive Impairment due to Alzheimer’s Disease: Recommendations from the National Institute on Aging-Alzheimer’s Association Workgroups on Diagnostic Guidelines for Alzheimer’s Disease,” *Alzheimer’s Dement.*, vol. 7, no. 3, pp. 270–279, 2011, doi: 10.1176/appi.focus.11.1.96.
- [12] G. M. McKhann *et al.*, “The diagnosis of dementia due to Alzheimer’s disease: Recommendations from the National Institute on Aging- Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease,” *Alzheimer’s Dement.*, vol. 7, no. 3, pp. 263–269, 2011, doi: 10.1016/j.jalz.2011.03.005.
- [13] R. D. Pascual-Marqui, “Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details,” *Methods Find. Exp. Clin. Pharmacol.*, vol. 24 Suppl D, pp. 5–12, 2002, doi: 841 [pii].
- [14] R. S. Desikan *et al.*, “An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest,” *Neuroimage*, vol. 31, no. 3, pp. 968–980, 2006, doi: 10.1016/j.neuroimage.2006.01.021.
- [15] J. Poza *et al.*, “Spectral and Non-Linear Analyses of Spontaneous Magnetoencephalographic Activity in Alzheimer’s Disease,” *J. Healthc. Eng.*, vol. 3, no. 2, pp. 299–322, 2012.
- [16] S. J. Ruiz-Gómez *et al.*, “Computational modeling of the effects of EEG volume conduction on functional connectivity metrics. Application to Alzheimer’s disease continuum,” *J. Neural Eng.*, vol. 16, no. 6, 2019, doi: 10.1088/1741-2552/ab4024.
- [17] M. Rubinov *et al.*, “Complex network measures of brain connectivity: Uses and interpretations,” *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010, doi: 10.1016/j.neuroimage.2009.10.003.

Modelo de *deep learning* basado en la combinación de redes neuronales convolucionales y recurrentes para clasificar eventos de apnea e hipopnea mediante la señal de oximetría

F. Vaquerizo Villar^{1,3}, D. Álvarez González^{1,2,3}, G.C. Gutiérrez Tobal^{1,3}, J. Jiménez García¹, C. A. Arroyo Domingo², F. del Campo^{2,3}, R. Hornero^{1,3}

¹ Grupo de Ingeniería Biomédica, Universidad de Valladolid, Valladolid, España, fernando.vaquerizo@gib.tel.uva.es

² Servicio de Neumología, Hospital Universitario Río Hortega, Valladolid, España, fsas@telefonica.net

³ Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina, (CIBER-BBN), España

Resumen

En este estudio se evalúa un clasificador automático de eventos respiratorios mediante la aplicación de una arquitectura de deep learning sobre la señal de saturación de oxígeno en sangre (SpO₂) en adultos diagnosticados con apnea obstructiva del sueño (AOS). Para ello, se han analizado 398 registros de SpO₂ de sujetos adultos. Se ha diseñado una arquitectura CNN+RNN compuesta por: (i) una red neuronal convolucional (CNN), que permite extraer características de los eventos respiratorios, y (ii) una red neuronal recurrente (RNN), encargada de modelar la distribución temporal de los eventos respiratorios durante el sueño. Esta arquitectura recibe a su entrada secuencias de segmentos de 30 s de la señal de SpO₂, clasificando cada segmento en 3 categorías: respiración normal, apnea e hipopnea. A continuación, se empleó un modelo de regresión para estimar el índice de apnea-hipopnea (IAH) de cada paciente. La arquitectura CNN+RNN propuesta alcanzó una precisión global del 80% y un Cohen's kappa de tres clases igual a 0.53, superando a una arquitectura CNN convencional. El IAH obtenido a partir del modelo CNN+RNN alcanzó un alto rendimiento diagnóstico, mostrando un Cohen's kappa de 0.649, así como precisiones del 90.8% y 89.2% para los umbrales del IAH de 15 y 30 eventos por hora. Estos resultados sugieren que la combinación de CNN+RNN permite extraer información sobre la distribución temporal de los eventos respiratorios, que por sí solas las CNN no pueden discernir, siendo así una herramienta útil en la detección de la AOS a partir de la señal de oximetría.

1. Introducción

La apnea obstructiva del sueño (AOS) se ha convertido en los últimos años en un importante problema sanitario a nivel mundial [1]. La AOS se caracteriza por la repetición recurrente de eventos de ausencia total (apnea) o parcial (hipopnea) de respiración durante el sueño, que derivan en una arquitectura del sueño fragmentada y no reparadora [2]. Aunque la prevalencia de la AOS es muy elevada en la población adulta (9%-38%), se trata de una enfermedad infradiagnosticada, lo cual contribuye a un mayor riesgo de alteraciones cardiovasculares, metabólicas y psiquiátricas, como la hipertensión, accidentes cerebrovasculares, diabetes y depresión [1].

A pesar de ser la técnica de referencia en el diagnóstico de la AOS, la polisomnografía (PSG) nocturna presenta importantes limitaciones. Se trata de una prueba costosa, incómoda para el paciente y de elevada complejidad, tanto en la realización como en su posterior interpretación, lo cual puede derivar en errores e inconsistencias en el

diagnóstico [3]. Ante estas limitaciones, múltiples investigaciones se han centrado en el uso de herramientas automáticas aplicadas a la detección de la AOS mediante un conjunto reducido de señales cardiorrespiratorias. Entre ellas, destaca el análisis de la señal de saturación de oxígeno en sangre (SpO₂) procedente de la oximetría nocturna, debido a su facilidad de adquisición e interpretación [4].

Diferentes estudios centrados en la señal de SpO₂ han abordado la detección automática de eventos respiratorios (apneas e hipopneas) y el diagnóstico automático de la AOS utilizando metodologías de *feature engineering*, basadas en la extracción y selección de características [4], [5]. Sin embargo, estas metodologías requieren determinar, *a priori*, las características a extraer de la señal, lo cual limita su capacidad para aprovechar toda la información contenida en el registro de SpO₂ relacionada con los eventos respiratorios. Esta limitación se puede solventar mediante los métodos de *deep learning*, que son capaces de aprender de manera automática las particularidades inherentes de los datos mediante el uso de arquitecturas multicapa con múltiples niveles de representación [6].

Estudios recientes han demostrado la superioridad de las técnicas de *deep learning* sobre los métodos de *feature engineering* para analizar señales cardiorrespiratorias en la detección de eventos apnéicos [7]. Sin embargo, los estudios centrados en el análisis de la señal de SpO₂ han abordado únicamente la detección automática de eventos respiratorios frente a episodios de respiración normal, sin diferenciar entre apneas e hipopneas [8]–[10]. Además, estos estudios han utilizado redes neuronales convolucionales (CNNs) que, a pesar de ser útiles para aprender las características de los eventos respiratorios, no consideran su distribución temporal durante el sueño [6], [11]. En cambio, las redes neuronales recurrentes (RNNs) aprenden las dependencias temporales de los datos, permitiendo así mejorar la detección de secuencias de eventos de apnea e hipopnea [11], [12], que en muchas ocasiones se presentan agrupados en forma de *clusters*.

Por tanto, el presente estudio parte de la hipótesis de que una arquitectura de *deep learning* basada en una combinación de CNN y RNN (CNN+RNN) podría extraer nueva información presente en la señal de SpO₂ capaz de mejorar la identificación y diferenciación de eventos de

apnea e hipopnea. Consecuentemente, el objetivo principal del estudio consiste en diseñar y evaluar un modelo de *deep learning* para la clasificación automática de eventos respiratorios basado en una arquitectura CNN+RNN aplicada a la señal de SpO₂. Además, se propone como objetivo secundario evaluar la utilidad del modelo CNN+RNN para la estimación del índice de apnea-hipopnea (IAH), que es el parámetro clínico para el diagnóstico de la AOS.

2. Sujetos y señales

En este estudio se ha analizado una población compuesta por 398 sujetos adultos, donde todos fueron diagnosticados con AOS (IAH ≥ 5 e/h). Todos ellos fueron remitidos a la unidad del sueño del Hospital Universitario Río Hortega (Valladolid, España), donde se les realizó la PSG. El protocolo del estudio fue aprobado por el Comité Ético y de Investigación con medicamentos (CEIm 47/16).

De acuerdo con las reglas de la Academia Americana de Medicina del Sueño, se ha considerado como evento de apnea una disminución del 90% en el flujo aéreo (FA) durante al menos 10 segundos. Asimismo, se ha considerado como hipopnea una disminución del FA de al menos un 30%, con una duración de 10 segundos o más, y que esté acompañada de una desaturación de oxígeno igual o superior al 3% y/o por un micro-despertar [2]. A partir de estas anotaciones, se obtuvo el IAH de cada sujeto como el número de eventos de apnea e hipopnea por hora de sueño (e/h). A partir del IAH estimado, se evaluó la capacidad del modelo para diagnosticar AOS leve (5 ≤ IAH < 15 e/h), AOS moderado (15 ≤ IAH < 30 e/h) y AOS severo (IAH ≥ 30 e/h).

Los registros de SpO₂ se obtuvieron durante la PSG a una frecuencia de muestreo de 16 Hz y posteriormente se remuestrearon a una frecuencia de 1 Hz, que ha demostrado ser suficiente para caracterizar la señal de SpO₂ a la vez que permite reducir el coste computacional [12]. Cada señal se dividió en segmentos de 30 s y cada segmento se etiquetó como respiración normal (N), apnea (A) o hipopnea (H) según las anotaciones proporcionadas por los especialistas al analizar la PSG. La población fue distribuida de manera consecutiva en tres grupos: entrenamiento (199 sujetos, 50%), empleado para entrenar la arquitectura CNN+RNN; validación (79 sujetos, 20%), utilizado para la optimización de hiperparámetros, y test (120 sujetos, 30%), empleado para evaluar la metodología propuesta. La tabla 1 muestra las características clínicas y polisomnográficas de la población bajo estudio.

3. Metodología

3.1. Arquitectura CNN+RNN

La figura 1 muestra la arquitectura *deep learning* propuesta, que consiste en la combinación de una CNN y una RNN. La sección de entrada consiste en una secuencia de 100 épocas consecutivas de 30 s de la señal de SpO₂. La arquitectura propuesta procesa de manera independiente cada época con una capa de distribución temporal que contiene una CNN. La CNN está compuesta por 6 bloques convolucionales con el objetivo de aprender las características de los eventos de apnea e hipopnea que contiene cada época. Cada uno de estos bloques está

	Todos	Entrenamiento	Validación	Test
Características de los pacientes en cada subconjunto de datos				
Sujetos (n)	398	199	78	120
Edad (años)	56 [47-65]	56 [47-64]	56 [47-67]	55 [44-65]
Varones (n)	278 (69.9%)	144 (72.4%)	59 (74.6%)	75 (62.5%)
IMC (kg/m²)	29.1 [26.1-33.1]	29.4 [26.0-33.5]	28.9 [26.8-31.7]	29.2 [25.6-34.0]
IAH (e/h)	35.0 [17.3-59.4]	33.2 [15.9-59.7]	35.0 [17.3-59.4]	36.2 [18.3-61.4]
Número y tipo de eventos en cada subconjunto de datos				
Normal (n)	250669 (72.5%)	127292 (73.2%)	49932 (72.5%)	73445 (71.3%)
Apnea (n)	40838 (11.8%)	20174 (11.6%)	7972 (11.6%)	12692 (12.3%)
Hipopnea (n)	54165 (15.7%)	26350 (15.2%)	10989 (15.9%)	16826 (16.4%)

Los datos se presentan como mediana [rango intercuartil] o n (%). IMC: Índice de masa corporal; IAH: índice de apnea-hipopnea

Tabla 1. Datos clínicos y polisomnográficos de los sujetos

formado por una etapa convolucional, que extrae los mapas de características mediante 64 filtros de tamaño 5, una etapa de normalización (*batch normalization*) y una función de activación ReLU (*Rectified Linear Unit*) [6]. Después del último bloque convolucional, la última etapa de la CNN, *flattening*, convierte los mapas de características 2D en series de datos 1D [6].

La capa de distribución temporal CNN es procesada mediante una RNN, para así aprender la distribución

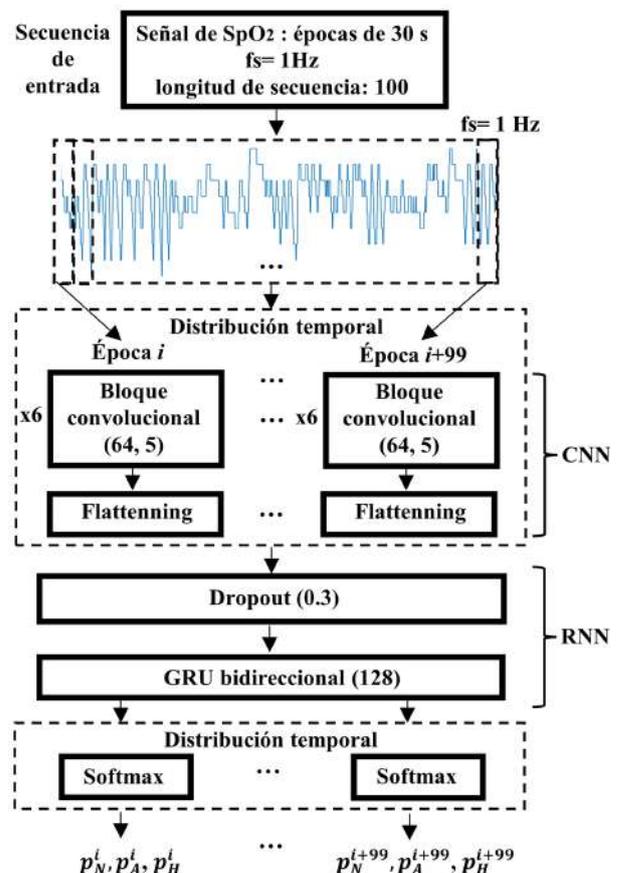


Figura 1. Arquitectura de *deep learning* basada en una combinación de CNN y RNN (CNN+RNN).

temporal de los eventos respiratorios que ocurren durante la secuencia. En primer lugar, una capa de regularización (*dropout*) se encarga de eliminar conexiones entre elementos de la red con una probabilidad de 0.3 para reducir el sobreentrenamiento [6]. A continuación, una capa recurrente bidireccional compuesta por 128 celdas de tipo GRU (*gated recurrent units*) se encarga de modelar la dependencia temporal de la secuencia de épocas de entrada, decidiendo qué información debe ser retenida y cuál debe ser olvidada de la red [6]. Finalmente, una capa de distribución temporal que contiene una función de activación *softmax* es utilizada para obtener la probabilidad de cada época i de la secuencia de ser un evento de apnea (p_A^i), hipopnea (p_H^i) o respiración normal (p_N^i).

La arquitectura CNN+RNN se ha implementado en la librería *Keras*. La inicialización de pesos de la arquitectura se hizo con el método *He-normal*. El entrenamiento se llevó a cabo con la función de pérdida *categorical cross-entropy* y el algoritmo Adam, con una tasa de aprendizaje inicial de 0.0001, para la optimización de pesos [6].

3.2. Estimación del IAH

A partir de las probabilidades de salida del modelo CNN+RNN para cada segmento de 30 s de la señal de SpO₂, se ha obtenido el IAH de cada sujeto. Para ello, se han considerado como evento apneico ($y^i = 1$) aquellos segmentos donde es mayor la probabilidad de apnea o hipopnea que de respiración normal:

$$y^i = \begin{cases} 1, & p_A^i > p_N^i \text{ o } p_H^i > p_N^i \\ 0, & \text{otros} \end{cases} \quad (1)$$

A continuación, dado un sujeto con N segmentos de 30 s de la señal de SpO₂, el IAH es estimado mediante regresión lineal, cuyos coeficientes β (pendiente) y ε (sesgo) son optimizados en el conjunto de validación [12]:

$$IAH = (\beta \cdot \sum_{i=1}^N \frac{y^i}{N}) + \varepsilon \quad (2)$$

3.3. Comparación con una CNN convencional

Se ha implementado una arquitectura de *deep learning* clásica basada en CNN para comparar el rendimiento diagnóstico del modelo CNN+RNN. La sección de entrada de esta arquitectura consiste en una época de 30 s de la señal de SpO₂. A continuación, la arquitectura CNN consta de 6 bloques convolucionales, una capa de *flattening* y una unidad de activación *softmax*, que permiten obtener las probabilidades de salida para esa época (p_A^i , p_H^i y p_N^i). A partir de las probabilidades p_A^i , p_H^i y p_N^i obtenidas para cada época, se ha estimado el IAH de cada sujeto a través de (1) y (2).

3.4. Análisis estadístico

La concordancia de las arquitecturas CNN+RNN y CNN en la clasificación automática de eventos respiratorios se evaluó por medio de la matriz de confusión (de 3 clases), la precisión global (suma de la diagonal principal) y el coeficiente *kappa* (κ). Además, se calculó, para ambas arquitecturas, el número total de eventos de apnea, hipopnea y apnea+hipopnea de cada sujeto y se compararon estos valores con los obtenidos durante la PSG

mediante el coeficiente de correlación intraclase (CCI). Asimismo, el rendimiento diagnóstico del IAH obtenido por el modelo CNN+RNN se evaluó en términos de precisión global, κ , sensibilidad (S), especificidad (E), valor predictivo positivo y negativo (VPP y VPN), razón de verosimilitud positiva y negativa (LR+ y LR-) y precisión (P).

4. Resultados

4.1. Clasificación orientada a eventos

La Figura 2 muestra la matriz de confusión en el grupo de test de los modelos CNN+RNN y CNN para clasificar automáticamente eventos respiratorios. Es importante destacar que el modelo CNN+RNN clasificó correctamente el 80.0% (83222/104043) de las épocas de 30 s, con un κ de 0.53, mientras que el modelo CNN alcanzó una precisión global del 78.8% (82087/104043) y un κ de 0.48. Además, el modelo CNN+RNN superó al modelo CNN en el CCI correspondiente a los eventos de apnea (0.931 vs 0.893), hipopnea (0.689 vs 0.543) y apnea+hipopnea (0.921 vs 0.851).

4.2. Clasificación orientada a sujetos

La figura 3 muestra la matriz de confusión del IAH estimado por el modelo CNN+RNN en el grupo de test. El modelo CNN+RNN alcanzó una precisión global del 80.0% (96/120) a la hora de determinar el grupo de severidad de la AOS, así como un κ de 0.65. Asimismo, se obtuvieron precisiones del 90.8% y 89.2% en el conjunto de test para los umbrales de 15 e/h y 30 e/h, respectivamente, como se puede ver en la tabla 2.

		CNN+RNN			CNN			
PSG	N	68878 0.93	1773 0.02	3755 0.05	N	69741 0.94	1621 0.02	3044 0.04
	A	2807 0.22	7755 0.61	2202 0.17	A	3511 0.28	6820 0.53	2433 0.19
	H	6750 0.40	3534 0.21	6589 0.39	H	8654 0.51	2763 0.16	5456 0.32
		N	A	H	N	A	H	

Figura 2. Matriz de confusión de las arquitecturas CNN+RNN y CNN en la clasificación orientada a eventos en el grupo de test.

		CNN+RNN		
PSG	1	18 0.72	7 0.28	0 0.00
	2	4 0.17	14 0.58	6 0.25
	3	0 0.00	7 0.10	64 0.90
		1	2	3

Figura 3. Matriz de confusión de la arquitectura CNN+RNN en la clasificación orientada a sujetos en el grupo de test. 1: AOS leve ($5 \leq IAH < 15$ e/h); 2: AOS moderado ($15 \leq IAH < 30$ e/h); 3: AOS severo ($IAH \geq 30$ e/h).

Método	IAH (e/h)	S (%)	E (%)	VPP (%)	VPN (%)	LR+	LR-	P (%)
CNN+RNN	15	95.8	72.0	92.9	81.8	3.42	0.06	90.8
	30	90.1	87.8	91.4	86.0	7.36	0.11	89.2

Tabla 2. Rendimiento diagnóstico del modelo CNN+RNN en el grupo de test para los puntos de corte de 15 y 30 e/h.

5. Discusión y conclusiones

En este estudio se ha evaluado un clasificador automático de eventos respiratorios (apnea, hipopnea y respiración normal) mediante la aplicación de una arquitectura novedosa de *deep learning* compuesta por una combinación CNN+RNN sobre la señal de SpO₂. El modelo CNN+RNN propuesto ha alcanzado una concordancia notable para la clasificación de eventos de apnea e hipopnea, superando a una red CNN convencional. Por tanto, la información sobre la dependencia temporal de los datos proporcionada por la RNN permite mejorar la detección de secuencias de eventos de apnea e hipopnea. En las matrices de confusión de la figura 2, se observa cómo la arquitectura CNN+RNN ha detectado con mayor precisión los eventos de apnea (61%) que los eventos de hipopnea (39%). Esto puede ser debido a la presencia de hipopneas asociadas a un micro-despertar que no están ligadas a ningún cambio en la señal de SpO₂ [2]. En la clasificación orientada a sujetos, el IAH estimado por el modelo CNN+RNN mediante un procedimiento basado en ponderación y regresión lineal ha obtenido un elevado rendimiento diagnóstico, con precisiones cercanas al 90% en los puntos de corte de 15 y 30 e/h.

Estudios previos han demostrado la utilidad de los métodos de *deep learning* para detectar automáticamente eventos respiratorios a partir de señales biomédicas, superando a los métodos clásicos de *feature engineering* [7]. Dentro de estos estudios, Vaquerizo-Villar *et al.* [8] y Mostafa *et al.* [9], [10] aplicaron redes CNN a segmentos de 60 s de la señal de SpO₂ para detectar eventos respiratorios frente a episodios de respiración normal, alcanzando precisiones en un rango 85%-95%. En nuestro estudio, se ha obtenido una precisión de 2 clases (normal vs. apnea/hipopnea) en ese rango (85.5%) con un tamaño de segmento de 30 s, que es más adecuado para detectar *clusters* de eventos respiratorios [2]. Además, en el presente estudio se aborda por primera vez la diferenciación entre apneas e hipopneas mediante una arquitectura CNN+RNN.

Este estudio presenta ciertas limitaciones. En primer lugar, la base de datos empleada en este trabajo no incluye sujetos no-AOS (IAH < 5 e/h). La inclusión de estos sujetos sería útil para medir el rendimiento diagnóstico de la arquitectura CNN+RNN en sueño no patológico. Además, la aplicación de técnicas *explainable Artificial Intelligence* podría ayudar a identificar las características de la señal de oximetría extraídas por el modelo CNN+RNN que están relacionadas con los eventos de apnea y de hipopnea, así como las diferencias entre ellas.

En resumen, una arquitectura de *deep learning* basada en una combinación de CNN y RNN ha demostrado ser muy útil para la clasificación automática de eventos de apnea e hipopnea a partir de la señal de SpO₂, superando a la arquitectura CNN convencional. Además, el modelo

propuesto ha alcanzado un elevado rendimiento diagnóstico para los umbrales de severidad de la AOS de 15 y 30 e/h. Estos resultados sugieren que las arquitecturas CNN+RNN pueden ser utilizadas para extraer información adicional sobre la distribución temporal de los eventos de apnea e hipopnea la señal de oximetría en la ayuda a la detección automática de la AOS.

Agradecimientos

Este estudio ha sido financiado por el Ministerio de Ciencia e Innovación-Agencia Estatal de Investigación (MICN-AEI) y FEDER (PID2020-115468RB-I00 y RTC-2017-6516-1), por el CIBER-BBN (ISCIII), por la Sociedad Española de Neumología y Cirugía Torácica (SEPAR 649/2018) y por la Sociedad Española de Sueño (Beca de Investigación SES 2019). F. Vaquerizo Villar cuenta con una ayuda para la Formación de Profesorado Universitario del Ministerio de Educación, Cultura y Deporte (FPU16/02938). D. Álvarez es beneficiario de una ayuda Ramón y Cajal (RYC2019-028566-I) del MICN-AEI y el Fondo Social Europeo. J. Jiménez cuenta con un contrato predoctoral concedido por la Universidad de Valladolid.

Referencias

- [1] C. V. Senaratna *et al.*, "Prevalence of obstructive sleep apnea in the general population: A systematic review," *Sleep Med. Rev.*, vol. 34, pp. 70–81, 2017.
- [2] R. B. Berry *et al.*, "Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events: deliberations of the sleep apnea definitions task force of the American Academy of Sleep Medicine," *J. Clin. sleep Med.*, vol. 8, no. 5, p. 597, 2012.
- [3] S. M. *et al.*, "Sleep apnea: a review of diagnostic sensors, algorithms, and therapies," *Physiol. Meas.*, vol. 38, no. 9, pp. R204–R252, 2017.
- [4] F. del Campo, *et al.*, "Oximetry use in obstructive sleep apnea," *Expert Rev. Respir. Med.*, vol. 12, no. 8, pp. 665–681, 2018.
- [5] M. B. Uddin, *et al.*, "Classification methods to detect sleep apnea in adults based on respiratory and oximetry signals: A systematic review," *Physiol. Meas.*, vol. 39, no. 3, 2018.
- [6] I. Goodfellow, *et al.*, *Deep Learning*. MIT Press, 2016.
- [7] S. S. Mostafa, F. Mendonça, A. G. Ravelo-García, and F. Morgado-Dias, "A systematic review of detecting sleep apnea using deep learning," *Sensors (Switzerland)*, vol. 19, no. 22, pp. 1–26, 2019.
- [8] F. Vaquerizo-Villar *et al.*, "Convolutional Neural Networks to Detect Pediatric Apnea- Hypopnea Events from Oximetry," in *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2019, pp. 3555–3558.
- [9] S. S. Mostafa, *et al.*, "Greedy based convolutional neural network optimization for detecting apnea," *Comput. Methods Programs Biomed.*, vol. 197, p. 105640, 2020.
- [10] S. S. Mostafa, *et al.*, "Multi-Objective Hyperparameter Optimization of Convolutional Neural Network for Obstructive Sleep Apnea Detection," *IEEE Access*, vol. 8, pp. 129586–129599, 2020.
- [11] S. Nikkonen *et al.*, "Automatic Respiratory Event Scoring in Obstructive Sleep Apnea Using a Long Short-Term Memory Neural Network," *IEEE J. Biomed. Heal. Informatics*, vol. 25, no. 8, pp. 2917–2927, 2021.
- [12] F. Vaquerizo-Villar *et al.*, "A convolutional neural network architecture to enhance oximetry ability to diagnose pediatric obstructive sleep apnea," *IEEE J. Biomed. Heal. Informatics*, vol. 25, no. 8, 2021.

Surgical phase recognition for non-intrusive surgical skills' assessment in educational contexts based in speech and video

C. Guzmán-García¹, P. Sánchez-González^{1,2}, I. Oropesa¹, E. J. Gómez^{1,2}

¹ Biomedical Engineering and Telemedicine Centre, ETSI Telecomunicación, Center for Biomedical Technology, Universidad Politécnica de Madrid, Madrid, Spain; carmen.guzmang@upm.es, i.oropesa@upm.es, p.sanchez@upm.es, enriquejavier.gomez@upm.es

² Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina, Madrid, Spain

Summary

Automatic surgical workflow analysis (SWA) plays an important role in the modelling of surgical processes. This can eventually be used as an assessment method for procedural and cognitive surgical skills. Current automatic approaches for SWA use videos (with accuracies varying from 0.8 and 0.9), but they do not incorporate speech (inherently linked to the ongoing cognitive process). The approach followed in this study uses both video and speech to classify the phases of laparoscopic cholecystectomy, based on neural networks and machine learning. The system was tested with a video (and its corresponding speech signal), resulting in an F1-score of 0.912. The potential of this algorithm may be further explored for the definition of training and assessment systems of surgeons' cognitive and procedural skills.

1. Introduction

Minimally invasive surgery (MIS) has evolved as the gold standard for highly accurate, sensitive, and less invasive surgical procedures [1]. These procedures have gained clinical acceptance due to their advantages compared to open surgery, such as less pain, less scarring, less damage to healthy tissue and a faster recovery [1]. Although MIS can indeed be beneficial, surgical skills for these techniques are generally difficult to acquire [2].

Surgical expertise can be very complex to define [3,4], being characterized by different types of skills, ranging from technical (e.g. dexterity, depth perception, efficiency, autonomy, etc.); to nontechnical (e.g. stress management, decision-making, etc.) and cognitive skills (e.g. conceptual and procedural knowledge) [2].

Residents and educators alike have been particularly concerned about advanced cognitive skills' acquisition and have been looking at strategies to combat this lack of experience [5]. Advanced cognitive skills are defined as the "core principles that guide the thoughts, judgements, decisions, and actions during surgical performance" [3].

One of the most extended strategies for the training and assessment of advanced cognitive skills is the analysis of surgical processes, targeted at the automatic recognition of surgical phases. Surgical processes are defined as "a set of one or more linked procedures or activities that collectively realise a surgical objective within the context of an organisational structure" [6]. One of the steps involved in surgical process modelling (SPM) is surgical workflow analysis (SWA), which is the deconstructing process by

which a surgical process is divided into a list of different phases, activities or tasks [7].

Currently, SWA focuses on analysing the laparoscopic video to recognize surgical phases and actions. Probabilistic methods have proven to be of great interest in video-based SWA [8,9]. However, they are usually trained with simple features based on low-level visual cues, typically obtained through a manual annotation process, which is virtually impossible to perform for real-time applications [10]. Thus, recent video-based SWA studies have shifted towards the recognition of surgical phases using a combination of deep learning techniques and probabilistic methods (mostly with Hidden Markov Models –HMMs–) [10–13]. The best results obtain accuracies varying from 0.8 to 0.9 approximately [10,12,14,15].

A current line of research considers monitoring audio cues during surgical procedures from the attending surgical staff, motivated by the Think Aloud theory by which verbalization of tasks are inherently linked to the ongoing cognitive processes [16]. To our knowledge, the only study in this field was reported by Suzuki et al., which attempted to use audio information to find patterns of conversation (i.e., correlating the amount of conversation to the probability of transitioning from one phase of the procedure to another) [17]. However, their recognition algorithm reached only a 51.6% accuracy. We explored this avenue in a previous study [18], where we used speech to classify surgical phases using natural language processing (NLP) and machine learning, reaching a 80.7% accuracy. The algorithm will be disclosed in section 2.2.

The main goal of this study is to obtain an algorithm capable of classifying laparoscopic surgery phases combining visual information and speech from a surgical procedure, such that our previous speech-based algorithm can be improved. Specifically, laparoscopic cholecystectomy (LC), or the removal of the gallbladder, was selected as the main procedure to analyze since it is a widespread and convenient procedure for SWA (due to the unique characteristics of each of its phases), as well as one of the interventions with the highest training rate [10,19]. The final purpose of this study is oriented towards developing intelligent systems to train and assess surgeons' advanced cognitive skills by means of comparing the normal workflow of the procedure with the actual residents' performance.

2. Materials and methods

The implemented algorithm for the recognition of surgical phases uses two different classification models: one for the endoscopic video feed, and one for the surgeon-in-training’s voice (obtained through of a clip-on microphone).

2.1. Materials

To train the endoscopic video model, the dataset from the “Surgical Workflow and Skill Analysis” sub-challenge of the Endoscopic Vision Challenge [20] was used. This dataset consists of 24 endoscopic videos obtained during LC procedures at the University Hospital of Heidelberg and its affiliate hospitals. Surgical phases for all interventions were annotated frame-wise by surgical experts. Annotated phases are shown in Figure 1.

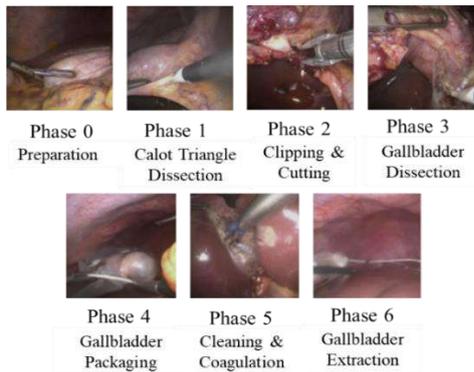


Figure 1. Laparoscopic cholecystectomy surgical phases.

To train the speech model, a database was created using 15 online surgical educational videos, in which the steps of the LC were verbally disclosed by a single trainer surgeon equipped with a clip-on microphone [18]. The audio from these videos was extracted and fragmented into smaller pieces according to the silences in the recording, using Python’s speech_recognition module [31]. This allowed to have smaller audio sequences which could be easily associated with a specific LC phase. In this case, apart from the 7 phases shown in Figure 1, an “extra” pseudo-phase was defined to account for the speech not necessarily related to the procedure itself (i.e., normal conversation between surgeons during the educational procedure). Each of the obtained audio chunks was manually transcribed into text and assigned to one of these phases.

To evaluate the performance of the algorithm combining the video- and speech-based models, a single educational video was used. This video, similar to the ones in which the speech dataset is based, consists of a high-quality endoscopic video feed on top of which the surgeon verbally discloses the ongoing procedure. The audio and image channels were separated using FFmpeg (FFmpeg, Sofia, Bucharest). The audio was chunked and automatically transcribed using Google transcriber. In our tests, this was considered to be the most efficient transcriber compared with other automatic transcribers, such as CMUSphinx (Carnegie Mellon University, Pittsburgh, Pennsylvania, USA), Microsoft Azure Speech (Microsoft, Redmond, Washington, USA) and IBM Watson Speech To Text (IBM, Armonk, Nueva York, USA). This resulted in a total

of 51 samples. 367 video samples (frames) were obtained from the video analysis. Both samples were then manually tagged independently following the same guidelines as for the individual datasets.

2.2. Speech

The speech signals are preprocessed to be divided into fragments and automatically transcribed into text using Google transcriber. Next, text data is transformed into feature vectors, using Word2Vec (i.e., a neural network which predicts the context of the word, and returns a feature vector representing said context [21,22]). Finally, the obtained features are passed through a model which combines the semantic information obtained from Support Vector Machines (SVM) with the temporal information obtained from Hidden Markov Models (HMM). The model obtained an average accuracy of 80.7% against the speech dataset described in section 2.1. Further information on the architecture and training approach of the speech-based model can be found in [18].

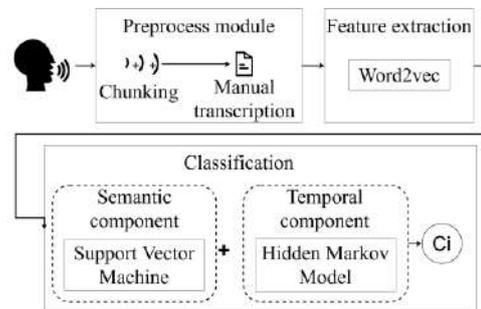


Figure 2. Workflow of the speech-based model. C_i represents the predicted class.

2.3. Video

The endoscopic video is classified by means of a Convolutional Neural Network (CNN) with a Long Short-Term Memory (LSTM) unit (Figure 3). Specifically, the CNN is trained using a two-stage approach based on VGG-16 [23].

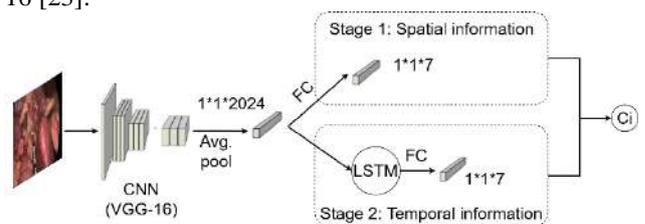


Figure 3. Workflow of the video-based model. FC represents a fully-connected layer; C_i , the predicted class.

The first training stage learns the spatial information using transfer learning methodologies. This first stage takes frames as input and outputs the predicted class.

This model is used as a feature extractor, followed by a LSTM unit, such that the temporal information is learnt. LSTM units require sequences instead of individual frames to properly analyze the temporal dependencies. To accomplish this, 9-length sequences including the frame of interest and its 4 previous and 4 following frames were extracted. The features from the individual frames included in these sequences were extracted from the model resulting

from the first training stage and used to train the LSTM unit.

2.4. Evaluation

Predictions from speech- and video-based models are synchronized in time (i.e., if an audio fragment took place between seconds 1-5, the phase predictions of the frames corresponding to these seconds were combined with the phase predictions of said audio fragment). The final prediction averages those of the speech and video models, rounded up by default. Predictions from both models are given equal weights (i.e., 50% audio, 50% video). Thus, whenever there is more than one frame for an audio fragment, the weight for the frames is equally divided to sum up to 50%. In addition, whenever the predictions of the audio-based model correspond to the pseudo-phase used to contain sentences not related to the procedure itself, they are eliminated from the analysis.

In order to assess the complete algorithm for phase recognition, performance was analyzed based on the F1 score, which is the harmonic mean of precision and recall. The F1 score of the complete algorithm was compared with that obtained by the video-based model and the speech-based model alone.

In addition, we carried out an error analysis (1) analyzing the model performance per phase by means of a confusion matrix, and (2) analyzing the error rate per phase (i.e., number of incorrectly classified samples per phase) and the average error rate (i.e., number of incorrectly classified samples). This allowed us to find the strongest and weakest points of the best performing model.

Finally, to analyze the temporality of the model, a comparison between the actual and the predicted timeline was performed. This gave further insights as to whether the model can be used in real-time surgical-educational environments, and what the model’s main flaws were.

3. Results and discussion

Table 1 represents the F1 scores and errors obtained for the test video for the video-based model, the speech-based model and the complete algorithm.

Table 1. F1-scores obtained for the video to test the algorithm of combination of signals, of the video-based model and the speech-based model.

Signal	F1-score	Avg. Error/phase	Avg. Error
Video-based model	0.743	36.01%	38.54%
Speech-based model	0.7	23.08 %	20.69%
Complete algorithm	0.912	9.42%	13.73%

The combination of audio and video models resulted in a higher F1-score than for both individually (0.912). In addition, the average error per phase was found to be 9.42% and the average error, 13.73%, improving with respect to the individual models and the state of the art (around 20%). This suggests that the algorithm created to combine the two models was effective in its purpose and yields a more accurate prediction of surgical phases. This suggests that visual information can reinforce speech

predictions (and viceversa), reducing the possibility of errors introduced by one of the models alone.

The analysis of the confusion matrix (Figure 4) shows that phases 0, 3, 5 and 6 are correctly classified in 100% of the cases. Phase 1 is wrongly classified as phase 0 in 21% of cases. The endoscopic frames of these phases are sufficiently different, but the spoken words share common elements. This may be the reason for the misclassifications in these phases. On the other hand, 11% of phase 2 samples are classified as phase 3, being the words in speech different but the images similar. Nevertheless, the instruments employed in these phases (and thus seen in the video frames) are different. A potential solution for this could be introducing an extra model with the instruments’ information to diffuse this bias.

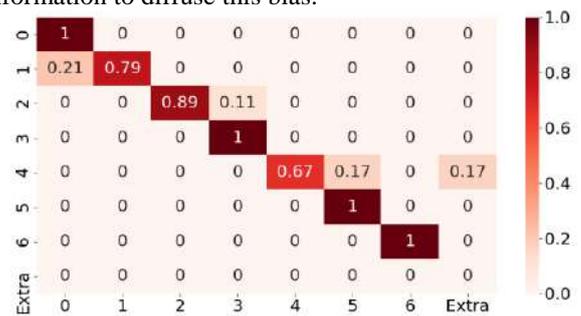


Figure 4. Confusion matrix resulting of testing the video with the combination of audio- and video-based models.

The algorithm confuses phase 4 with phase 5, due to the averaging between the video and speech samples (i.e., when the video model predicts phase 4 and the speech model phase 5, the algorithm rounds up to 5 by default). This could be easily solved by applying restrictions based on the neighbouring predictions when rounding (e.g., taking the previous and next 2 predictions and rounding up or down depending on their average). Phase 4 is also misclassified as the pseudo-phase, although only in one case, where there are no video samples for the video-based model (since the video signal was lost for the 3 seconds that this speech fragment lasts) and the speech sample was predicted not to belong to any phase per se.

Figure 5 represents the timeline comparison between the ground truth and the predicted phases in sequential order. The classification of the test dataset reveals errors which slightly disrupt the normal workflow of LC.

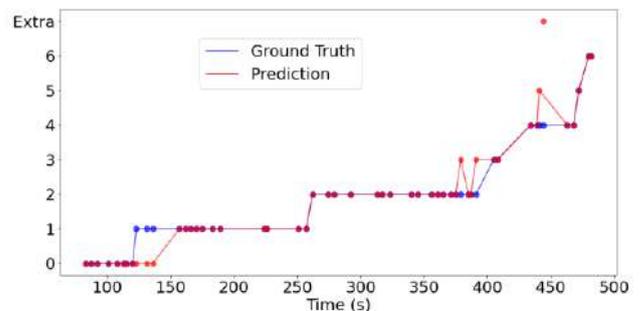


Figure 5. Timeline comparison of the ground truth (blue) and predicted phases by the SVM-HMM model (red) for the test dataset.

The predicted workflow starts phase 1 at a later instance, phase 2 appears interrupted abruptly by phase 3 (sample 38), and phase 4 by phase 5 (45). None of these affect the normal workflow (i.e., phase 1 always follows phase 0 despite the time spent, phase 3 can take place in between clips –phase 2– if the anatomy makes it necessary, and cleaning may be necessary at any given step, even more so while dissecting).

The main limitations of our study are the treatment of the pseudo-phase and the averaging between predictions. As part of our efforts to deal with these issues, we propose to apply further restrictions to the allowed predictions based on neighboring predictions. In addition to this, the algorithm could only be evaluated with a video due to the lack of public videos and the difficulties in generating new ones (because of COVID-19 restrictions in hospitals). For future studies we aim to gather new custom-made video and audio recordings at surgical training centers, to effectively extract the speech of the surgeon-in-training.

4. Conclusions

In this study, we developed an algorithm to combine the predictions of video- and speech-based phase recognition models, reaching a F1 score of 0.912. This proves that it is feasible to recognize surgical phases using a combination of (1) the surgeon-in-training speech (endorsing the Think Aloud paradigm), and (2) the endoscopic video of the procedure (as a supporting algorithm).

Our final aim is to test the validity of the proposed algorithm to obtain the workflow of the surgeon's performance and compare it with the normal workflow of the procedure. This will allow to find disruptions within the natural workflow of the procedure, which can be further analyzed as part of the assessment process of advanced cognitive skills. We also aim to test the feasibility to conduct real-time phase recognition to explore the potential of the algorithm, to provide residents with hints whenever disruptions from the normal workflow occur.

References

- [1] Verdura J, Maureen E, Ek S, et al. Systems, methods, and instruments for minimally invasive surgery. *Relat US Appl Data*. 2000;09:722.
- [2] Uemura M, Jannin P, Yamashita M. Procedural surgical skill assessment in laparoscopic training environments. *Int J Comput Assist Radiol Surg*. 2016;11:543–552.
- [3] Madani A, Vassiliou MC, Watanabe Y, et al. What Are the Principles That Guide Behaviors in the Operating Room? *Ann Surg*. 2017;
- [4] Yule S, Flin R, Paterson-Brown S, et al. Non-technical skills for surgeons in the operating room: A review of the literature. *Surgery*. 2006.
- [5] Vassiliou MC, Feldman LS, Andrew CG, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg*. 2005;190:107–113.
- [6] Neumuth T, Trantakis C, Eckhardt F, et al. Supporting the analysis of intervention courses with surgical process models on the example of fourteen microsurgical lumbar discectomies. *Int J Comput Assist Radiol Surg*. 2007;2:436–438.
- [7] Gentric JC, Trelhu B, Jannin P, et al. Development of workflow task analysis during cerebral diagnostic angiographies: Time-based comparison of junior and senior tasks. *J Neuroradiol*. 2013;40:342–347.
- [8] Padoy N, Blum T, Feussner H, et al. On-line recognition of surgical activity for monitoring in the operating room. *Proc Natl Conf Artif Intell*. 2008.
- [9] Dergachyova O, Bouget D, Huauilmé A, et al. Automatic data-driven real-time segmentation and recognition of surgical workflow. *Int J Comput Assist Radiol Surg*. 2016;
- [10] Twinanda AP, Shehata S, Mutter D, et al. EndoNet: A Deep Architecture for Recognition Tasks on Laparoscopic Videos. *IEEE Trans Med Imaging*. 2017;36:86–97.
- [11] Cadene R, Robert T, Thome N, et al. M2CAI workflow challenge: convolutional neural network with time smoothing and hidden Markov model for video frames classification. *arXiv e-prints*. 2016;
- [12] Jin Y, Dou Q, Chen H, et al. SV-RCNet: Workflow recognition from surgical videos using recurrent convolutional network. *IEEE Trans Med Imaging*. 2018;
- [13] Nakawala H, Bianchi R, Pescatori LE, et al. “Deep-Onto” network for surgical workflow and context recognition. *Int J Comput Assist Radiol Surg*. 2019;14:685–696.
- [14] Jin Y, Dou Q, Chen H, et al. EndoRCN: Recurrent Convolutional Networks for Recognition of Surgical Workflow in Cholecystectomy Procedure Video [Internet]. Hong Kong; 2016. p. 4. Available from: <http://camma.u-strasbg.fr/m2cai2016/reports/Jin-Workflow.pdf>.
- [15] Lea C, Choi JH, Reiter A, et al. Surgical Phase Recognition: from Instrumented ORs to Hospitals Around the World. *Work Challenges Model Monit Comput Assist Interv (M2CAI)*, Held Conjunction with Int Conf Med Image Comput Assist Interv. Athens, Greece; 2016.
- [16] Cowan J. The potential of cognitive think-aloud protocols for educational action-research. *Act Learn High Educ*. 2019;20:219–232.
- [17] Suzuki T, Sakurai Y, Yoshimitsu K, et al. Intraoperative multichannel audio-visual information recording and automatic surgical phase and incident detection. 2010 *Annu Int Conf IEEE Eng Med Biol Soc EMBC'10*. 2010.
- [18] Guzmán-García C, Gómez-Tome M, Sánchez-González P, et al. Speech-based surgical phase recognition for non-intrusive surgical skills' assessment in educational contexts. *Sensors (Switzerland)*. 2021;
- [19] Weede O, Dittrich F, Worn H, et al. Workflow analysis and surgical phase recognition in minimally invasive surgery. 2012 *IEEE Int Conf Robot Biomimetics, ROBIO 2012 - Conf Dig*. 2012;3:1068–1074.
- [20] Speidel S. EndoVis - Grand Challenge [Internet]. 2015. Available from: <http://www.endovis-grand-challenge.org>.
- [21] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space. *arXiv e-prints*. 2013;
- [22] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality. *Proc of the 26th International Conf Neural Inf Process Syst*. 2013;2:3111–3119.
- [23] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 3rd *Int Conf Learn Represent ICLR 2015 - Conf Track Proc*. 2015.

La importancia dinámica como indicador de comportamiento de *hub* en la región parietal de la red neuronal por defecto

A. Maturana-Candelas¹, J. Poza^{1,2,3}, R. Hornero^{1,2,3}, V. Gutiérrez-de Pablo¹, C. Gómez^{1,2}

¹ Grupo de Ingeniería Biomédica, Universidad de Valladolid, Valladolid, España, aaron.maturana@gib.tel.uva.es

² Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), España

³ IMUVA, Instituto de Investigación en Matemáticas de la Universidad de Valladolid, Valladolid, España

Resumen

Obtener información sobre la activación de regiones cerebrales específicas requiere generalmente de costosas técnicas de imagen médica. Por ello, el objetivo de este estudio es presentar la importancia dinámica (DI) como una medida de centralidad de la red cerebral. DI es un parámetro de bajo coste computacional que puede aplicarse a matrices de conectividad obtenidas a partir del electroencefalograma (EEG). Para evaluar su eficacia, se calculó la DI con el fin de detectar la red neuronal por defecto (DMN) empleando la actividad beta del EEG, la cual está implicada de manera fundamental en esta red. Para ello, se registró el EEG en la banda beta de 51 sujetos para, posteriormente, calcular la conectividad mediante la medida amplitud envelope correlation (AEC). A continuación, se calculó la DI asociada a cada región cerebral para comprobar el grado de disrupción de la red al eliminar cada nodo. Los resultados obtenidos no mostraron diferencias significativas entre los valores de DI del DMN y los del resto del cerebro. Sin embargo, la región parietal del DMN (pDMN) sí mostró valores de DI significativamente más elevados que en las regiones no pertenecientes a la pDMN (p -valor < 0.001 , test de permutaciones, $p = 1000$). Ello podría indicar diferencias en la frecuencia de procesamiento de cada región del DMN. Estas observaciones sugieren que el DI puede ser útil para el estudio de las alteraciones del conectoma cerebral. Además, corroboran que la banda beta juega un papel determinante en la activación a nivel de *hub* en la pDMN.

1. Introducción

El estudio de la conectividad funcional de la red cerebral permite analizar las alteraciones dinámicas neuronales. Estas variaciones pueden ser provocadas por diversas causas, desde resolución de tareas cognitivas hasta estados patológicos provocados por diferentes enfermedades [1]. El conectoma a nivel funcional permite estudiar la estructura anatómica cerebral [1], pues da la posibilidad de examinar en qué grado los estímulos, la concentración o los estados patológicos pueden afectar al cerebro a nivel global [1].

Uno de los protocolos más empleados para analizar la red cerebral es el estudio de la actividad neuronal adquirida en estado de reposo. Este estado se caracteriza por la ausencia de tareas cognitivas o estimulación alguna. Este procedimiento cuenta con varias ventajas, como no requerir una colaboración activa por parte del sujeto en el registro y una mayor relación señal/ruido [2]. Además, la información obtenida en estado de reposo es más

significativa que la obtenida bajo tareas a la hora de reflejar la funcionalidad global cerebral [2]. Estas características son especialmente útiles en el estudio del impacto de enfermedades neurodegenerativas, como la enfermedad de Alzheimer o de Parkinson, en los patrones de conectividad.

El estudio del cerebro en estado de reposo ha revelado diversas redes, llamadas redes en estado de reposo (*resting-state networks*), que representan funciones específicas y una topología espacial variada [2]. Entre ellas destaca la red neuronal por defecto (DMN, *default mode network*), que se caracteriza por una particular activación en estado de reposo. La DMN se puede dividir en tres regiones principales: dos prefrontales compuestas por la corteza prefrontal medial ventral (CPMV) y la corteza prefrontal medial dorsal (CPMD); y una parietal (pDMN) compuesta por la región precúnea, la corteza posterior del cíngulo y la corteza parietal lateral [3]. Estas regiones parecen estar asociadas a procesos vinculados con procesamiento emocional, actividad mental autorreferencial y el recuerdo de experiencias previas, respectivamente [3]. Sin embargo, es habitual el estudio de estas redes por medio de técnicas de imagen médica como la resonancia magnética funcional (fMRI) o la tomografía por emisión de positrones (PET). Estos procedimientos, aunque precisos, son costosos y requieren un equipo muy sofisticado. Por ello, es clave proponer medidas a partir de técnicas más sencillas que permitan obtener información sobre redes neuronales específicas.

Una de las medidas más empleadas para el estudio de los patrones eléctricos cerebrales es el electroencefalograma (EEG). El EEG mide los campos eléctricos provocados por los potenciales de acción sincronizados de las neuronas piramidales de la corteza cerebral [4]. El EEG cuenta con diversas ventajas como su portabilidad, su bajo coste y su elevada resolución temporal. Previamente, se han empleado parámetros derivados del EEG, junto con técnicas de imagen médica, para categorizar el comportamiento de la DMN. Por ejemplo, Mantini *et al.* [5] y Jann *et al.* [6] emplearon EEG junto con fMRI en su estudio de las dinámicas eléctricas de esta red funcional, obteniendo diferencias significativas en la banda beta. Además, se halló una asociación entre actividad en esta banda de frecuencia y la activación de la pDMN [6]. Por otro lado, puesto que las redes cerebrales se definen por las relaciones entre grupos de neuronas trabajando

síncronamente para llevar a cabo procesos de mayor complejidad, el cálculo de conectividad es un punto crucial en el estudio de los patrones dinámicos cerebrales.

En anteriores estudios se ha comprobado que, a partir del máximo autovalor de una matriz de adyacencia, se puede extraer una gran cantidad de información sobre la estructura de una red y de cómo esas estructuras pueden afectar a la dinámica de sus procesos [7]. A partir de esta premisa, proponemos en este estudio el uso de una medida llamada importancia dinámica (DI) [7] con el fin de evaluar la capacidad de detección de la DMN a partir de la actividad en la banda beta del EEG. Usando esta métrica, proponemos identificar las regiones cerebrales asociadas a una mayor gestión de información (o *hubs*) que, hipotéticamente, deberían corresponder a áreas pertenecientes a la DMN. De este modo, validaremos el uso de la DI como un método potencial para caracterizar redes neuronales a partir de señales EEG. Además, se comprobará en qué medida la banda beta es capaz de discriminar regiones pertenecientes a la DMN. Así pues, el objetivo de este trabajo es calcular la DI a partir de la banda beta del EEG en cada región cerebral de interés (ROI) y comprobar las diferencias entre los valores obtenidos en la DMN y en el resto del cerebro.

2. Materiales y métodos

2.1. Sujetos

En este estudio participaron 51 sujetos cognitivamente sanos (26 hombres y 25 mujeres) con una media de edad de 80.14 ± 7.09 años (media \pm desviación típica, SD). Se obtuvo el consentimiento informado de cada participante según las recomendaciones del Código Ético de la Asociación Médica Mundial (Declaración de Helsinki). El protocolo fue aprobado por el Comité de Ética de la Universidad de Oporto (Oporto, Portugal. Informe nº 38 / CEUP / 2018).

2.2. Registro y preprocesado de EEG

Por cada sujeto, se registraron 5 minutos de actividad cerebral en estado de reposo. El dispositivo empleado fue un equipo de adquisición de EEG Nihon Kohden Neurofax JE-921A de 19 canales, en los electrodos F3, F4, F7, F8, Fp1, Fp2, T3, T4, T5, T6, C3, C4, P3, P4, O1, O2, Fz, Cz y Pz, del sistema internacional 10-20. Se estableció a 500 Hz la frecuencia de muestreo. La adquisición se realizó en un entorno libre de ruido, con los ojos cerrados y en una posición relajada. Se vigiló la somnolencia de cada sujeto durante el procedimiento.

Los datos obtenidos de cada registro fueron preprocesados según el siguiente protocolo [9]: (i) eliminación de media; (ii) filtro FIR (*finite impulse response*) tipo *notch* de ventana Hamming a 50 Hz; (iii) filtro FIR paso banda de ventana Hamming en la banda beta (β , 13-30 Hz); (iv) análisis de componentes independientes (ICA) para eliminar componentes asociadas a ruido; (v) segmentación del registro en épocas de 5 s y rechazo visual de las contaminadas con artefactos. La media de épocas libres de artefactos fue de 43.63 ± 11.07 épocas (media \pm SD). A continuación, se obtuvo la actividad cerebral a nivel de fuente mediante el algoritmo sLORETA (tomografía

electromagnética de baja resolución estandarizada) [10]. Este procedimiento asume una correlación máxima entre fuentes neuronales próximas para resolver el problema inverso [10]. La localización de fuentes se realizó considerando 15000 posibles fuentes para después agruparlas en 68 ROIs. Estas regiones fueron configuradas acorde al atlas cerebral de Desikan-Killiany, el cual clasifica cada giro según su función dominante [11]. Los análisis de EEG, la etapa de procesado y la visualización de resultados se realizaron con MATLAB® (versión R2018a, Mathworks, Natick, MA) y *Brainstorm toolbox*.

2.3. Procesado de EEG

2.3.1. Cálculo de AEC

En primer lugar, se estimó los valores de conectividad mediante la medida *amplitude envelope correlation* ortogonalizada (AEC) del EEG filtrado en la banda beta. La AEC permite cuantificar el grado de correlación entre las envolventes de dos señales [12]. Valores altos de AEC indican fluctuaciones sincrónicas de las envolventes entre oscilaciones. El uso de la AEC ha sido recomendado anteriormente por su consistencia y por su menor sensibilidad al ruido frente a otras métricas de conectividad [13]. Se calculó la AEC para cada época libre de artefactos en cada registro. A continuación, se promediaron los resultados por épocas, obteniendo una matriz de dimensiones 68x68 por cada sujeto.

2.3.2. Cálculo de DI

A partir de los resultados de AEC, se calculó la DI para cada matriz de conectividad. La DI cuantifica el grado de disrupción de la red al eliminar un nodo, de modo que puede estimarse su importancia en la integridad funcional de la misma [7]. La DI de un nodo se define como la diferencia normalizada del autovalor máximo de la matriz de conectividad eliminando ese nodo y el de la matriz original. La DI se obtiene a partir de una matriz de adyacencia; en nuestro estudio, usaremos una matriz de conectividad procedente del EEG para estudiar la estructura funcional del cerebro. Ello permite cuantificar objetivamente la importancia relativa de cada elemento de la red. La DI se representa matemáticamente como [7]:

$$DI_k = -\frac{\Delta\lambda_k}{\lambda}$$

siendo λ el autovalor máximo de la matriz de conectividad original, λ_k el autovalor máximo de la matriz de conectividad estableciendo como ceros todos los valores de la fila y la columna del nodo k y $\Delta\lambda_k$ la diferencia entre λ_k y λ . Aplicando este procedimiento para cada nodo, se obtuvo un vector de 68 valores por sujeto. Finalmente, se promediaron los valores de DI de cada nodo entre sujetos.

3. Resultados

Se calculó la DI para cuantificar la importancia de cada ROI de la red cerebral en la banda beta. Con el fin de comprobar la significancia estadística de las diferencias entre la DI de las ROIs pertenecientes y no pertenecientes a la DMN, se usó el test de permutaciones. Las regiones del atlas de Desikan-Killiany clasificadas como pertenecientes a la DMN se describen a continuación.

Región pDMN: región precúnea (ROIs 51 y 52), corteza del cíngulo posterior (47 y 48), istmo del cíngulo (21 y 22), parietal lateral inferior (15 y 16), supramarginal (63 y 64) y paracentral (33 y 34). Regiones CPMV y CPMD: superior frontal (57 y 58), frontal medio rostral (55 y 56), cíngulo anterior rostral (53 y 54), parsorbitalis (39 y 40), orbitofrontal medial (29 y 30) y temporal medio (31 y 32). Estas regiones fueron consideradas según la literatura al respecto [14, 15]. Las diferencias en los valores de DI de las regiones del DMN con respecto al resto del cerebro no fueron estadísticamente significativas. Sin embargo, los valores de DI de la pDMN fueron más elevados de manera estadísticamente significativa que los valores de DI de las ROIs no pertenecientes a la pDMN (p -valor < 0.001 , test de permutaciones, $p = 1000$). Lo opuesto sucedió con los valores de DI obtenidos en las regiones CPMV y CPMD, siendo significativamente inferiores estadísticamente (p -valor < 0.001 , test de permutaciones, $p = 1000$). Los valores de DI obtenidos en la banda beta se muestran en la figura 1.

4. Discusión

En este estudio se ha cuantificado el grado de disrupción en la red cerebral que provoca la eliminación de cada ROI. Con ello, se puede estimar en qué medida cada región juega un papel más o menos fundamental en los patrones espontáneos de conectividad. Se ha observado en la banda beta unos valores significativamente más elevados en la región pDMN que en el resto del cerebro.

La DMN ha sido estudiada previamente por su particular relación con el estado de reposo. En estados de vigilia relajada y en ausencia de estímulos que requieran de atención, la DMN muestra una mayor activación [14]. Además, ésta cesa en el momento en el que se realiza una tarea cognitiva o los niveles de estrés aumentan [15]. Sin embargo, la medida propuesta en este estudio sólo detecta una activación significativamente más alta en la pDMN, siendo consistente con resultados anteriores [6]. Ello puede deberse a la especialización de cada una de las regiones principales que componen esta red. Por un lado, las áreas pertenecientes a la CPMV son regiones críticas en el procesamiento de la información sensitiva del mundo exterior [15]. Por otro lado, la CPMD ha sido asociada al desarrollo de juicios autorreferenciales sobre sucesos externos [15]. Puesto que los registros se realizaron en total ausencia de estímulos, es poco probable que estos procesos ocurriesen o, posiblemente, se dieran lugar en otras bandas de frecuencia. Así pues, lo más probable es que la tercera región del DMN, el pDMN, sea la región más activa, dadas las condiciones de registro. Además, la pDMN se ha asociado consistentemente con el recuerdo de elementos previamente experimentados [15], lo cual es más razonable que ocurra durante la adquisición del EEG.

Otra razón por la que puede deberse la ausencia de actividad beta en las regiones frontales de la DMN puede ser por el comportamiento de *hub*; esta característica refleja en qué medida un nodo es un elemento central de la red. Estos nodos son considerados puntos de alta vulnerabilidad ya que, al ser atacados, provocan mayores disrupciones en

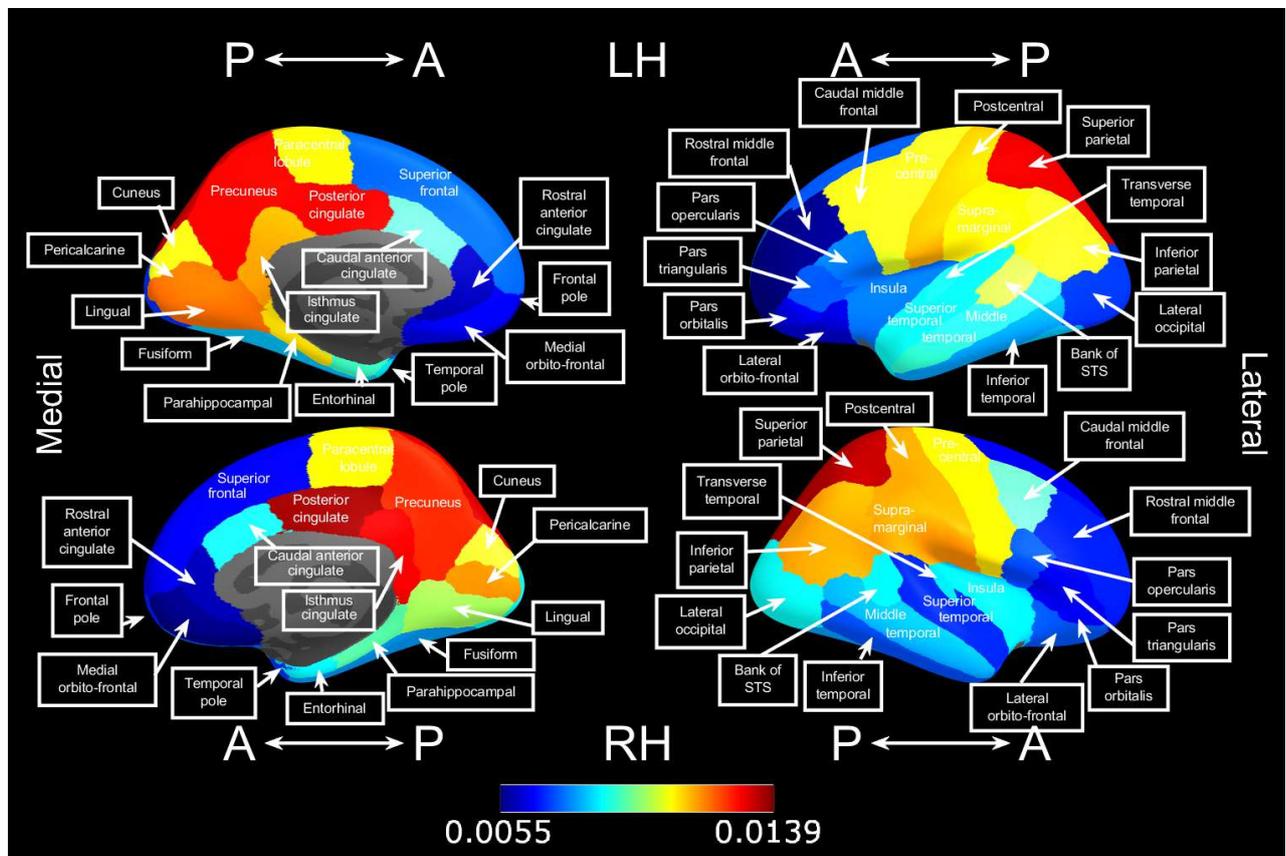


Figura 1. Valores de DI en la banda beta por cada ROI según el atlas de Desikan-Killiany. Fila superior: hemisferio izquierdo (LH). Fila inferior: hemisferio derecho (RH). A: región anterior. P: región posterior.

la estructura funcional de la red. En un estudio previo, se calculó el grado de interrupción de cada ROI mediante un análisis multiplex [16]. En pacientes que sufrían la enfermedad de Alzheimer se comprobó una vulnerabilidad mayor en las regiones parietales de la DMN que en las frontales [16]. Puesto que el DI podría ser sensible, no sólo a la actividad en cuanto a intensidad de procesamiento, sino también a la gestión del flujo de información (una característica de los *hubs*), podría ser esa la razón por la que sólo se visualiza la pDMN con este parámetro. Por otro lado, a pesar de que los estados de relajación suelen estar asociados a frecuencias más reducidas, es posible que los procesos cognitivos durante la adquisición hayan jugado un papel importante en los resultados. Por ejemplo, previamente se ha asociado la banda beta a procesos de memoria activa [17], los cuáles podrían estar directamente relacionados con las particularidades funcionales de la pDMN.

Finalmente, es preciso destacar ciertas limitaciones metodológicas. En primer lugar, este estudio cuenta con una base de datos reducida, por lo que existe la posibilidad de que valores espurios puedan generar un sesgo en nuestros resultados. Para subsanar esta limitación, es preciso aumentar la base de datos con nuevos registros de EEG. En segundo lugar, la edad media de los participantes es relativamente elevada. Aunque el estado cognitivo promedio es saludable, una edad avanzada podría repercutir en alteraciones en el EEG debido a senescencia no patológica. Por ello, para comprobar la validez de la DI en la detección de redes neuronales, será necesario replicar el estudio en bases de datos con edades diversas. Por último, limitar el estudio a una sola banda podría estar mostrando evidencia de sólo una parte del fenómeno subyacente. Estudiar cómo interactúa esta medida en otras bandas de frecuencia, o incluso en interacciones interbanda, es necesario.

5. Conclusiones

En este trabajo, se ha propuesto una medida que no se había empleado previamente para el análisis de señales EEG. Nuestros resultados sugieren que la DI en la banda beta es sensible a la activación de regiones asociadas al estado de reposo, como la pDMN. Esta medida podría ser útil en el estudio de los patrones de actividad en situaciones anómalas, como condiciones patológicas. Ello permitiría evaluar en qué medida las alteraciones cognitivas se podrían ver reflejadas en interrupciones de la red cerebral de manera eficiente.

Agradecimientos

Esta investigación ha sido financiada por la ‘Comisión Europea’ y el ‘Fondo Europeo de Desarrollo Regional’ (FEDER) a través del proyecto ‘Análisis y correlación entre la epigenética y la actividad cerebral para evaluar el riesgo de migraña crónica y episódica en mujeres’ (‘Programa de Cooperación Transfronteriza Interreg V-A España-Portugal 2014-2020’), por el ‘Ministerio de Ciencia e Innovación - Agencia Estatal de Investigación’ y FEDER a través del proyecto PGC2018-098214-A-I00, y por el ‘CIBER en Bioingeniería, Biomateriales y

Nanomedicina (CIBER-BBN)’ a través del ‘Instituto de Salud Carlos III’ cofinanciado con fondos FEDER.

Referencias

- [1] Buckner, R. L., *et al.* Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. *Journal of Neuroscience*, 29(6), 2009, 1860-1873.
- [2] Smitha, K. A., *et al.*, Resting state fMRI: A review on methods in resting state connectivity analysis and resting state networks. *The Neuroradiology Journal*, 30(4), 2017, 305-317.
- [3] Raichle, M. E. The brain's default mode network. *Annual Review of Neuroscience*, 38, 2015, 433-447.
- [4] Kirschstein, T., & Köhling, R. What is the source of the EEG?. *Clinical EEG and Neuroscience*, 40(3), 2009, 146-149.
- [5] Mantini, D., *et al.* Electrophysiological signatures of resting state networks in the human brain. *Proceedings of the National Academy of Sciences*, 104(32), 2007, 13170-13175.
- [6] Jann, K., *et al.* Topographic electrophysiological signatures of fMRI resting state networks. *PloS One*, 5(9), 2010, e12945.
- [7] Restrepo, J. G., *et al.* Characterizing the dynamical importance of network nodes and links. *Physical Review Letters*, 97(9), 2006, 094102.
- [8] Folstein, M. F., *et al.* “Mini-mental state”: a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3), 1975, 189-198.
- [9] Maturana-Candelas, A., *et al.* EEG characterization of the Alzheimer's disease continuum by means of multiscale entropies. *Entropy*, 21(6), 2019, 544.
- [10] Pascual-Marqui, R. D. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods & Findings in Experimental & Clinical Pharmacology*, 24(Suppl D), 2002, 5-12.
- [11] Desikan, R. S., *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, 31(3), 2006, 968-980.
- [12] Bruns, A., *et al.* Amplitude envelope correlation detects coupling among incoherent brain signals. *Neuroreport*, 11(7), 2000, 1509-1514.
- [13] Colclough, G. L., *et al.* How reliable are MEG resting-state connectivity metrics?. *Neuroimage*, 138, 2016, 284-293.
- [14] Raichle, M. E., *et al.* A default mode of brain function. *Proceedings of the National Academy of Sciences*, 98(2), 2001, 676-682.
- [15] Raichle, M. E. The brain's default mode network. *Annual Review of Neuroscience*, 38, 2015, 433-447.
- [16] Yu, M., *et al.* Selective impairment of hippocampus and posterior hub areas in Alzheimer's disease: an MEG-based multiplex network study. *Brain*, 140(5), 2017, 1466-1485.
- [17] Hsu, C. C., *et al.* Analyze the beta waves of electroencephalogram signals from young musicians and non-musicians in major scale working memory task. *Neuroscience Letters*, 640, 2017, 42-46.

Sub-surface laser engraving application for hexagonal SiPM

David Pérez-Benito¹, Rigoberto Chil¹, Leandro A. Hidalgo-Torres¹, Juan José Vaquero^{1,2}

1: Departamento de Bioingeniería e Ingeniería Aeroespacial, Universidad Carlos III de Madrid, Madrid, España, david.perez.benito@uc3m.es

2: Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, España

Abstract

Positron Emission Tomography detectors are commonly dense pixelated inorganic scintillators. These are embedded in a reflector whose thickness hinders the sensitivity. Pixel arrays manufacturing is laborious and expensive. The extensively known laser processing material can be employed as an alternative. A grid of laser induced microcracks at the focal spot creates a semitransparent optical barrier to guide scintillation light towards the photosensor. In this work, a cost-effective nanosecond laser has been used to validate this approach. Its design flexibility has been exploited to develop hexagonal patterns of 1.25 mm of edge to serve a hexagonal photodetector, building block of an innovative quasi-spherical scanner. Field flood images of individual readout in LYSO blocks of 5 mm, 7 mm and 10 mm of thickness are presented alongside the pixel resolvability. Crystals have been also tested with four channels multiplexing readout. Depth of interaction has been developed with two shifted layers of 5 mm in a 10 mm crystal. It has been also proved that sensitivity is not affected by measuring the count rate with a collimated beam throughout the crystal surface. The reduction of analog inputs in the multiplexing scheme together with the crystal resolution confirms the validity of the engraving solution in a complete system.

1. Introduction

Positron Emission Tomography (PET) camera detectors rely on dense inorganic scintillators to halt gamma photons resulting in the decay of injected radioactive tracers. The interaction of ionizing radiation with the scintillator brings about an isotropic burst of light that propagates throughout the crystal until it reaches the interface with the photosensor, it escapes through any of the other sides of the volume, or it is absorbed within the scintillator bulk. The unceasing seek of improvement of spatial resolution led, since the early PET stages, to segmented detector design in which millimetric scintillators crystals were embedded in a reflective array. Thus, scintillation light is partially confined in one crystal and its neighbors and, subsequently, the scintillation position in the plane of the detector can be estimated by simple centroid computation.

The depth at which the scintillation takes place, known as Depth of Interaction, or DOI for short, is lost since the crystal surface is uniformly illuminated. Depth decoding is of great value since it can be used to reduce the loss of spatial resolution due to imprecise gamma positioning especially at the edges of the scanner. To the date, several DOI approaches have been developed. Among others, offset of two crystal array relative to each other [1] and scintillators with different decay time or light output [2].

Various materials have been employed to create the optical guide. The most common materials are Teflon, Vikuiti™ and Barium sulfate $BaSO_4$. The thickness to achieve the desired reflectance is in the range of few dozens of micrometers. This finite space between crystals reduces unavoidably the detection capabilities since gamma photons can freely go in the intervening reflective material. The term packing factor is defined as the ratio of the total crystal surface to the surface of the entire array. Common values spans from 0.8 to 0.9 [3].

Another doable alternative is the application of the extensively known material laser processing to imprint scintillation blocks with the desired discrete pattern [4] [5], known as sub-surface laser engraving. In a nutshell, the electric field of the laser's wave can drive the crystal lattice into vibration that it is eventually transmitted around the solid by the linking of molecules. Under high irradiance conditions, the molecular bonding is stretched beyond the mechanical strength and the material melts or fracture. In the scintillator used in PET, melting temperatures can be reached with nanosecond laser pulses. For ultrashort pulses, picosecond or shorter, the deposition of energy rate is much faster than thermal diffusion, the molecular bonding can be further loosened beyond the melting point and the material is evaporated or solid-state plasma appears. The flexibility in the engraving position, geometry and optical barriers permeability of this technique has been exploited in different designs. Several authors has revised the common square grid in a pitch range from 2 mm to 0.8 mm [5]. DOI has been estimated with varying permeability along the depth direction [6] among others. The thickness of the scintillation blocks varies from 5 mm to 9 mm in the nanosecond lasers [4] to 20 mm with picosecond pulses.

However, this technique is commonly prohibited by the cost of the equipment. This work devises the application of sub-surface laser engraving with a cost-effective nanosecond laser in LYSO scintillation blocks up to 10 mm thick. Furthermore, the flexibility in the geometry of the crystals is exploited to serve a hexagonal photodetector, building block of an innovative quasi-spherical PET scanner design [7]. Field flood images and pixel resolvability in crystals of 5 mm, 7 mm and 10 mm thick with individual readout are presented. Field flood imaged have been also acquired with four channels multiplexing along the diagonals of the hexagonal photodetector. Two shifted layers detector with DOI capabilities has been also developed in a crystal of 10 mm of height. Finally, it has been proved that sensitivity is not affected by estimated the count rate along the crystal with a collimated beam.

2. Methods

The system setup is schematically shown in (left). The sample is accurately laid on the XY linear stage. The depth of the focal spot is controlled by a stage in the beam direction that moves the lens vertically. The laser is Nd-YAG with wavelength of 532 nm and pulse width of 6-8 nsec. The lens is a LINOS F-Theta-Ronar with focal length of 100 mm, design wavelength of 532 nm and spot diameter of 15 μm . The complete system was purchased to the company Wuhan Syntony Laser Co., Ltd. Its commercial focus is the souvenir engraving rather than the scientific field, which makes it a competitive solution pricewise but poses a challenge in terms of spatial and laser energy precision. This has been partially overcome by a stringent calibration protocol based on visual inspection of test pulses under the microscope. The linear stages were adjusted until the focal spot was within the system accuracy of 10 μm in the three axis and a fine spot diameter of about 15 μm in average was achieved. The protocol was often repeated to prevent misalignment.

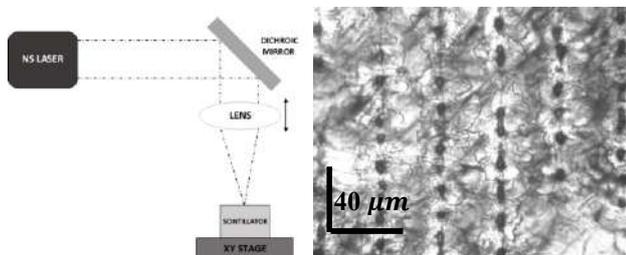


Figure 1. Left: Engraving experimental setup. The engraving height is achieved by the vertical linear stage that controls the position of the lens with respect to the sample. Right: Square engraving grid. The vertical orientation is the beam direction.

To produce a crack in the medium, the pulse power must exceed the laser induced damage threshold. Power density depends on the repetition pulse rate, pulse intensity, focal spot diameter, laser mode and wavelength. The last three are intrinsic characteristics of the equipment and they cannot be modified. The first two are set empirically closely above the damage threshold so that thermal stress and fatigue is kept at a minimum. Extensive testing has been conducted to find the most adequate parameters. For this system and LYSO samples, the diode pump voltage is set at 5.0 V and pulse repetition rate was set a 500 Hz.

(right) shows a square grid with separation of 40 μm where the top-down orientation of the image is the beam direction. It can be seen that pulses are correctly engraved in the perpendicular direction to the laser beam (horizontal direction of the figure) but some points are closer than others and some are even missing along the beam direction (lower right corner). The presence of mechanical polishing faults of the order of micrometers in the entrance face of the sample alters the laser focal spot. Unfortunately, this is unavoidable unless further chemical polishing by acid etching is done. On the other hand, the induced damage is static in nature so that aiming right above the damage threshold might not result in a microcrack in some

locations. These two reasons pose the main precision limitation of this technique.

The optical permeability of the borders depends on the engraving density. Dense engraving is desired for PET applications, but the probability of fatal failure of the crystal is higher since wide fissure might arise. Therefore, density is selected by carefully analysis of crack morphology under the microscope. Testing has been conducted in the range of 40 μm to 100 μm . Proper density and optical border capabilities has been found in the range of 50 μm to 60 μm . Contiguous points have not been engraved in the same pulse scan. The goal is to engrave always in thermally stable surroundings so that thermal stress is not propagated along the crystal.

Energy deposition by the laser pulse in transparent material is restraint to the focal spot. The internal temperature distribution will depend mainly on the thermal conductivity of the material and the rate of temperature change. This causes a temperature rise at the center and radial thermal stress due to local temperature gradients. Temperature gradients are higher in the engraving points close to the surface of the specimen. Cooling of the focal spot bring the surface into a state of tension. The entrance surface of the laser radiation is also exposed a thermal fatigue due to repeated and fluctuating thermal stress. This is one of the reasons why the fluence should be kept as low as possible over the damage threshold. Crystals of 7 mm and 10 mm are engraved both sides half of their thickness to alleviate thermal fatigue.

The engraving presented in this study is a honeycomb pattern of 35 hexagons of 1.25 mm of edge that better suits a custom-made hexagonal MPPC detector of 2.25 mm, (right). A sketch of the layout is displayed in (left). The scintillator used are square LYSO blocks of 15x15 mm with thickness of 5 mm, 7 mm and 10 mm. The LYSO was coupled to the hexagonal MPPC by optical grease BC-630. Digitalization and readout has been performed with 32 channel 125 MHz PicoDigitizer Nutaq system.

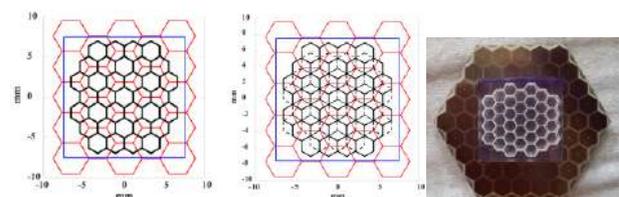


Figure 2. Left: sketch of the MPPC channels (red), engraved crystals (black) and square scintillator (blue). Middle: DOI design. The solid line is the upper layer and the dashed lines represent the lower half. Right: engraved sample over the hexagonal MPPC.

3. Results

3.1. Pixel Resolvability

Pixel resolution studies are conducted with uniform irradiation with ^{22}Na point radioactive source placed at 15 cm from the detector's front. There was no wrapping around the crystal. The 2-D position histograms were computed as the center of mass with the integral of the pulses sensed by each channel for each event. Figure 3

shows the 2-D position histogram and the profile along the three diagonals for a block of 7 mm. lists the maximum, minimum and average peak-to-valley ratios along the diagonals and rows for the three engraved thickness.

The 35 pixels are clearly visible. It is worth mention that the four edges are prominent in the histogram because light gets trapped between the borders of the crystal and the far side of the engraving. It is also notable the presence of another two points at both sides of diagonal 1. It is due to the reflections at the convex side of the hexagons and the borders of the LYSO block.

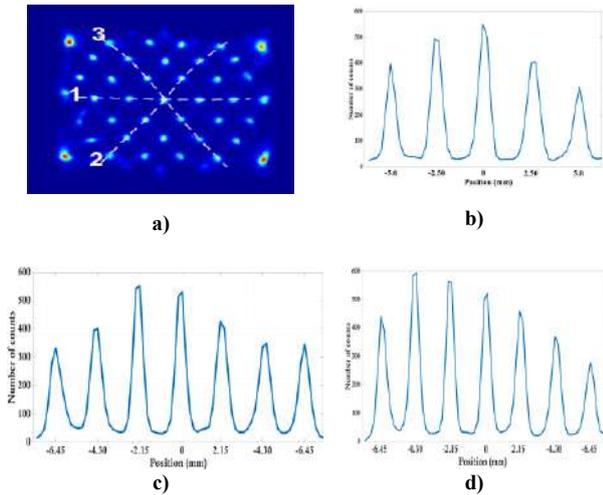


Figure 3. a) Field flood image in a crystal 7 mm thick. Figures b), c) and d) show profiles along diagonals 1, 2 and 3 respectively.

		Max Peak/Valley	Min Peak/Valley	Avg Peak/Valley
5 mm	Diagonal	18.7	5.0	13.7
	Row	23.0	5.7	10.9
7 mm	Diagonal	16.3	5.4	11.1
	Row	18.6	5.1	9.6
10 mm	Diagonal	14.9	5.6	9.8
	Row	16.7	5.4	8.1

Table 1. Peak to valley ratio for the three engraved crystals of 7mm, 7 mm and 10 mm.

3.2. Multiplexing

Uniform irradiation experiment was also conducted with a multiplexed readout scheme of four channels along two diagonals of the hexagon (positive and negative direction) as depicted in Figure 4 (right). Positioning is done with the common center of mass according to the energy sensed by each channel:

$$X = X_{BS} - X_{BD}$$

$$Y = Y_{RS} - Y_{RD}$$

$$E = (X_{BS} + X_{BD} + Y_{RS} + Y_{RD})/4$$

where the subindices stand blue solid, blue dashed and red solid and red dashed according to Figure 4 (right).

Figure 4 (left) is the 2-D position histograms computed as the center of mass with the pulse integral on a 5 mm thick crystal. No energy filter has been done. The 35 crystals are clearly distinguishable.

3.3. Energy resolution

The spectrum of the 10 mm thick crystal is shown in Figure 5. The FWHM of the photopeak is 72 keV which corresponds to 14%.

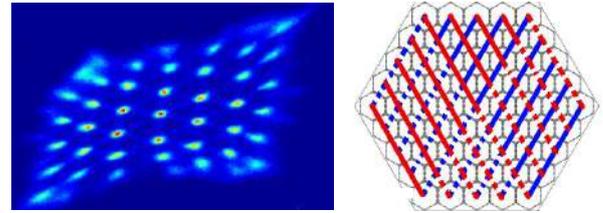


Figure 4. Left: field flood image of a 5 mm crystal with 4 channels multiplexing. Right: multiplexing encoding along the diagonals of the hexagon. Solid blue and red lines are two channels and dashed lines are the other two output channels.

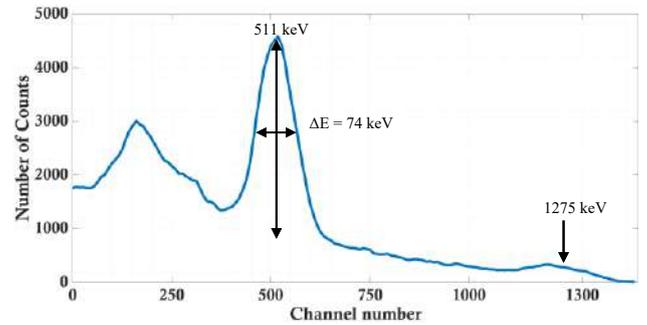


Figure 5. Energy Spectrum of the 10 mm crystal. The FWHM is 14%.

3.4. Depth of Interaction

A detector with the hexagonal pattern and depth of interaction capabilities with two shifted layers of 5 mm thick in a 10 mm thick LYSO square block of 15x15 mm of edge has been engraved. The engraving density is 60 μm in x,y and z direction. Figure 2 (center) is a sketch of the arrangement of the two layers with respect to the MPPC detector. shows the result of uniform irradiation experiment with the source in the upper face of the crystal. Zig zag line in (left) goes from crystals in the upper and lower layer. Profiles of this line is shown in (right) where the upper layer is plotted in solid blue line and lower layer in dashed black line.

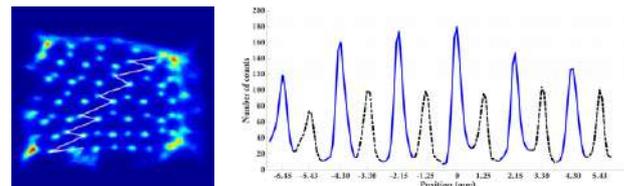


Figure 6. Left: field flood image. Right: profiles along the highlighted diagonal where solid blue lines are the upper layer and dashed black lines correspond to the lower layer.

3.5. Sensitivity

To prove whether the engrave affects the sensitivity, a collimated fan beam of pinhole of 1.8 mm of diameter scanned the surface of the 10 mm thick crystal before and after the engraving. Special attention was paid to the most unfavorable positions, like the corners of three hexagons. The count rate capability was constant and similar to the monolithic sample all around the detector with fluctuations within the statistical nature of the emission and detection.

4. Discussion and Conclusions

It has been proved possible the use of a cost-effective laser engraving machine to create optical barriers for the scintillation light of PET detectors. Extensive observation under the microscope of the morphology of the produced microcracks, has resulted in a fine calibration of the machine, both in terms of spatial precision and pulse energy to set the induced damage closely above the damage threshold. Maximum spatial engraving density with this equipment and LYSO samples has been found at 50 μm . Beyond this limit, uncontrolled cracking is common. For most brittle crystalline materials, crack formation and propagation correspond to repeated rupture of atomic bonds along specific crystallographic planes. Thermal shock is likely to be triggered by tensile stress due to rapid cooling of a brittle body that ultimately can lead to crack formation. Although LYSO is nowadays the most common scintillator due to its good light output properties, its monoclinic crystal structure, exhibits an anisotropy that causes unbalance tensile stress and contribute to crack formation.

The achieved spatial density is also given by the limited precision of the equipment. Furthermore, the presence of faults or previous microcracks in the crystal act as stress risers. The magnitude of this amplification is higher at small tip radius of curvature, so that it is intended that contiguous microcracks are as close as possible but not overlap. The presence of imperfection at the surface crystal might alter the beam shape and the focal spot.

In order to better control the focal spot and the absorption of the energy of the pulse, the inclusion of metal-type absorbers such a nanoparticles of gold is a common solution in laser defect-induced techniques. However, the crystallization method of scintillator makes it unfeasible. On the other hand, it could be considered the combination of ultrashort pulses in the range of picoseconds to induce multi-absorption processed and create small plasma balls of high absorption. A synchronized nanosecond pulse might follow to induce damage by thermal stress.

The area of the engraved hexagons is 2.6 times smaller than the area of the photocell. However, due to the excellent peak to valley ratios, the size of the hexagons can be further reduced. Ongoing experiments attempts a ratio of 6 by engraving triangles on the extent of a hexagonal photocell.

The multiplexing result proves the feasibility of the technique to be integrated in a complete scanner since it would allow to reduce the electronic complexity from 61 channels to 4 without a significant detriment in the crystal

identification. Thus, the number of analog inputs is reduced which ease the data transfer and the creation of bottlenecks in the system.

The count rate capability of the system has been proved not to be modified. Although it is believed that the fracture at the atomic level hampers the scintillation properties of the material, the engraving width is too small to be appreciated. In comparison to embedded crystal, it is also possible that detection probability is affected in lessen degree since engraving is not perfectly straight. Thus, gamma photons will likely always encounter a piece of crystal with unaltered scintillation properties.

Depth of interaction has been developed. Results were also confirmed by the collimated irradiation from one side. DOI development is critical to improve the spatial resolution of the system, especially at the border of the scanner. This allows more compact scanner designs which in turn might lead to cheaper systems by reduction of its transversal extent.

References

- [1] J. J. Vaquero, J. Seidel, S. Siegel, and M. V. Green, "Depth-encoding PET detector module with improved spatial sampling," *IEEE Nucl. Sci. Symp. Med. Imaging Conf.*, vol. 2, pp. 1255–1258, 1999.
- [2] J. J. Vaquero, J. J. Sánchez, E. Lage, J. M. Udías, P. Guerra, and M. Desco, "Design of DOI PET detector modules using phoswich and SiPMs: First results," *IEEE Nucl. Sci. Symp. Conf. Rec.*, pp. 3311–3313, 2012.
- [3] S. K. Jin *et al.*, "Performance measurement of the microPET focus 120 scanner," *J. Nucl. Med.*, vol. 48, no. 9, pp. 1527–1535, 2007.
- [4] T. Moriya, K. Fukumitsu, T. Yamashita, and M. Watanabe, "Fabrication of finely pitched LYSO arrays using sub-surface laser engraving technique with picosecond and nanosecond pulse lasers," *IEEE Nucl. Sci. Symp. Conf. Rec.*, vol. 61, no. 2, pp. 1032–1038, 2013.
- [5] H. Sabet, H. Kudrolli, B. Singh, and V. V. Nagarkar, "Fabricating high-resolution and high-sensitivity scintillator arrays using Laser Induced Optical Barriers," *IEEE Nucl. Sci. Symp. Conf. Rec.*, pp. 4080–4084, 2012.
- [6] W. C. J. Hunter, R. S. Miyaoka, L. MacDonald, W. McDougald, and T. K. Lewellen, "Light-sharing interface for dMiCE detectors using sub-surface laser engraving," *IEEE Trans. Nucl. Sci.*, vol. 62, no. 1, pp. 27–35, 2015.
- [7] D. Perez-Benito, R. Chil, J. M. Udías, M. Desco, and J. J. Vaquero, "SiPM-based PET detector module for a 4 π span scanner," *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.*, vol. 936, no. August 2018, pp. 18–21, 2019.

Methodologies for the quantification and classification of PET neuroimaging of patients with neurodegenerative diseases for the clinical decision support

A.P. Seiffert¹, A. Gómez-Grande^{2,3}, A. Villarejo-Galende³⁻⁶, E.J. Gómez^{1,7},
P. Sánchez-González^{1,7}

¹ Grupo de Bioingeniería y Telemedicina, ETSI Telecomunicación, Centro de Tecnología Biomédica, Universidad Politécnica de Madrid, Madrid, España; {ap.seiffert, enriquejavier.gomez, p.sanchez}@upm.es

² Servicio de Medicina Nuclear, Hospital Universitario 12 de Octubre, Madrid, España; adolfo.gomez@salud.madrid.org

³ Facultad de Medicina, Universidad Complutense de Madrid, Madrid, España; alberto.villarejo@salud.madrid.org

⁴ Servicio de Neurología, Hospital Universitario 12 de Octubre, Madrid, España

⁵ Grupo de Enfermedades Neurodegenerativas, Instituto de Investigación Sanitaria (i+12), Hospital Universitario 12 de Octubre, Madrid, España

⁶ Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Madrid, España

⁷ Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina, Madrid, España

Abstract

Neurodegenerative diseases are characterized by neuronal decline and death, and the resulting neurodegeneration results in reduced cerebral metabolism and perfusion. Additionally, they are characterized by anormal protein deposition like amyloid beta (A β) in Alzheimer's disease. Both can be visualized by PET imaging. In this work, methodologies for the quantification and classification of PET neuroimages for clinical decision support systems are designed and validated in different clinical problems related to neurodegenerative diseases. Specifically, these are the classification of A β PET images to determine the A β load, the validation of an alternative perfusion-like early-phase A β PET image technique, and the detection of altered metabolic patterns in [¹⁸F]fluorodeoxyglucose (FDG) PET images. The designed methodologies are validated in all three clinical problems and the results show the usefulness of quantitative PET neuroimage analysis.

1. Introduction

Neurodegenerative diseases are a group of diseases characterised by the progressive degeneration or death of neurons, leading to dementia, and are a major cause of mortality and morbidity. The most common cause of dementia is Alzheimer's disease (AD), which presents an increasing socio-economic, as well as a socio-sanitarian problem [1]. PET imaging is one of the main functional imaging tools for the diagnosis of AD. Different radiotracers allow the in vivo evaluation of neurodegeneration by studying the cerebral metabolism ([¹⁸F]fluorodeoxyglucose, FDG), and abnormal A β and tau deposition, defined as AD biomarkers [2].

On one hand, in neurodegenerative diseases, due to the degeneration or death of neurons, less glucose is consumed by the brain, which, consequently, reduces the [¹⁸F]FDG uptake during PET imaging. Dementias and other neurodegenerative diseases present characteristic patterns of reduced [¹⁸F]FDG uptake in PET images, mainly affecting

cortical brain regions, which allows for an improved differential diagnosis. On the other hand, A β -binding PET radiotracers were developed to be able to evaluate especially cortical A β burden [3].

In this work, a methodology for the processing and analysis of PET neuroimages used as AD imaging biomarkers, is designed. This methodology is validated by applying it to quantitatively analyze PET neuroimages to create new clinical decision support systems (CDSS) in current research questions. Concretely, it is applied (i) to validate the usefulness of an alternative quantitative image analysis technique for the classification of A β PET images, (ii) to quantitatively validate the perfusion-like information of the first-minute-frame (FMF) after ¹⁸F-labeled A β -binding radiotracer injection compared to [¹⁸F]FDG PET, (iii) to quantitatively describe metabolic uptake patterns in [¹⁸F]FDG PET. All image databases are provided by the Department of Nuclear Medicine of the Hospital Universitario 12 de Octubre, Madrid, Spain.

2. Methodology

The objective of the designed methodology is to quantitatively describe the radiotracer uptake in the PET neuroimages and employ these measures to draw clinically relevant conclusions and support the clinical decision-making based on the data. In broad terms, the methodology can be divided into three consecutive steps: (i) preprocessing, (ii) feature extraction, and (iii) statistical data analysis. The preprocessing is implemented in Statistical Parametric Mapping 12 (SPM12) [4], the feature extraction in MATLAB (The MathWorks Inc., Natick, MA, USA), and the statistical analysis in SPSS (IBM Corp., Armonk, NY, USA).

2.1. Preprocessing

The main objective of the preprocessing is to reduce interpatient anatomical variabilities and prepare the images

for the quantitative analysis. First, PET images are co-registered to its corresponding anatomical reference image. Then, spatial normalisation is performed to reduce the variability across the study cohort, allowing for direct and easier inter-subject comparisons. In spatial normalisation, the neuroimages are converted from native space to a standard space. Briefly, a template in standard space is aligned to an anatomical reference image using affine registration and non-linear deformations. The inverse deformation field can be then used on the co-registered functional images to spatially normalize them to the standard space.

2.2. Feature extraction

Quantitative image features are extracted from specific cortical and subcortical grey matter brain regions or volumes of interest (VOI). Brain atlases with predefined segmentations are employed, specifically the Automatic Anatomical Labelling (AAL2) atlas [5, 6].

For the analysis of PET neuroimages, the most used quantitative parameter is the standardised uptake value ratio (SUVR). The regional mean is normalized to the mean value of a reference region that is typically not affected by the disease, resulting in the SUVR.

Apart from the SUVR, the image texture can be quantified, measuring the spatial distribution of grey levels. Here, 6 texture features are extracted. These are computed from the from the grey-level co-occurrence matrix and are Energy, Contrast, Entropy, Homogeneity, Correlation, and Dissimilarity [7].

2.3. Statistical analysis

Statistical analyses are performed using the extracted features to draw clinically relevant conclusions and support the clinical decision-making. The main tests used in this work are Mann-Whitney U-tests, Kruskal-Wallis tests and correlation analyses calculating Pearson's correlation coefficient (r). In all analyses, a significance threshold of $\alpha = 0.05$ is defined.

3. Results and discussion

3.1. Texture analysis of A β PET

To evaluate cortical A β burden, one of the biomarkers of AD, A β PET images are usually interpreted visually in clinical practice. The main feature for the visual classification is the grey-to-white matter contrast. Cases of abnormally elevated cortical A β deposition are considered amyloid positive (A β +). The difference between the grey matter (GM) and white matter (WM) is difficult to identify and the image intensities are more homogeneous. Amyloid negative (A β -) cases are characterized by WM uptake without high GM uptake. WM can be clearly differentiated and the image intensity distribution between both tissues is heterogeneous [8].

The goal of this study is to evaluate the usefulness of textural image features for the classification of A β + and A β - scans, especially compared to the more common SUVR. Textural image analysis can describe the spatial distribution of grey levels in an image and may be able to quantitatively represent the grey-to-white matter contrast.

Patients who had undergone an A β PET scan are included in the retrospective study cohort. A β PET images were acquired either with [18 F]florbetapir (FBP) or [18 F]flutemetamol (FMM). A total of 66 patients (mean age, 66.11 ± 8.06 years; 34 men/32 women; 41 A β + / 25 A β -) are included.

Images are preprocessed following the methodology defined in section 2. A global VOI composed of the frontal, temporal, parietal and occipital lobes, the anterior and posterior cingulate cortices (ACC and PCC, respectively) and the precuneus is defined. SUVRs and 6 texture features are extracted. Differences of quantitative image features between positive and negative images are evaluated using Mann-Whitney U-tests and the area under the curve (AUC) of receiver operating characteristic curves of each feature is calculated.

When comparing image features between A β + and A β -, all features show statistically significant differences. Additionally, AUCs are over 0.7 except for Correlation, indicating acceptable discrimination. The results are summarized in Table 1.

Image Feature	p -Value	AUC
SUVR	< 0.001	0.758
Energy	< 0.001	0.785
Contrast	< 0.001	0.797
Entropy	< 0.001	0.819
Homogeneity	< 0.001	0.813
Correlation	0.037	0.654
Dissimilarity	< 0.001	0.806

Table 1. Differences of quantitative features between A β + and A β - PET images. AUC: Area under the curve; SUVR: Standardized Uptake Value Ratio.

Contrast, entropy, and dissimilarity are related to A β -, with higher values describing heterogeneous spatial intensity distributions that are correlated negatively to A β +. On the other hand, energy, homogeneity, and correlation are positively correlated to A β + and their values increase with higher homogeneity of the image intensity values. While non-SUVR, and specifically texture features, have previously been used to quantitatively analyze A β PET images [9], this is not the case to perform the classification task into A β + and A β -. The full results and regional, as well as radiotracer-specific analyses are published in [10].

3.2. The first-minute-frame (FMF) of A β PET as a surrogate for neuronal injury

Currently the most widely used radiotracer in clinical routine practice in the assessment of neurodegeneration is [18 F]FDG [3]. Together with A β PET, two imaging exams are required for two of the main AD biomarkers, i.e. neurodegeneration and A β deposition [11]. The dual-phase A β PET protocol acquires one perfusion-like early-phase image and the standard late-phase A β image using only one radiotracer injection and imaging exam [12]. However, the early-phase time windows are radiotracer-dependent and usually dynamic scans compared to a faster and simpler static scan are acquired.

The aim of this study is to quantitatively evaluate the perfusion-like information of static brain A β PET images obtained in the first minute after radiotracer injection, hereafter called FMFs (first-minute-frame), with ^{18}F -labeled A β -binding radiotracers. These proposed images are compared to the cerebral metabolism as measured in [^{18}F]FDG PET images.

A retrospective study cohort is defined including all patients transferred for suspected neurocognitive disease between February 2016 and July 2019 for A β and [^{18}F]FDG PET imaging. A total of 60 patients (mean age, 66.27 ± 8.26 years; 29 men/31 women; 27 [^{18}F]FBP/27 [^{18}F]FMM/6 [^{18}F]Florbetaben) with available FMF and [^{18}F]FDG PET scans comprise the final retrospective study cohort.

Images are preprocessed using SPM12 as described in section 2. Regional SUVRs are calculated to quantify radiotracer uptake. The correlation of regional SUVRs of FMFs and [^{18}F]FDG PET images is calculated to quantitatively analyze the similarity between the images.

The intensity profiles of the FMF and the [^{18}F]FDG PET images are shown in Figure 1. These are computed averaging the regional SUVRs among the study cohort. Both intensity profiles are also strongly correlated ($r = 0.98$).

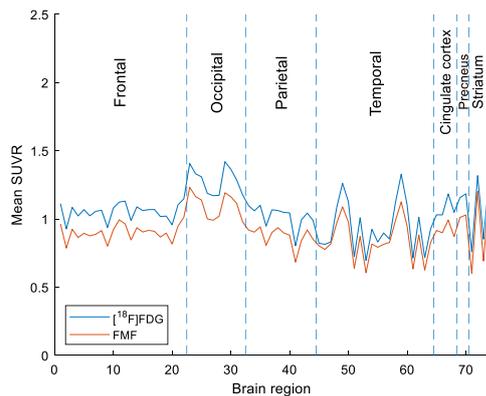


Figure 1. Intensity profiles of FMF and [^{18}F]fluorodeoxyglucose (FDG) PET.

The interpatient correlation is calculated of the SUVRs of each region. The results are summarized in Table 2. Overall, high correlations are shown in all regions, especially in the striatum and the temporal VOI.

Brain region	r (left/right)
Frontal	0.83/0.76
Occipital	0.71/0.73
Parietal	0.78/0.80
Temporal	0.88/0.88
ACC	0.86/0.83
PCC	0.81/0.86
Precuneus	0.79/0.79
Striatum	0.92/0.92

Table 2. Regional interpatient correlation coefficients. ACC: Anterior Cingulate Cortex; PCC: Posterior Cingulate Cortex.

In addition to evaluate the correlation among all patients for each region, inpatient correlation coefficients based on the regional SUVRs are calculated. These are also high,

resulting in a mean (\pm standard deviation) correlation coefficient of $r = 0.93 \pm 0.05$ for the study cohort.

The results show quantitative similarities of the FMF to [^{18}F]FDG brain PET images. Strong correlations to [^{18}F]FDG PET images are revealed using a common early-phase time window for the three ^{18}F -labeled A β -binding radiotracers to acquire a static perfusion-like PET image. The results are also comparable or superior to previous studies [13, 14]. These and additional results corresponding to A β status and radiotracer-specific subcohorts are published [15].

3.3. Pattern detection in [^{18}F]FDG PET

As described in the introduction, neurodegenerative diseases present characteristic patterns of alterations in [^{18}F]FDG PET images and other metabolism or perfusion images compared to the healthy brain. These affect mainly cortical regions, but may also be present in subcortical grey matter structures like the striatum.

While in the two previous sections the designed methodology is applied to quantitatively describe A β PET images, in this section [^{18}F]FDG PET images are processed and analyzed. It is studied if the quantitative parameters can describe cerebral metabolic patterns in primary progressive aphasia (PPA), and Parkinson's disease (PD) and Parkinsonisms.

Primary progressive aphasia

PPA is a neurological syndrome characterized by the degeneration of the language-related brain regions, and therefore causing an impairment of language faculties. PPA symptoms may differ depending on the affected brain regions and the most common classification is defined by Gorno-Tempini et al. [16]: (i) non-fluent/agrammatic variant PPA (nfvPPA), (ii) semantic variant PPA (svPPA), (iii) logopenic variant PPA (lvPPA), and (iv) unclassifiable PPA (ucPPA).

A database of 19 patients is used to quantitatively detect the metabolic patterns of the variants of PPA. Given that only 1 case with svPPA was available, patients are grouped into nfvPPA, lvPPA, and other PPAs (svPPA and ncPPA). Due to possible overlaps in affected brain areas (e.g. anterior temporal vs. posterior temporal lobe), 94 VOIs directly defined by the AAL2 atlas are used to calculate SUVRs.

Observing the distribution of regional SUVRs in each variant, regions of reduced values, i.e. hypometabolism, can be identified. In both the nfvPPA and lvPPA groups, regions corresponding to the temporal and parietal lobes show low SUVRs. In the nfvPPA group, the frontal lobe also presents lowered SUVRs. In the last group, the SUVRs are more disperse due to the heterogeneity of the group, however, mainly the temporal and parietal lobes are affected. All these results correspond to the typical clinical presentations of the variants of PPA.

Parkinson's disease and Parkinsonisms

Lastly, both PD and Parkinsonisms, neurodegenerative disorders that share clinical symptoms with PD, are characterized by alterations of the striatal synaptic activity, which

are also visible in [^{18}F]FDG PET images. A retrospective study cohort of 3 PD patients, 29 patients with a Parkinsonism syndrome and 25 subjects with normal striatal metabolic activity was selected.

Kruskal-Wallis tests revealed that all 6 VOIs (caudate, putamen, and striatum) except the right caudate show significantly higher activity in PD than Parkinsonisms, as well as in the controls than Parkinsonisms ($p < 0.03$). Moreover, caudate-putamen and hemispherical differences characteristic of specific Parkinsonisms are found in the corresponding subgroups. Lastly, a software application (Figure 2) is developed to easily visualize the striatal uptake (red points in the graph) compared to the control group (blue points in the graph).

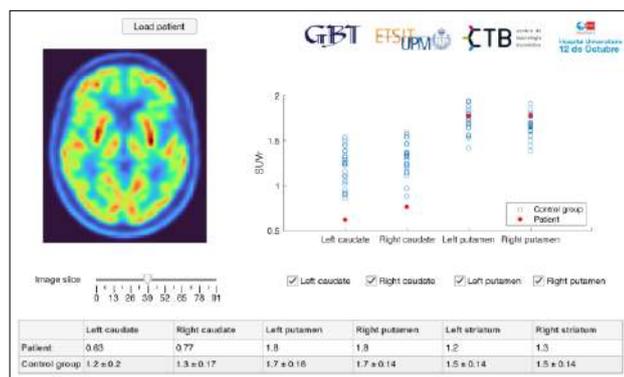


Figure 2. Application to visualize striatal [^{18}F]FDG uptake.

4. Conclusions

In this work, a methodology for the quantification and classification of PET neuroimages is designed. The methodology is validated by applying it to three clinical problems. In all cases, the objectives are fulfilled, and it is shown that using the designed methodology, CDSS can be created for clinical use.

Acknowledgements

This work is partly supported by an FPU fellowship (*Beca de Formación de Profesorado Universitario*) from the Spanish Ministry of Science, Innovation and Universities (FPU16/06487).

References

- [1] Hickman RA *et al.* Alzheimer Disease and Its Growing Epidemic: Risk Factors, Biomarkers, and the Urgent Need for Therapeutics. *Neurologic Clinics*, vol 34, no 4. W.B. Saunders, pp 941–953, 01-Nov-2016.
- [2] La Joie R *et al.* Multisite study of the relationships between antemortem [^{11}C]PIB-PET Centiloid values and postmortem measures of Alzheimer's disease neuropathology. *Alzheimer's and Dementia*, vol 15, no 2, 2019, pp 205–216, (ISSN: 15525279).
- [3] Tiepolt S *et al.* Current radiotracers to image neurodegenerative diseases. *EJNMMI Radiopharmacy and Chemistry*, vol 4, no 1, 2019, p 17, (ISSN: 2365-421X).
- [4] Friston KJ *et al.* Statistical parametric mapping: the analysis of functional brain images. Elsevier/Academic Press, 2006, (ISBN: 9780080466507).
- [5] Tzourio-Mazoyer N *et al.* Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, vol 15, no 1, 2002, pp 273–289, (ISSN: 10538119).
- [6] Rolls ET *et al.* Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *NeuroImage*, vol 122, 2015, pp 1–5, (ISSN: 10959572).
- [7] Haralick RM *et al.* Textural Features for Image Classification. *IEEE Transactions on Systems, Man and Cybernetics*, vol SMC-3, no 6, 1973, pp 610–621, (ISSN: 21682909).
- [8] Minoshima S *et al.* SNMMI procedure standard/EANM practice guideline for amyloid PET imaging of the brain 1.0. *Journal of Nuclear Medicine*, vol 57, no 8, 2016, pp 1316–1322, (ISSN: 2159662X).
- [9] Campbell DL *et al.* Application of Haralick texture features in brain [^{18}F]florbetapir positron emission tomography without reference region normalization. *Clinical Interventions in Aging*, vol 12, 2017, pp 2077–2086, (ISSN: 11781998).
- [10] Seiffert AP *et al.* Texture-Based Analysis of ^{18}F -Labeled Amyloid PET Brain Images. *Applied Sciences*, vol 11, no 5, 2021, p 1991.
- [11] Jack CR *et al.* NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's and Dementia*, vol 14, no 4, 2018, pp 535–562, (ISSN: 15525279).
- [12] Valentina G *et al.* Dual-phase amyloid PET: hitting two birds with one stone. *European Journal of Nuclear Medicine and Molecular Imaging*, vol 43, no 7, 2016, pp 1300–1303, (ISSN: 16197089).
- [13] Hsiao IT *et al.* Correlation of early-phase ^{18}F -florbetapir (AV-45/Amyvid) PET images to FDG images: preliminary studies. *European journal of nuclear medicine and molecular imaging*, vol 39, no 4, 2012, pp 613–20, (ISSN: 1619-7089).
- [14] Schmitt J *et al.* Dual-Phase β -Amyloid PET Captures Neuronal Injury and Amyloidosis in Corticobasal Syndrome. *Frontiers in Aging Neuroscience*, vol 13, 2021, (ISSN: 16634365).
- [15] Seiffert AP *et al.* High Correlation of Static First-Minute-Frame (FMF) PET Imaging after ^{18}F -Labeled Amyloid Tracer Injection with [^{18}F]FDG PET Imaging. *Sensors*, vol 21, no 15, 2021, p 5182.
- [16] Gorno-Tempini ML *et al.* Classification of primary progressive aphasia and its variants. *Neurology*, vol 76, no 11, 2011, pp 1006–1014, (ISSN: 1526632X).

La Importancia de la Aurícula Izquierda en la Alteración del Sustrato Auricular tras Ablación por Catéter de Fibrilación Auricular Paroxística

A. Vraka¹, A. Quesada Dorador², F. Hornero Sos³,
V. Bertomeu-González⁴, R. Alcaraz Martínez⁵, J.J. Rieta Ibáñez¹

¹ BioMIT.org, Departamento de Ingeniería Electrónica, Universitat Politècnica de Valencia, España, {aivra, jjrieta}@upv.es

² Servicio de Cardiología, Hospital General Universitario de Valencia, España, quesada_aur@gva.es

³ Servicio de Cirugía Cardiovascular, Hospital Clínico Universitario de Valencia, España, hornero_fer@gva.es

⁴ Departamento de Medicina Clínica, Universidad Miguel Hernández, España, vbortomeu@umh.es

⁵ Grupo de Inv. en Electrónica, Telecom. y Bioingeniería, Univ. de Castilla-La Mancha, España, raul.alcaraz@uclm.es

Resumen

El efecto de la ablación por catéter (AC) de la fibrilación auricular (FA) puede ser diferente en la aurícula derecha (AD) que en la izquierda (AI). Aunque las ondas P contienen información que corresponde tanto a la AD como a la AI por separado, los estudios insisten en analizar la duración de onda P (DOP) de manera indivisible. Se utilizaron registros de derivación II obtenidos durante cinco minutos antes y después de la AC de pacientes con FA paroxística. Las características calculadas fueron: DOP, duración de primera y segunda parte de la onda P (DOP_1 y DOP_2), reflejando la despolarización de la AD y AI, respectivamente y el intervalo P-R. Las variaciones de la frecuencia cardíaca se controlaron mediante corrección. Los resultados pre- y postablativos se compararon con la prueba U de Mann-Whitney. Se calcularon los valores de la mediana y el porcentaje de la variación provocada por la AC y se compararon entre el DOP y cada una de las características restantes con la correlación de Pearson. La DOP y la DOP_2 disminuyeron significativamente después de la AC, independientemente de la corrección (DOP : $\leq -17,96\%$, $p < 0,0442$, DOP_2 : $\leq -27,77\%$, $p < 0,0268$). El acortamiento de DOP estaba más correlacionado con lo de DOP_2 antes de la corrección (61,4%, $p = 0,0031$). Después de la corrección, el acortamiento de la DOP se correlacionó mejor con la DOP_1 (76,8%, $p < 0,0001$). Aunque la DOP se aproxima más al tiempo de despolarización de la AD, la alteración de la onda P es provocada por el acortamiento de la AI. Por tanto, los estudios enfocados a la alteración del sustrato auricular deberían centrarse en las alteraciones de la AI.

1. Introducción

La fibrilación auricular (FA) es la arritmia cardíaca más frecuente y la ablación por catéter (AC) se considera su tratamiento estrella [1, 2]. Los procedimientos de AC consisten en el aislamiento de la fuente principal de FA, que se cree que son las venas pulmonares (VPs) [3]. Sin embargo, en muchos casos, la ablación de sitios adicionales que presentan actividad de FA se lleva a cabo para lograr su terminación [4].

Durante la AC, se obtienen los registros del electrocardiograma (ECG) y se analizan posteriormente para evaluar el sustrato auricular y predecir el resultado de la AC [5–9].

Las ondas P son de gran interés en los estudios de la FA, ya que representan la despolarización de las aurículas [10]. La duración de las ondas P es el índice más popular, ampliamente utilizado para este fin. Las ondas P largas o cortas se han asociado a la remodelación eléctrica de las aurículas y a la recurrencia de la FA [5–7]. Por lo tanto, el acortamiento de las ondas P es un efecto deseable que se cree que predice un resultado satisfactorio de la AC [8, 9].

La onda P puede revelar información sobre la despolarización de la aurícula derecha (AD) y de la aurícula izquierda (AI) correspondiente a su primera y segunda parte, respectivamente [11, 12]. Aunque se ha comprobado que ambas aurículas forman el sustrato auricular [13–16], su respuesta al procedimiento de AC puede que sea diferente. No obstante, los análisis se centran en el estudio de las modificaciones en su totalidad, ignorando la posibilidad de diferente grado de alteración entre la AD y la AI tras la AC. El objetivo del presente estudio es centrarse en esta cuestión y ofrecer una visión más específica de la alteración del sustrato auricular, aprovechando los componentes separados de la onda P correspondientes a la AD y la AI.

2. Materiales

Se emplearon los registros de la derivación II de 29 pacientes con FA paroxística sometidos por primera vez a AC mediante radiofrecuencia (RF). Los registros se adquirieron durante cinco minutos antes y después de la AC con una frecuencia de muestreo de 1 kHz.

3. Métodos

3.1. Preprocesado

El preprocesamiento comenzó con la eliminación de ruido. Se adoptó un método basado en ondas para eliminar las interferencias de las líneas eléctricas, seguido de la eliminación del ruido muscular y de la fluctuación de la línea base mediante filtros bidireccionales de paso bajo y paso

Tabla 1. Correlación entre DOP y las demás características (antes y después de la AC) y correlación del efecto de la AC ente DOP y las demás características. Las correlaciones estadísticamente significativas se muestran en color gris.

Feature	Antes de la AC		Después de la AC		Efecto de la AC	
	CP	valor p	PCC	valor p	CP	valor p
DOP_1	74,7 %	< 0,0001	74,5 %	< 0,0001	54,0 %	0,0114
DOP_2	47,7 %	< 0,0001	41,9 %	< 0,0001	61,4 %	0,0031
$P_{on} - R$	77,2 %	< 0,0001	75,2 %	< 0,0001	24,0 %	0,2942
$P_{off} - R$	-50,3 %	< 0,0001	-23,2 %	< 0,0001	-40,1 %	0,0716
$C(DOP_1)$	80,5 %	< 0,0001	80,2 %	< 0,0001	76,8 %	< 0,0001
$C(DOP_2)$	54,1 %	< 0,0001	53,1 %	< 0,0001	71,0 %	0,0003
$C(P_{on} - R)$	83,3 %	< 0,0001	85,5 %	< 0,0001	78,5 %	< 0,0001
$C(P_{off} - R)$	9,9 %	< 0,0001	26,9 %	< 0,0001	32,5 %	0,1508

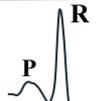
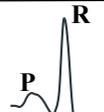
DOP	Duración de la onda P; despolarización de toda la aurícula.	
DOP1	Duración de la primera parte de la onda P; despolarización de la AD.	
DOP2	Duración de la segunda parte de la onda P; despolarización de la AI.	
Pon-R	Intervalo P-R. Despolarización de aurículas hasta unión Purkinje-músculo	
Poff-R	Intervalo desde fin de despolarización de AI hasta unión Purkinje-músculo	

Figura 1. Las características calculadas. Las líneas rojas corresponden al intervalo que se calcula en los ECGs.

alto, respectivamente [17, 18]. Finalmente, la cancelación de los latidos ectópicos se realizó mediante interpolación lineal [19]. La presencia de latidos ectópicos fue inferior al 4 % del total de latidos de cada registro y no todos los registros mostraron presencia ectópica. La detección de las ondas P y R, así como la delineación de las ondas P, siguieron el preprocesamiento [20, 21].

3.2. Análisis principal

A continuación, se calcularon las características de duración íntegra y parcial de la onda P (DOP , DOP_1 , DOP_2) y distancia entre la onda P y el pico R, como se describe en la figura 1.

Dado que los componentes temporales del ECG se ven influidos por las fluctuaciones del ritmo cardiaco (RC) observadas a lo largo de la AC [22], todas las características se multiplicaron por un factor de corrección (FC)

$$C(característica_i) = \frac{1000}{R_i - R_{i-1}}, \quad (1)$$

donde $R_i - R_{i-1}$ es el intervalo de dos picos R consecutivos dentro de los cuales se encuentra la activación i .

3.3. Análisis estadístico

Para el análisis estadístico, en primer lugar, se realizó la prueba de normalidad de Kolmogorov-Smirnov y, según los resultados, se compararon las características antes y después de la AC con una prueba U de Maan-Whitney (UMW). A continuación, se calculó la correlación entre la DOP y las demás características mediante la correlación de Pearson (CP). Por último, se utilizó adicionalmente la CP para definir si el efecto de la AC sobre la DOP estaba relacionado con el efecto de la AC sobre alguna de las demás características. Esto se realizó correlacionando el porcentaje de alteración tras la AC de cada rasgo con el porcentaje de alteración de la DOP.

4. Resultados

La figura 2 muestra los valores de la mediana de las características antes y después de AC. También puede apreciarse el porcentaje de variación. La DOP disminuyó significativamente con la AC (-9,84 %). Las DOP_1 y DOP_2 también se redujeron. No obstante, el acortamiento que mostró DOP_1 no fue significativo. Por otro lado, la reducción de la DOP_2 después de la AC fue notablemente mayor (-22,03 %), mostrando además significación estadística.

La corrección no alteró las tendencias mencionadas, que siguieron mostrando un acortamiento estadístico para DOP y DOP_2 pero no para DOP_1 . En cuanto a los intervalos P-R, ninguna de las características calculadas varió significativamente después de la AC.

Las correlaciones entre DOP y el resto de características se muestran en la tabla 1. Para las correlaciones medidas antes y después de AC, se observan correlaciones de moderadas a altas con significación estadística en todas las características. $P_{off} - R$ y $C(P_{off} - R)$ fueron las características que mostraron las correlaciones más débiles (-50,3 % a -23,2 % y 9,9 % a 26,9 %, respectivamente, $p < 0,0001$ en todos los casos). Por el contrario, $P_{on} - R$ y $C(P_{on} - R)$ mostraron las mayores correlaciones (75,2 % a 77,2 % y 83,3 % a 85,5 %, respectivamente, $p < 0,0001$ en todos los casos). Para el análisis de la onda P parcial, DOP_1 se correlacionó notablemente más que DOP_2 independientemente de la corrección (hasta 80,5 % para DOP_1 y hasta 53,1 % para DOP_2 , $p < 0,0001$ para ambos).

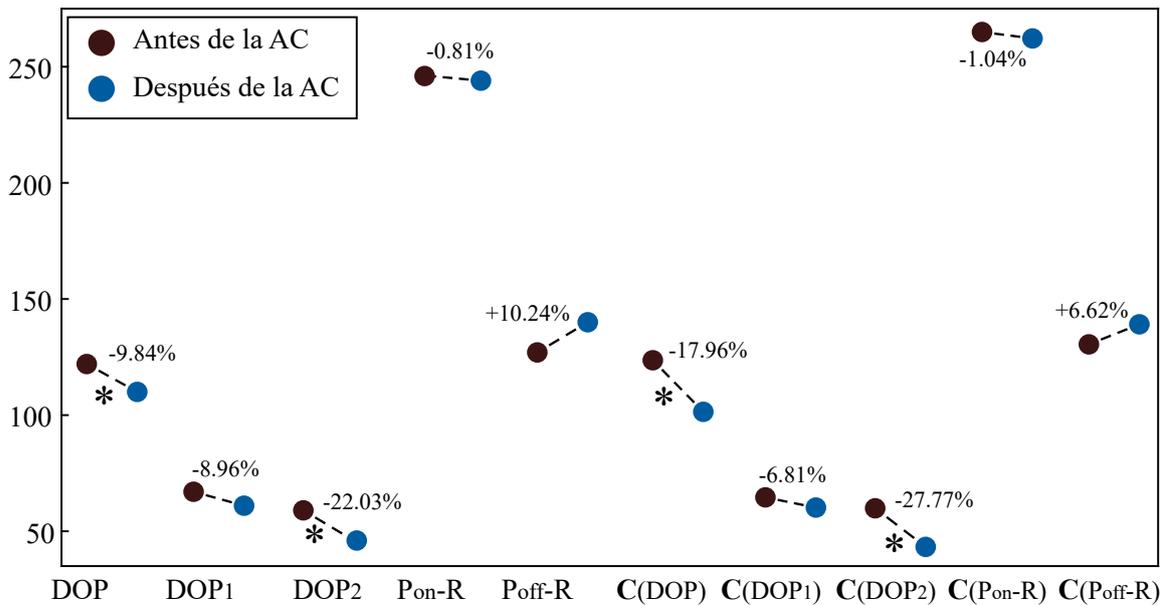


Figura 2. Valores medianos de las características antes (marrón) y después (azul) de la AC. Se ilustra también el porcentaje de variación y variaciones estadísticamente significativas están marcadas con un asterisco (*).

Por último, la correlación entre la DOP y las restantes características disminuyó tanto en magnitud como en potencia estadística cuando se examinó el efecto de la AC. La correlación con $P_{on} - R$ resultó ser baja y no significativa (24,0%, $p = 0,2942$) antes de la corrección, mientras que tras la aplicación del FC, volvió a ser alta y significativa (78,5%, $p < 0,0001$). La correlación con DOP_2 fue mayor que la correlación con DOP_1 antes de la corrección (61,4%, $p = 0,0031$ y 54,0%, $p = 0,0114$, respectivamente). Sin embargo, el efecto de la corrección hizo que la correlación con DOP_2 volviera a ser ligeramente inferior a la correlación con DOP_1 (71,0%, $p = 0,0003$ y 76,8%, $p < 0,0001$, respectivamente).

5. Discusión

El análisis de la DOP es de indudable importancia para la evaluación del sustrato auricular, ampliamente utilizado para predecir la recurrencia de la FA [5–9]. Aunque los componentes de la DOP pueden analizarse más a fondo para obtener una idea más específica de los mecanismos de la FA y de cómo afectan a cada aurícula, tienden a omitirse.

En trabajos anteriores, las características medidas a partir de los registros de AD y AI se han utilizado para detectar signos de fibrosis en la aurícula correspondiente [13–16]. Las ondas P también se han correlacionado con características que indican remodelación estructural en la AI [23]. Sin embargo, la contribución de cada aurícula, medida a partir de los componentes de las ondas P, a la alteración del sustrato observada después de la AC está aún por descubrir. Este ha sido el objetivo principal del presente estudio.

Se comprobó que el acortamiento de la onda P era diferente en la AD que en la AI. Más concretamente, fue la duración de la onda P medida a partir de la AI la que mostró una notable reducción, independientemente de la corrección. Esta reducción fue significativa tanto en magnitud como en potencia estadística. Por el contrario, aunque la parte de la

onda P que corresponde a la AD sí mostró tendencia de disminución, ésta fue menor. La proximidad de la AI a las VPs puede ser la razón de esta diferencia en el efecto de la AC.

La correlación entre la DOP y las restantes características reveló dos puntos principales: 1. El intervalo $P_{on} - R$ es la característica más relacionada con la DOP y 2. DOP_1 está correlacionada en mayor grado con la DOP que DOP_2 . Teniendo en cuenta que $P_{on} - R$ implica realmente la medición de la DOP, el primero coincide con observaciones anteriores [24]. El estudio de las alteraciones del intervalo P-R, en consecuencia, se ve muy afectado desde el componente auricular. Por ello, se anima a los estudios a realizar cálculos sin incluir la parte de la onda P para evaluar la alteración del tiempo que necesita la onda para propagarse desde las aurículas a los ventrículos.

En cuanto al segundo supuesto, en principio, el tiempo de despolarización de la AD es mayor que el de la AI y, por tanto, contribuye en mayor medida a la medición de la DOP. Además, el nodo sinusal, cuyo correcto funcionamiento es crucial para el mantenimiento del ritmo sinusal, se encuentra en la AD. Estas observaciones pueden explicar este hallazgo. No obstante, el efecto de la AC se ilustra mejor y más claramente a partir de la AI, que parece provocar el famoso acortamiento de la DOP.

6. Conclusiones

El análisis parcial de las ondas P ofrece una perspectiva más detallada del estado auricular tras la AC. La AI se ve afectada en mayor medida por la AC y la alteración de las ondas P procede principalmente de la alteración de la parte de la onda P que corresponde a la AI. En consecuencia, se debe priorizar el análisis de la AI durante y después de las sesiones de AC. El intervalo P-R depende en gran medida de la despolarización auricular. Por lo tanto, para estudiar el tiempo de propagación de las aurículas a los ventrículos

los, es necesario realizar mediciones sin tener en cuenta el componente de la onda P.

Agradecimientos

Trabajo cofinanciado por los proyectos de investigación competitiva DPI2017-83952-C3 de MINECO-AEI-FEDER-UE, SBPLY/17/180501/000411 de la JCCLM y AICO/2021/286 de la GVA.

Referencias

- [1] Oral H, Ozaydin M, Chugh A, Scharf C, Tada H, Hall B, Cheung P, Pelosi F, Knight BP, Morady F. Role of the coronary sinus in maintenance of atrial fibrillation. *Journal of cardiovascular electrophysiology* Dec 2003;14:1329–36.
- [2] Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, et al. 2020 esc guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the european association for cardiothoracic surgery (eacts): The task force for the diagnosis and management of atrial fibrillation of the european society of cardiology (esc) developed with the special contribution of the european heart rhythm association (ehra) of the esc. *European heart journal* February 2021;42:373–498. ISSN 1522-9645.
- [3] Haissaguerre M, Jaïs P, Shah DC, Takahashi A, Hocini M, Quiniou G, Garrigue S, Le Mouroux A, Le Métayer P, Clémenty J. Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins. *New England Journal of Medicine* 1998;339(10):659–666.
- [4] Santangeli P, Marchlinski FE. Techniques for the provocation, localization, and ablation of non-pulmonary vein triggers for atrial fibrillation. *Heart Rhythm* 2017;14(7):1087–1096.
- [5] Cheng S, Keyes MJ, Larson MG, McCabe EL, Newton-Cheh C, Levy D, Benjamin EJ, Vasan RS, Wang TJ. Long-term outcomes in individuals with prolonged pr interval or first-degree atrioventricular block. *Jama* 2009;301(24):2571–2577.
- [6] Chen Q, Mohanty S, Trivedi C, Gianni C, Della Rocca DG, Canpolat U, et al. Association between prolonged p wave duration and left atrial scarring in patients with paroxysmal atrial fibrillation. *Journal of cardiovascular electrophysiology* October 2019;30:1811–1818. ISSN 1540-8167.
- [7] Auricchio A, Özkartal T, Salghetti F, Neumann L, Pezzuto S, Gharaviri A, Demarchi A, Caputo ML, Regoli F, De Asmundis C, Chierchia GB, Brugada P, Klersy C, Moccetti T, Schotten U, Conte G. Short p-wave duration is a marker of higher rate of atrial fibrillation recurrences after pulmonary vein isolation: New insights into the pathophysiological mechanisms through computer simulations. *Journal of the American Heart Association* January 2021;10:e018572. ISSN 2047-9980.
- [8] Maan A, Mansour M, Ruskin JN, Heist EK. Impact of catheter ablation on p-wave parameters on 12-lead electrocardiogram in patients with atrial fibrillation. *Journal of electrocardiology* 2014;47:725–733. ISSN 1532-8430.
- [9] Kizilirmak F, Demir GG, Gokdeniz T, Gunes HM, Cakal B, Guler E, Karaca bO, Omaygenç MO, Yılmaz F, Olgun FE, Kilicaslan F. Changes in electrocardiographic p wave parameters after cryoballoon ablation and their association with atrial fibrillation recurrence. *Annals of noninvasive electrocardiology the official journal of the International Society for Holter and Noninvasive Electrocardiology Inc* November 2016;21:580–587. ISSN 1542-474X.
- [10] Simpson RJ, Foster JR, Gettes LS. Atrial excitability and conduction in patients with interatrial conduction defects. *The American journal of cardiology* December 1982;50:1331–1337. ISSN 0002-9149.
- [11] Baranchuk A, de Luna AB. The p-wave morphology: what does it tell us? *Herzschrittmachertherapie Elektrophysiologie* 2015;26(3):192–199.
- [12] Van Steenkiste G, Vera L, Decloedt A, Schauvliege S, Boussy T, van Loon G. Endocardial electro-anatomic mapping in healthy horses: Normal sinus impulse propagation in the left and right atrium and the ventricles. *Veterinary journal London England* 1997 April 2020;258:105452. ISSN 1532-2971.
- [13] de Groot NM, Allesie MA. Pathophysiology of atrial fibrillation: focal patterns of activation. *Pacing and Clinical Electrophysiology* 2019;42(10):1312–1319.
- [14] Jadidi A, Nothstein M, Chen J, Lehrmann H, Dössel O, Allgeier J, Trenk D, Neumann FJ, Loewe A, Müller-Edenborn B, Arentz T. Specific electrogram characteristics identify the extra-pulmonary vein arrhythmogenic sources of persistent atrial fibrillation – characterization of the arrhythmogenic electrogram patterns during atrial fibrillation and sinus rhythm. *Sci Rep* 2020;10. ISSN 2045-2322.
- [15] Kharbanda RK, Knops P, van der Does LJME, Kik C, Taverne YJHJ, Roos-Serote MC, Heida A, Oei FBS, Bogers AJJC, de Groot NMS. Simultaneous endo-epicardial mapping of the human right atrium: Unraveling atrial excitation. *Journal of the American Heart Association* 2020;9. ISSN 2047-9980.
- [16] Li CY, Zhang JR, Hu WN, Li SN. Atrial fibrosis underlying atrial fibrillation. *International Journal of Molecular Medicine* 2021;47(3):1–1.
- [17] Sörnmo L, Laguna P. Ecg signal processing. In *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. Academic Press, 2005; 453–566.
- [18] García M, Martínez-Iniesta M, Ródenas J, Rieta JJ, Alcaraz R. A novel wavelet-based filtering strategy to remove powerline interference from electrocardiograms with atrial fibrillation. *Physiological measurement* November 2018;39:115006. ISSN 1361-6579.
- [19] Martínez A, Alcaraz R, Rieta JJ. Detection and removal of ventricular ectopic beats in atrial fibrillation recordings via principal component analysis. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual International Conference* 2011;2011:4693–4696. ISSN 2694-0604.
- [20] Martínez A, Alcaraz R, Rieta JJ. A new method for automatic delineation of ecg fiducial points based on the phasor transform. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual International Conference* 2010;2010:4586–4589. ISSN 2375-7477.
- [21] González F, Alcaraz R, Rieta JJ. Electrocardiographic p-wave delineation based on adaptive slope gaussian detection. In *Computing in Cardiology, CinC 2007, Rennes, France, September 24-27, 2017*. www.cinc.org, 2017; 1–4.
- [22] Toman O, Hnatkova K, Smetana P, Huster KM, Šišáková M, Barthel P, Novotný T, Schmidt G, Malik M. Physiologic heart rate dependency of the pq interval and its sex differences. *Scientific reports* February 2020;10:2551. ISSN 2045-2322.
- [23] Kasser I, Kennedy JW. The relationship of increased left atrial volume and pressure to abnormal p waves on the electrocardiogram. *Circulation* March 1969;39:339–343. ISSN 0009-7322.
- [24] Park J, Kim TH, Lee JS, Park JK, Uhm JS, Joung B, Lee MH, Pak HN. Prolonged pr interval predicts clinical recurrence of atrial fibrillation after catheter ablation. *Journal of the American Heart Association* October 2014;3:e001277. ISSN 2047-9980.

Optimización preoperatoria de la configuración de los electrodos de estimulación con un modelo 3D de médula espinal personalizado al paciente

C. Solanes¹, J. L. Durá¹, M. A. Canós², J. De Andrés³, L. Martí-Bonmatí⁴, J. Saiz¹

¹ Ci2B, Universitat Politècnica de València, Valencia, España, carsogal@etsii.upv.es

² Unidad del Dolor, Hospital Universitari i Politècnic La Fe, Valencia, España, canos_marver@gva.es

³ SART, Hospital General de Valencia, Valencia, España, deandres_jos@gva.es

⁴ GIBI230-PREBI, Hospital Universitari i Politècnic La Fe - IIS La Fe, Valencia, España, marti_lui@gva.es

Resumen

La terapia de Estimulación de Médula Espinal (EME) es una técnica mínimamente invasiva que se utiliza para tratar el dolor crónico. El desarrollo y uso de modelos computacionales de EME ha permitido entender el efecto de los parámetros de estimulación en la activación neuronal para producir una adecuada cobertura de parestesia al paciente. Sin embargo, todavía no se ha considerado el uso de modelos de EME en la práctica clínica. Con el objetivo de estudiar la capacidad de un modelo de EME como herramienta de trabajo en un escenario clínico, en este trabajo se ha diseñado una aplicación clínica, que consiste en la optimización de la configuración de los electrodos antes del implante de los componentes de EME. La metodología empleada consiste en el uso de un modelo de médula espinal específico de un paciente desarrollado en COMSOL Multiphysics® combinado con un modelo de fibra nerviosa mielínica sensitiva implementado en MATLAB® R2017a. Los resultados muestran que, de las seis configuraciones de electrodos consideradas, la configuración 3 es la que permitiría optimizar la estimulación para solapar la cobertura de parestesia en los dermatomas dolorosos del paciente. Por tanto, el uso de modelos de EME específicos del paciente podría llegar a ser una futura herramienta clave para el médico en la práctica clínica para optimizar la terapia de EME.

1. Introducción

La terapia de Estimulación de Médula Espinal (EME) es una técnica ampliamente utilizada para el tratamiento del dolor crónico [1]. Esta técnica consiste en estimular eléctricamente las fibras nerviosas de la médula espinal, con el objetivo de producir una sensación de hormigueo (parestesia) en los dermatomas donde el paciente siente el dolor. De acuerdo con la teoría de la puerta del control [2], el objetivo de esta terapia es activar las fibras responsables de la parestesia, que son las fibras sensitivas Aβ grandes, las cuales se encuentran en los cordones posteriores (DC, por sus siglas en inglés) y en las raíces dorsales (DR, por sus siglas en inglés) de la médula espinal.

Durante los últimos 40 años, el modelado computacional de la terapia de EME ha sido una herramienta clave para entender el efecto de los parámetros de estimulación en la activación neuronal [3], [4], para estudiar el efecto del diseño de los electrodos [5] e incluso para establecer hipótesis sobre los posibles mecanismos de acción que hay detrás de la estimulación eléctrica [6]. Además, la necesidad de mejorar el manejo de los parámetros de

estimulación para producir una adecuada cobertura de parestesia en cada paciente en la práctica clínica ha favorecido el desarrollo de modelos de EME personalizados al paciente en los últimos años [7], [8]. Sin embargo, a pesar de que los modelos personalizados han demostrado ser más precisos que los modelos generalizados [9], todavía no se ha estudiado su utilidad en la práctica clínica.

Con el objetivo de conocer la capacidad de un modelo de EME personalizado al paciente en un escenario clínico, el presente estudio se centra en el desarrollo de una aplicación clínica que consiste en la optimización de la configuración de los electrodos antes del implante de los componentes de EME.

2. Materiales y métodos

2.1. Datos clínicos

El paciente de este estudio fue diagnosticado de síndrome de cirugía de columna fallida y reportó dolor crónico en el miembro inferior izquierdo y en las lumbares de la parte izquierda. Después del implante del sistema de EME, en la sesión de programación se encontró que la configuración de cátodo guardado transverso (programado en los polos 5(+), 6(-), 7(+) y 14(+)) cubría los dermatomas dolorosos del paciente (ver Figura 1).

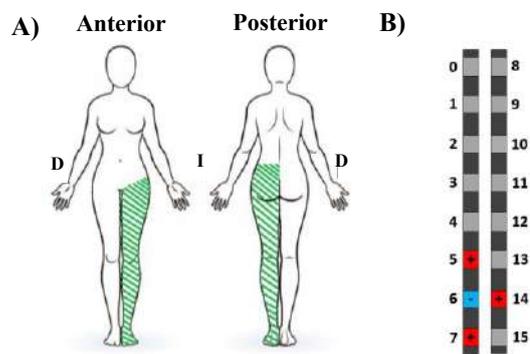


Figura 1. A) Dermatoma doloroso del paciente (izquierda). B) Configuración de electrodos programada para cubrir la zona de dolor del paciente (cátodo guardado transverso en polos 5(+), 6(-), 7(+) y 14(+)) (derecha). D: derecha; I: izquierda.

2.2. Modelo 3D de médula espinal

A partir de imágenes de resonancia magnética de alta resolución en T1 y T2* obtenidas del paciente de este

estudio, se ha desarrollado el modelo 3D de médula espinal específico del paciente, el cual incluye los niveles vertebrales T8, T9 y T10. Para ello, se ha utilizado el software *Spinal Cord Toolbox* (SCT) [10], con el que se ha segmentado automáticamente la sustancia blanca y el líquido cefalorraquídeo (CSF). La segmentación de los cuerpos vertebrales se ha realizado manualmente con el programa 3Dseg. Estas segmentaciones han sido exportadas a COMSOL Multiphysics®, donde se ha añadido los tejidos restantes: la duramadre, con un grosor de 0.3 mm [11]; y la grasa, que se encuentra en el espacio epidural. El modelo también incluye dos electrodos percutáneos de estimulación, de ocho polos cada uno, situados en el espacio epidural, tal y como se muestra en la Figura 2. Para calcular el campo eléctrico producido a partir de un estímulo eléctrico en el modelo de EME aplicando el método de elementos finitos en COMSOL Multiphysics®, se ha utilizado una malla tetraédrica adaptativa para evitar errores de cálculo en bordes o zonas estrechas. En total, el modelo considera alrededor de unos 12.300.000 de elementos.

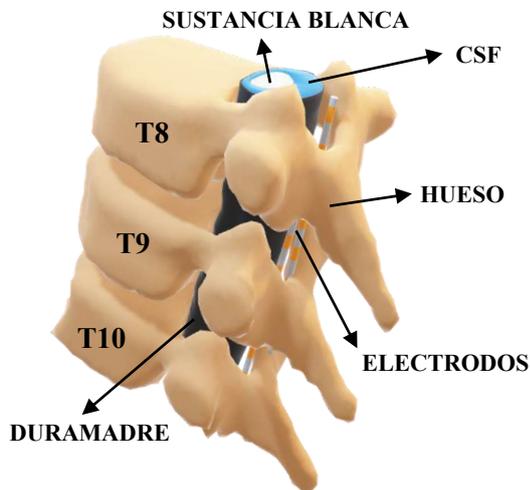


Figura 2. Modelo 3D de médula espinal específico del paciente y tejidos considerados (la grasa del espacio epidural no se muestra para poder visualizar los electrodos). CSF: líquido cefalorraquídeo.

2.3. Modelo de fibra nerviosa miélica

Para simular la activación neuronal, se ha implementado un modelo de fibra nerviosa miélica sensitiva A β en MATLAB® R2017a [12]. Este modelo incluye, en los nodos de Ranvier, los canales iónicos de sodio, potasio y de corriente de fuga; y los internodos son modelizados como un aislante imperfecto, considerando así las pérdidas de corriente que se producen al propagarse el potencial de acción. La forma del potencial de acción que se obtiene en un nodo de Ranvier con este modelo de fibra miélica sensitiva se puede observar en la Figura 3.

En cuanto a la distribución de fibras, en base a un estudio histológico sobre el tamaño y densidad de fibras presente en la médula espinal humana [13], en los DC del modelo 3D de médula espinal, se considera un total de 2.000 fibras sensitivas A β de 12.8 μm de diámetro. Cada fibra contiene 64 nodos de Ranvier. En las DR, el tamaño de fibra considerado es de 15 μm de diámetro y contienen entre 5 y 7 nodos de Ranvier.

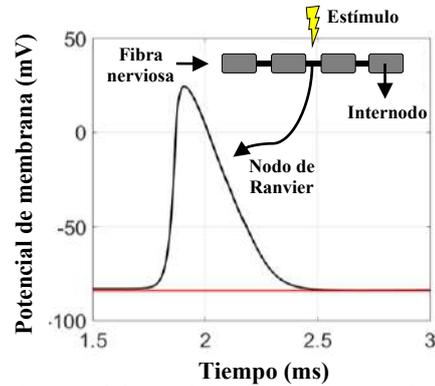


Figura 3. Potencial de acción producido en un nodo de Ranvier (negro) y potencial de reposo (-84 mV) de la fibra (rojo).

2.4. Procedimiento de simulación

Para calcular la activación neuronal, se aplica un pulso monofásico de estimulación de voltaje rectangular de 300 μs de duración. Tras el cálculo del campo eléctrico en COMSOL Multiphysics®, los valores del potencial eléctrico en los nodos de Ranvier e internodos se exportan como parámetros de entrada al modelo de fibra en MATLAB R2017a®. A partir de las simulaciones, se calculan los siguientes parámetros de evaluación para comparar los resultados obtenidos:

- **Umbral de percepción (PT):** se define como el voltaje mínimo necesario para activar la primera fibra en los DC (PT_{DC}) o en las DR (PT_{DR}). Este valor se corresponde con el voltaje programado en los polos del electrodo. Se expresa en voltios (V).
- **Número de fibras activadas de la parte izquierda o derecha de la médula espinal:** se define como el área transversal (izquierda o derecha) que se estimula por densidad de fibras de 12.8 μm en los DC ($0.11 \cdot 10^{-3} (\mu\text{m}^2)^{-1}$, de acuerdo con Feirabend et al. [13]). Este parámetro representa el número de fibras que se activarían bajo un estímulo determinado.

3. Resultados

Los electrodos se localizaron en una posición ideal, es decir, en el espacio epidural sobre la duramadre, paralelos a la sustancia blanca y con una separación entre centro y centro de cada electrodo de 2.5 mm. Dada esta posición, el objetivo es maximizar la activación neuronal en la parte izquierda de la médula espinal. La polaridad de los electrodos utilizada en las simulaciones es la de cátodo guardado transverso, que es la que clínicamente cubría los dermatomas dolorosos del paciente de este estudio. A continuación, simulamos el efecto de estimular con dicha polaridad utilizando todos los polos del electrodo izquierdo (de arriba a abajo), considerando seis configuraciones o casos en total (ver Figura 4). De esta manera se determina qué configuración permitiría optimizar la estimulación en el paciente considerado.

Tal y como se muestra en la Figura 4, como la geometría de la sustancia blanca varía en cada corte transversal, las fibras que se activan dependen de los polos que se seleccionen para estimular y de la posición del electrodo.

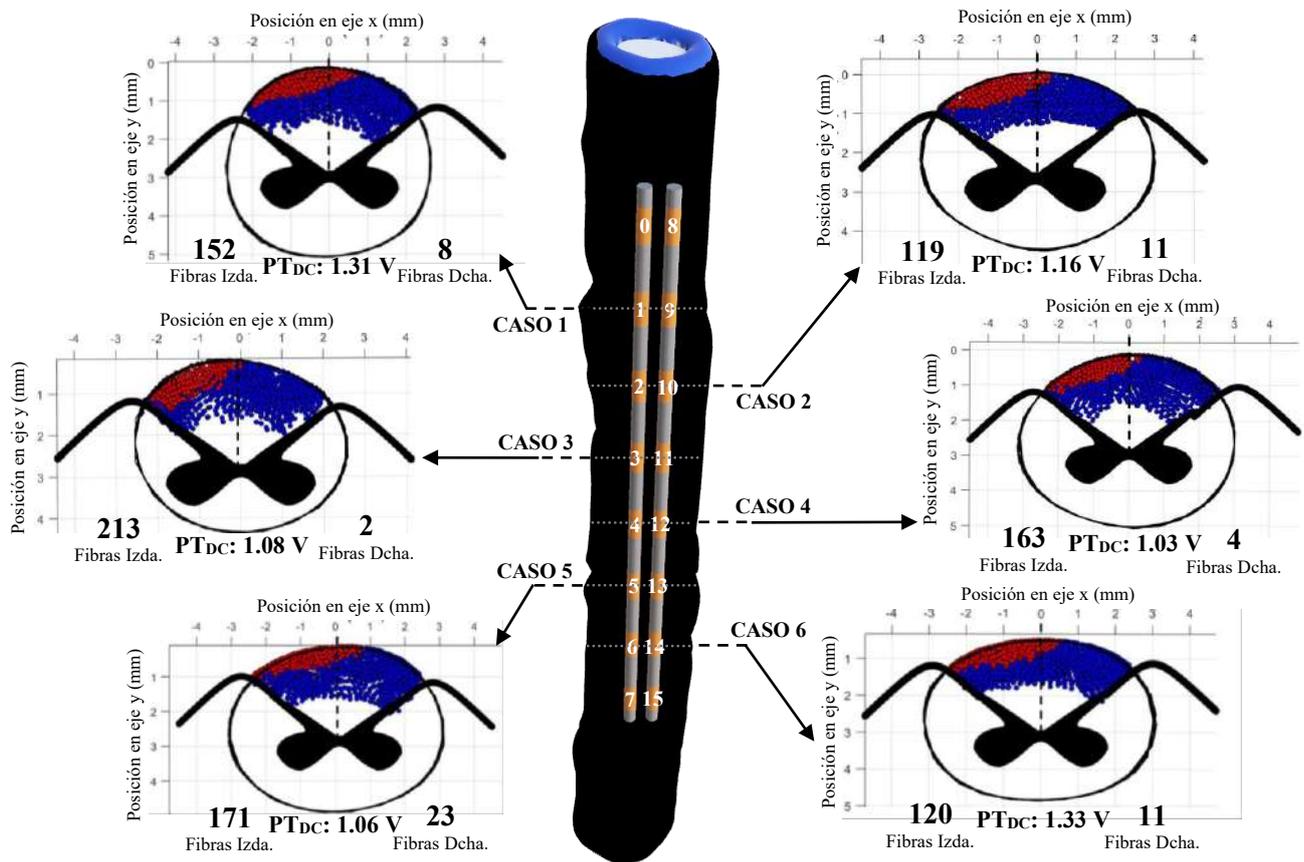


Figura 4. Activación neuronal en los cordones posteriores (DC) en seis configuraciones de electrodos: Caso 1: 0(+)-1(-)-2(+) y 9(+); Caso 2: 1(+)-2(-)-3(+) y 10(+); Caso 3: 2(+)-3(-)-4(+) y 11(+); Caso 4: 3(+)-4(-)-5(+) y 12(+); Caso 5: 4(+)-5(-)-6(+) y 13(+); Caso 6: 5(+)-6(-)-7(+) y 14(+). Izda: izquierda; Dcha: derecha; PTDC: umbral de percepción en los cordones posteriores.

Los parámetros de evaluación obtenidos en los seis casos se muestran en la Figura 4.

En general, el caso 3 maximiza la activación de las fibras de la izquierda (213 fibras frente a 2 fibras activadas en la parte derecha), mientras que el caso 5 es el que más fibras de la parte derecha activa (23 fibras).

En términos de umbrales de percepción, los casos 3, 4 y 5 presentan los valores más bajos (1.08, 1.03 y 1.06 V, respectivamente), ya que en estos casos los electrodos se encuentran más cerca de la duramadre, debido a la geometría de la médula espinal del paciente del estudio. En cambio, los casos 1, 2 y 6 requieren un PTDC más elevado para activar las fibras nerviosas (un 12-29% más que en los casos 3, 4 y 5).

4. Discusión y conclusiones

Para examinar la capacidad del uso de un modelo de EME específico del paciente, en este estudio se ha realizado una aplicación clínica sobre la optimización preoperatoria de la configuración de los electrodos.

El paciente de este estudio presentaba dolor crónico en el miembro inferior izquierdo y las lumbares de la parte izquierda. De acuerdo con Taghva et al. [14], los niveles T9 y T10 son las áreas vertebrales donde es más probable producir parestesia en la mayoría de los dermatomas del cuerpo humano. Sin embargo, en el nivel vertebral T8 es

más fácil cubrir las nalgas, la espalda y los muslos, y es menos probable cubrir las piernas y los pies. Como en T9 es menos probable estimular los dermatomas de la pierna de la parte anterior y en T10 no es probable cubrir la espalda baja que se encuentra por encima de la cintura, en el paciente de este estudio, para asegurar que la cobertura de parestesia se localizara en los dermatomas dolorosos (pierna y lumbares de la parte izquierda del cuerpo), los electrodos se localizaron idealmente en los niveles vertebrales T8 y T9 del modelo 3D de EME específico del paciente (5 polos por electrodo en T9 y 3 polos por electrodo en T8). Los electrodos percutáneos de ocho polos que se implantan actualmente (y que se han considerado en el modelo de EME) permiten estimular el nivel vertebral de interés usando diferentes polaridades y diferentes configuraciones de los electrodos (diferentes polos programados). Tener esta posibilidad es útil porque, en caso de fallo de algunos polos del electrodo, en vez de realizar una reintervención quirúrgica para reemplazar el electrodo dañado, el problema se puede resolver programando los polos funcionales que quedan.

Los resultados de este estudio (ver Figura 4) muestran que, si se programan los polos del caso 3, se consigue la mayor cobertura de parestesia de la parte izquierda del cuerpo del paciente. Sin embargo, si otros dermatomas que no son de interés fueran estimulados con esta configuración de electrodos o la estimulación fuera molesta, los polos programados en el caso 4 o en el caso 6

también permitirían maximizar la activación neuronal en la parte izquierda y, como estos polos se encuentran en T9, es más probable que la estimulación cubra los dermatomas dolorosos del paciente. Por tanto, esta aplicación clínica demuestra que es posible determinar los polos que deberían programarse primero para maximizar la cobertura de parestesia en el paciente mediante el uso de un modelo de EME específico del paciente.

Cabe destacar que, la falta de un mapa somato-tópico de los cordones posteriores de la médula espinal en cada nivel vertebral, además de la alta variabilidad en el posicionamiento de los electrodos durante el implante, hace que sea complicado implantar los electrodos en una posición ideal. Pero, tanto la distancia entre los electrodos como su alineación son parámetros que se pueden controlar durante el implante del sistema de EME. Este estudio evidencia que el uso de un modelo de EME específico del paciente junto con un atlas de cobertura de parestesia desarrollado por Taghva et al. [14], permitiría determinar estos parámetros, así como también la polaridad y la configuración de los electrodos. Esta nueva propuesta podría mejorar la cobertura de parestesia en cada paciente, es decir, de forma personalizada. Sin embargo, se debería realizar un estudio clínico con una cohorte de pacientes para validar esta hipótesis clínicamente.

En conclusión, los modelos computacionales de EME no solo son útiles para estudiar el efecto de los parámetros de estimulación, sino que presentan la capacidad de ser una futura herramienta de trabajo del facultativo para optimizar la terapia de EME en la práctica clínica. La aplicación clínica sobre la selección preoperatoria de la configuración de los electrodos podría ayudar a focalizar la cobertura de parestesia en los dermatomas dolorosos del paciente de forma más precisa, a través del desarrollo y uso de modelos de EME específicos del paciente.

Referencias

- [1] Y. Guan, “Spinal cord stimulation: Neurophysiological and neurochemical mechanisms of action,” *Curr. Pain Headache Rep.*, vol. 16, pp. 217–225, 2012.
- [2] R. Melzack and P. D. Wall, “Pain Mechanisms: A New Theory,” *Science (80-.)*, vol. 150, no. 3699, pp. 971–979, 1965.
- [3] D. Lee, B. Hershey, K. Bradley, and T. Yearwood, “Predicted effects of pulse width programming in spinal cord stimulation: A mathematical modeling study,” *Med. Biol. Eng. Comput.*, vol. 49, no. 7, pp. 765–774, 2011.
- [4] J. L. Durá, C. Solanes, J. De Andrés, and J. Saiz, “Computational Study of the Effect of Electrode Polarity on Neural Activation Related to Paresthesia Coverage in Spinal Cord Stimulation Therapy,” *Neuromodulation*, vol. 22, no. 3, pp. 269–279, 2019.
- [5] L. Manola, J. Holsheimer, and P. Veltink, “Technical performance of percutaneous leads for spinal cord stimulation: A modeling study,” *Neuromodulation*, vol. 8, no. 2, pp. 88–99, 2005.
- [6] S. F. Lempka, C. C. McIntyre, K. L. Kilgore, and A. G. Machado, “Computational Analysis of Kilohertz Frequency Spinal Cord Stimulation for Chronic Pain Management,” *Anesthesiology*, vol. 122, no. 6, pp. 1362–1376, 2015.
- [7] S. F. Lempka, H. J. Zander, C. J. Anaya, A. Wyant, J. G. Ozinga, and A. G. Machado, “Patient-Specific Analysis of Neural Activation During Spinal Cord Stimulation for Pain,” *Neuromodulation*, vol. 23, no. 5, pp. 572–581, 2020.
- [8] C. Solanes, J. L. Durá, M. Á. Canós, J. De Andrés, L. Martí-Bonmatí, and J. Saiz, “3D patient-specific spinal cord computational model for SCS management: potential clinical applications,” *J. Neural Eng.*, vol. 18, no. 036017, 2021.
- [9] C. Solanes, J. L. Durá, M. A. Canós, J. De Andrés, and J. Saiz, “Predicción de la activación neuronal en estimulación tónica en la terapia de EME: uso de modelos 3D de médula espinal personalizados al paciente frente a modelos generalizados,” in *XXXVIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica*, 2020, pp. 306–309.
- [10] B. De Leener, J. Cohen-Adad, and S. Kadoury, “Automatic Segmentation of the Spinal Cord and Spinal Canal Coupled with Vertebral Labeling,” *IEEE Trans. Med. Imaging*, vol. 34, no. 8, pp. 1705–1718, 2015.
- [11] H. J. Zander, R. D. Graham, C. J. Anaya, and S. F. Lempka, “Anatomical and technical factors affecting the neural response to epidural spinal cord stimulation,” *J. Neural Eng.*, vol. 17, no. 3, p. 036019, 2020.
- [12] C. Solanes, J. L. Dura, J. De Andres, and J. Saiz, “What Is the Role of Frequency on Neural Activation in Tonic Stimulation in SCS Therapy? A Computational Study on Sensory Aβ Nerve Fibers,” *IEEE Access*, vol. 9, pp. 107446–107461, 2021.
- [13] H. K. P. Feirabend, H. Choufoer, S. Ploeger, J. Holsheimer, and J. D. van Gool, “Morphometry of human superficial dorsal and dorsolateral column fibres: significance to spinal cord stimulation,” *Brain*, vol. 125, p. 1668, 2002.
- [14] A. Taghva, E. Karst, and P. Underwood, “Clinical Paresthesia Atlas Illustrates Likelihood of Coverage Based on Spinal Cord Stimulator Electrode Location,” *Neuromodulation*, vol. 20, no. 6, pp. 582–588, 2017.

P-Wave Detection using Latent Spaces from Autoencoders and Domain Description Classifiers

C. Plaza–Seco^{1*}, M. Blanco–Velasco¹, J.L. Rojo–Álvarez², and S. Muñoz–Romero²

¹ Departamento de Teoría de la Señal y Comunicaciones, Universidad de Alcalá, {carmen.plazas, manuel.blanco}@uah.es

² Departamento de Teoría de la Señal y Comunicaciones y Sistemas Telemáticos y Computación, Universidad Rey Juan Carlos, {joseluis.rojo, sergio.munoz}@urjc.es

Abstract

The P-wave of the electrocardiogram (ECG), which represents the atrial activation of the heart, is a subject of significant attention due to the incidence of atrial arrhythmia in cardiac disease. Recording and detecting this wave is an important task that presents two main challenges: (1) the difficulty of detecting the electrical activity that leads the P-wave in long-term ECG records; and (2) the importance of interpretability of machine learning engines in the clinical field. Therefore, the main goal of this project is to find ECG P-waves using Machine Learning and Deep Learning techniques and, at the same time, to make them visual and interpretable. The approach has been carried out with autoencoders (AEs), one of the models of pattern extraction from high dimensional datasets. These networks allow to extract relevant features from the data and to study how cardiac signals are translated into manifolds in latent spaces. In addition to that, this procedure facilitates the interpretability of the P-wave detection problem with classical kernel classifiers, like Support Vector Domain Description, which is a classical anomaly detection algorithm. Despite the low-dimensional spaces used in the AEs, the high interpretability and the promising results obtained, such as F1 score of 62.7%, show the interest and feasibility of this approach to delineate the entire heartbeat and to reliably interpret and detect cardiac anomalies in close steps through this pathway.

1. Introduction

An arrhythmia is any disturbance in the common rhythm of the heart. These alterations involve a disruption of the excitation-conduction system, which consists of a set of structures responsible of the genesis and distribution of the electrical impulse. The variety of cardiac arrhythmias is large, but one of the most usual is atrial arrhythmias [1], which corresponds to P-wave in the electrocardiogram (ECG). Most of them happen intermittently, requiring additional examinations. It is highly complex, though necessary to study the ECG when a dangerous arrhythmia is suspected, due to the permanent injuries they can cause.

Screening techniques are been used to study ECG recordings, like stress tests or Holter recordings, among others. However, there are some drawbacks to be taken into account, such as the amount of time spent by medical staff analyzing the data collected by these devices. In the 24-

hour Holter, a few minutes up to an hour can be spent reviewing the annotations made by current algorithms.

One of the most recent advances in this field has been the increasing application of Machine Learning (ML) and Deep Learning (DL) techniques in the health area. This increase, along with new tools to analyze large amounts of data efficiently, has resulted in ML taking an important role in healthcare innovation [2]. These kinds of systems have proven to be powerful tools to assist clinicians in their daily clinical routine to produce more consistent and faster detection. However, many of them do not provide detailed basis for understanding the classification results, like their interpretation.

Many contributions have been made in DL for cardiac signals [3], but less attention has been paid to the interpretation of the features extracted by neural networks. Therefore, this work studies how cardiac signals are translated into varieties in these spaces within the framework of a P-wave detection problem. For this purpose, autoencoders (AEs) have been used due to their high capacity to extract patterns from high dimensional data. Moreover, Support Vector Machines (SVMs) are used as classification tool, showing promising results for the discrimination of different types of cardiac anomalies [2]. Particularly, Support Vector Domain Description (SVDD), which is a classical anomaly detection algorithm [4], has been chosen.

2. Methods and Databases

2.1. Autoencoders

An AE is a neural network trained to replicate its input in an unsupervised manner [5]. Internally, it has a hidden layer that learns relevant features from a dataset and provides information about it. The learning process consists of minimizing, through the gradient descent method, a loss function $\mathcal{L}(f(x))$, such as the mean squared error:

$$\mathcal{L}(f(\mathbf{x})) = \frac{1}{2} \sum_{i=1}^{\ell} (\hat{\mathbf{x}}_i - \mathbf{x}_i)^2 \quad (1)$$

Sparse and contractive AEs are two of the regularized AEs which make the dimensionality reduction possible in the hidden layer. To this end, a penalty factor in the loss function is used to ensure that only some of the neurons in the hidden layer can be activated at a time. Thus, the AE is forced to learn the really relevant features, obtaining a better performance and avoiding overfitting. Besides, using contractive AEs allows to represent the cardiac signal in a visualizable and interpretable way. These kinds of AEs are typically used to learn features to be used in another task, such as in a classification problem [5].

2.2. Support Vector Domain Description

SVMs have become one of the most prominent supervised learning methods for solving regression, classification, or even domain detection problems. In this work, an SVDD model has been used for one-class classification due to its ability as a domain detector when features from multiple classes are present. This model is a variant of binary SVM which tries to identify only cases from a specific class among all the data in a given set. To do it, this method learns the distribution of the features from a training set that contains only events of the class to be studied [4]. Thus, the purpose of this model is to find the minimum sphere in the Reproducing Kernel Hilbert Space (RKHS) that contains all instances of the trained class, which is described by its centre c and its radius r . Thus, the model consists of an optimization problem described as follows:

$$\begin{aligned} \underset{\zeta}{\text{minimize}} \quad & r^2 + \frac{1}{\nu n} \sum_{i=1}^n \zeta_i, \\ \text{subject to} \quad & \|\Phi(\mathbf{x}_i) - c\|^2 \leq r^2 + \zeta_i, \end{aligned} \quad (2)$$

where $\Phi(x_i)$ refers to the kernel transformation for each sample and the parameter ν allows to control the behaviour of the sphere boundary. A sphere in the RKHS when using a nonlinear kernel corresponds to a possibly irregular shape of the described domain in the input space.

One of the downsides in the ECG segmentation task is the insufficient labelling of long duration recordings, so the use of one-class classifiers can be of great help because of its region detection potential.

2.3. Databases

In this work, two of the most used databases in the ECG field have been chosen.

QT Database. Training data are obtained from ECG signals from the publicly available QT Database (*qtdb*) [6]. The data consists of 15 minutes of two-channel ECG recordings, with a sample rate of 250 Hz, measured from a total of 105 patients. The database provides signal region labels generated by an automated expert system [7]. This database has been used in the same way as in [8].

MIT-BIH. Patients from MIT-BIH Arrhythmia Database (*mitdb*) [9] are used to test the trained SVDD classifier and make predictions due to its direct availability of beat type labeling. These patients were labeled with the waveform

segmentation tool introduced in [8]. This methodology has been chosen to generate labeled data as a first attempt to study the behaviour of the classifier and its performance.

All the time series used in the experiments have been pre-processed with basic routines to eliminate the different artifacts present in the signals, namely, trend elimination filter, notch filter (60Hz), and band-pass filter (0,5 – 40Hz).

3. Results

The main purpose of the experiments developed in this section is twofold: first, studying the features extracted by neurons from the hidden layer of AEs in the latent space; and second, exploring the effects of the parameters of the SVDD in the classification problem.

To carry out the experiments, we created a set of systems operating on continuous signal recordings that had been segmented into blocks of 0,8 ms with 97,5 % overlapping. Only the segments with a P-wave located in the central zone, i.e., with a P-wave crossing the center of the segment, are labeled as 1. All the segments are projected into the latent space by the AE, and the purpose of the classifier is to determine the region of the latent space in which the features of these vectors are found.

3.1. Parameter Experiments

This section is intended to determine which combination of SVDD parameters is better suited for this task. The Gaussian kernel is used due to the nonlinearity of the features in latent space, and two free parameters are checked: the kernel scale (h), equivalent to σ in regular SVM; and ν , which controls the boundary softness of the domain description (bubble) in the latent space. We aim to understand the behaviour of the SVDD model and the effect of using either a higher or lower kernel width, h , as well as a higher or lower value of ν to widen or narrow the boundary limits.

Whenever the latent space area in which the P-wave features are located is well defined, the classifier gives good results with a soft boundary ($\nu = 0,05$). However, when working with data from different patients, the boundary softness has to be restricted to higher values of ν to constrain

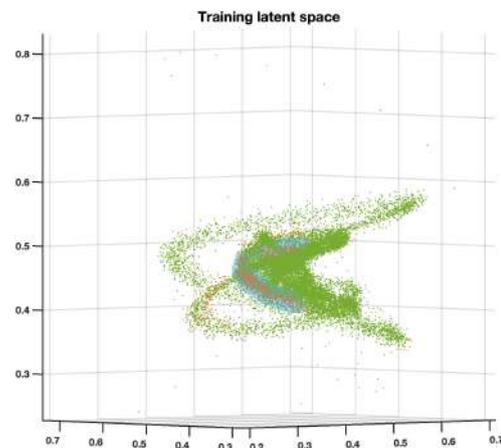


Figure 1. Training latent space from three different sinus patients and SVDD with $h = 0,01$ and $\nu = 0,95$.

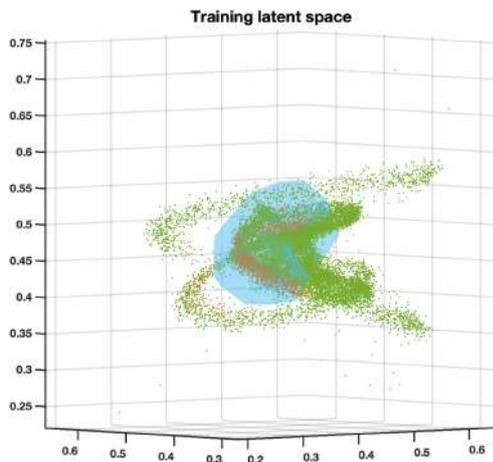


Figure 2. Training latent space from three different sinus patients and SVDD with $h = 0,05$ and $\nu = 0,95$.

the domain bubble to an area with more points of interest avoiding that features from other heartbeat parts enter in the delimited area.

The blue bubble depicted in Figure 1 represents the training domain obtained by the SVDD in the latent space for $h = 0,01$ and $\nu = 0,95$, which have been chosen for corresponding to a good trade-off to be fit into the P-wave area (red points). Training has been carried out with two minutes of three sinus rhythm patients from the *qtdb*, namely: *ecg1*, *ecg10*, and *ecg100*, according to notation in [8]. In general, increasing the value of h , the bubble becomes larger and covers a section outside the P-wave area, as it can be seen in Figure 2 ($h = 0,05$, $\nu = 0,95$), which is undesirable. As mentioned above, the same occurs with parameter ν , and as long as ν becomes smaller, the bubble boundary turns softer, which enlarges the bubble.

A testing set has been chosen with three patients not included in the training set, which consists of two minutes long signals taken from *mitdb* common sinus rhythm recordings 100, 101, and 222. An example is shown in Figure 3, which presents one fragment of the SVDD predictions on the *mitdb* 100 recording. As it can be seen, predictions match quite well the true labels assigned to the P-wave. The performance on this testing set is shown in the first row of Table 1 in terms of precision, recall, and F_1 score. For the computation of this metric, a detection is considered to be a true positive when any predicted label overlaps the real label area. Despite the AE was defined to be very simple, just for visual interpretation purposes, i.e, a shallow net with only 3 active cells in the latent space, the results show that the system exhibits some power of detection. Subsequently, an additional signal is incorporated in the test set, which was recording 102 from *mitdb* and corresponds to a paced patient, achieving the performance

Test dataset	Precision	Recall	F1 score
Patients in sinus rhythm	62 %	63,5 %	62,7 %
Sinus patients + 1 paced	46,7 %	47,7 %	47,2 %
Sinus training patient	48,6 %	92,5 %	63,7 %

Table 1. Performance of the trained SVDD (with three sinus patients) testing with three different testing sets.

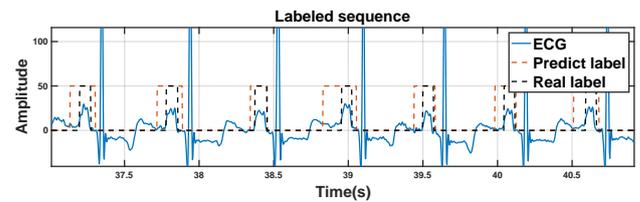


Figure 3. Predictions and real labels on *mitdb* 100 with trained SVDD classifier ($h = 0,01$ and $\nu = 0,95$).

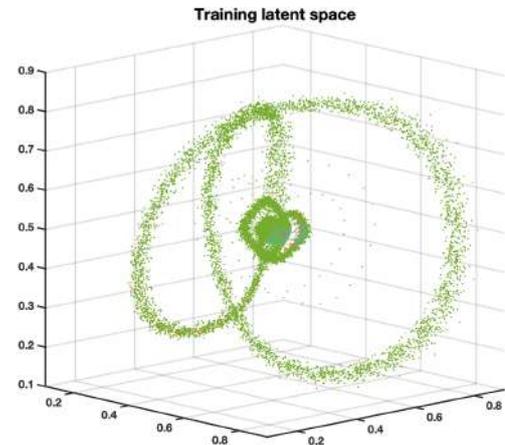


Figure 4. Training latent space with four patients: three in sinus rhythm plus another one paced.

shown in the second row of Table 1. We can see that the detector yields worse performance, which may be due to the paced patients does not presenting a clear P-wave pattern in their ECG morphology. This issue is checked in the next section, by studying the distribution of a paced recording.

3.2. Paced Patients in the Training Set

Join visual inspection of both the data projections in the latent space and the influence area of the classifier has been shown to gain interpretability of the system when working with real signals and few labels. Following this methodology, the distribution variability of the features extracted by the AE when different kinds of patients are selected is studied next.

Paced patient *ecg4* from *qtdb* is included in the training set described in Section 3.1, which results in the latent space shown in Figure 4. A central area almost identical to the illustration depicted in Figure 1 can be distinguished, which is mainly supported by the recordings included in the original training set. This space portion is surrounded by outer and farther rings that emerge due to the influence of the paced patient. Thus, the latent space of the AE has undergone significant changes, concluding that it has been highly affected by the morphology of the cardiac signal.

For this particular application of P-wave detection, paced recordings seem to be not suitable signals because the P-wave is masked by the stimulated beat and it can scarcely be recognized. Thus, the one-class classifier can be impaired to capture information of this area. In summary, it may not be convenient to choose recordings for the training set which do not report specific information of the phenomenon that has to be characterized by the one-class classifier.

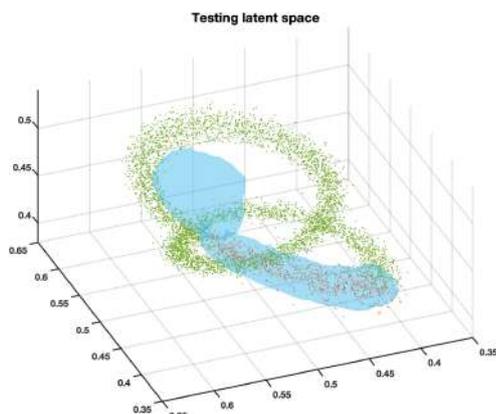


Figura 5. Testing latent space from a training patient.

3.3. Consequences of Overfitting

Finally, the performance of the AE trained in Section 3.1 has been checked on one signal of the training test, but in a separate section which was not used to train. Figure 5 shows the testing latent space from a two minutes long section of patient *ecg100*, where P-wave features, represented in red, are quite well enclosed inside the training bubble, standing for correct detections. Nevertheless, others related to other ECG beat parts, which are sketched as green dots in the graph, also enter into the bubble, which account for wrong detections. The performance on this testing approach is shown in the third row of Table 1, where the metrics seize the referred behavior, and despite the sensitivity is high, the number of false positives leads to decreased precision of the system.

4. Discussion and Conclusions

Latent spaces bring the opportunity to interpret how the SVDD classifier works and how the AE extract manifolds from the high dimensional data. As shown, the distribution of the features in the latent space strongly influences the ECG sub-wave detections, but SVDD model is capable of narrowing down the area of interest in latent spaces with a large volume of features.

An application like this one may be of great interest, where the delineation is accomplished with a very small set of labels relative to a lot of unlabeled data. The performance of the classifier can be readjusted through Active Learning over new predictions, i.e., by letting the classifier learn iteratively from the knowledge that it already has plus feedback received when making new predictions about new data or even additional patients. This could help to decrease the false positive rate, to control the overfitting and, at the same time, to improve the accuracy at the onsets and endings of the detections.

It has been found that, in spite of its high interpretability, low-dimensional spaces are usually insufficient for data analysis from a patient population. That is why deep networks with more layers and higher dimensionality are usually used to improve the performance. This first study shows the feasibility that a similar approach as the one followed in this work can be useful to generalize applications

like this and to understand what happens when DL models are used, as well as to be aware of the risk of overfitting.

Currently, there is no system capable of reliably analyzing signals of long duration while being at the same time easy to be understood. Therefore, additional advance research following this pathway could show the interest of this approach to delineate the entire heartbeat, and also to detect cardiac anomalies. Analysis of AEs with more than one layer would surely help to outperform results, especially with LSTM, convolutional, variational AEs, or even with topological embedding techniques, as they may report relevant advantages over conventional AEs and can be studied with similar tools such as those developed in this project.

Acknowledgments

This work was supported by the University of Alcalá and the Community of Madrid under grants COVID-19 2021 2020/00003/016/001/019 and EPU-INV/2020/002. Also, this work was supported by the Science and Innovation Ministry Grants meHeart-RisBi and Beyond (PID2019-104356RB-C42 and PID2019-106623RB-C41).

Referencias

- [1] L. S. Lilly, *Cardiología. Bases fisiopatológicas de las cardiopatías*. Wolters Kluwer, 6 ed., 2016.
- [2] J. L. Rojo-Álvarez, G. Camp-Valls, A. Caamaño-Fernández, and J. Guerrero-Martínez, *ECG Signal Processing, Classification and Interpretation. A Review of Kernel Methods in ECG Signal Classification*. Springer, 2012.
- [3] L. Bote-Curiel, S. Muñoz-Romero, A. Gerrero-Currieses, and J.L. Rojo-Álvarez, "Deep learning and big data in healthcare: A double review for critical beginners," *Applied Sciences*, vol. 9, no. 11, 2019.
- [4] Z. Zhang and X. Deng, "Anomaly detection using improved deep SVDD model with data structure preservation," *Pattern Recognition Letters*, vol. 148, pp. 1–6, 2021.
- [5] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. Book online source.
- [6] P. Laguna, R.G. Mark, A. Goldberger, and G.B. Moody, "A Database for Evaluation of Algorithms for Measurement of QT and Other Waveform Intervals in the ECG," *Computers in Cardiology*, vol. 24, p. 673–676, 1997. QT Database online source.
- [7] P. Laguna, R. Jané, and P. Caminal, "Automatic detection of wave boundaries in multilead ECG signals: Validation with the CSE database," *Computers and Biomedical Research*, vol. 27, no. 1, pp. 45–60, 1994.
- [8] MATLAB, "Waveform segmentation using deep learning - MATLAB & Simulink - MathWorks." Waveform segmentation online source.
- [9] G.B. Moody, and R.G. Mark, "The impact of the MIT-BIH Arrhythmia Database," *IEEE engineering in medicine and biology magazine: the quarterly magazine of the Engineering in Medicine & Biology Society*, vol. 20, no. 3, pp. 45–50, 2001.

“Time-Warping” Ponderado Sobre la Onda T Para Reducir el Impacto de Errores de Delineación: Implicaciones Clínicas

F. Palmieri^{1,2,3}, P. Gomis^{1,4,2}, J.E. Ruiz⁵, D. Ferreira³, E. Pueyo^{2,6}, J.P. Martínez^{2,6}, P. Laguna^{2,6}, J. Ramírez⁷

¹ Centre de Recerca en Enginyeria Biomèdica, Universitat Politècnica de Catalunya, Barcelona, España.

² CIBER en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Zaragoza, España.

³ Laboratorios Rubio, Castellbisbal, Barcelona, España

⁴ Universidad Internacional de Valencia, Valencia, España

⁵ Servicio de Nefrología, Hospital Clínico Universitario Lozano Blesa, Zaragoza, España.

⁶ BSICoS Group, I3A, IIS Aragón, Universidad de Zaragoza, Zaragoza, España.

⁷ William Harvey Research Institute, Queen Mary University of London, London, United Kingdom

Resumen

La morfología de la onda T (TW) se ha investigado para desarrollar marcadores de predicción de riesgo arritmico específicos como d_w , que cuantifica el nivel de “warping” necesario para alinear temporalmente una TW con respecto a otra de referencia a diferente ritmo cardiaco. Sin embargo, los errores de delineación de los límites de la TW pueden poner en peligro la capacidad de diagnóstico de d_w . Por lo tanto, se han propuesto dos funciones de ponderación (WF) a partir de (i) la TW de referencia (\mathcal{T}) y (ii) el valor absoluto de su primera derivada (\mathcal{D}). Se simuló errores de delineación desplazando los límites de dos TW diferentes. Se calculó el ratio de variación (\mathcal{R}), comparando d_w derivado, aplicando las dos WF con el obtenido en el caso sin ponderación (\mathcal{C}). Luego, se comparó la capacidad de d_w , con y sin WF, para monitorizar cambios en la concentración de potasio en sangre ($\Delta[K^+]$) mediante la correlación de Pearson (r) en 29 registros Holter de 48 horas de pacientes durante hemodiálisis (HD). Las simulaciones mostraron que los valores de \mathcal{R} para las dos TW fueron, respectivamente, 0,17 y 0,19 para \mathcal{C} , 0,05 y 0,08 para \mathcal{T} y 0,07 y 0,10 para \mathcal{D} . Sin embargo, se encontraron valores similares de mediana r [rango intercuartílico] para \mathcal{C} (0,90 [0,27]), \mathcal{T} (0,90 [0,25]) y \mathcal{D} (0,90 [0,29]) en el caso de HD. En conclusión, el uso de WF puede disminuir los efectos de los errores de delineación de las TW, aunque no se hallaron mejoras significativas en la monitorización de $\Delta[K^+]$.

1. Introducción

La morfología de la onda T (TW) en el electrocardiograma (ECG) se ha investigado ampliamente en un intento de identificar marcadores de riesgo sensibles y específicos en diferentes escenarios, como la estratificación del riesgo de muerte súbita cardíaca (en inglés, Sudden Cardiac Death, SCD) [1]. En particular, el marcador de morfología de TW, d_w , que cuantifica el nivel de deformación necesario para alinear temporalmente la morfología de una TW con respecto a la morfología de una TW de referencia [1], ha demostrado una buena correlación con la dispersión de la repolarización ventricular y el riesgo arritmico [2]. La delineación automática de los límites de las TW es un paso necesario en el cálculo de d_w , pero es muy susceptible al ruido, especialmente en casos con baja amplitud y

variabilidad morfológica de las TW [3]. Los errores de delineación de los límites de las TW pueden afectar en gran medida la capacidad de diagnóstico de d_w , lo que podría afectar su valor de estratificación de riesgo. La hipótesis de este estudio es que al introducir una etapa de ponderación en el cálculo de d_w se reduciría el efecto de los errores de delineación de los límites de las TW. Se proponen dos funciones de ponderación (WF): (i) la TW de referencia, $\mathcal{W}_{\mathcal{T}}(t)$, y (ii) el valor absoluto de su primera derivada, $\mathcal{W}_{\mathcal{D}}(t)$. Luego, se ha evaluado la robustez de los marcadores d_w resultantes frente a errores simulados en la delineación de los límites de TW, así como su capacidad para seguir los cambios en la concentración de potasio en sangre ($[K^+]$) en pacientes en tratamiento de hemodiálisis (HD), comparándolo con respecto al caso de control sin ponderación \mathcal{C} , $\mathcal{W}_{\mathcal{C}}(t) = 1$.

2. Poblaciones de estudio

Errores de delineación de los límites de las TW: Se extrajo una TW de referencia de un latido sin ruido, muestreado a 1 kHz, registrado en posición supina en un sujeto sano [4].

Pacientes sometidos a tratamiento de HD: Se analizaron 29 ECG Holter de 12 derivaciones de 48 horas (frecuencia de muestreo 1 kHz) de pacientes con enfermedad renal en etapa terminal sometidos a HD (ESRD-HD) en el Hospital Clínico Universitario Lozano Blesa (Zaragoza, España) [5]. Para cada paciente, se tomaron cinco muestras de sangre de donde se obtuvieron cinco $[K^+]$: una al inicio de la HD (h_0), tres cada hora siguiente (h_1 , h_2 y h_3) y una al final de la HD (h_4).

3. Métodos

Análisis de las componentes periódicas: El análisis de las componentes periódicas (πC) es una técnica de descomposición de valores propios cuyo objetivo es extraer las fuentes más periódicas de la señal.

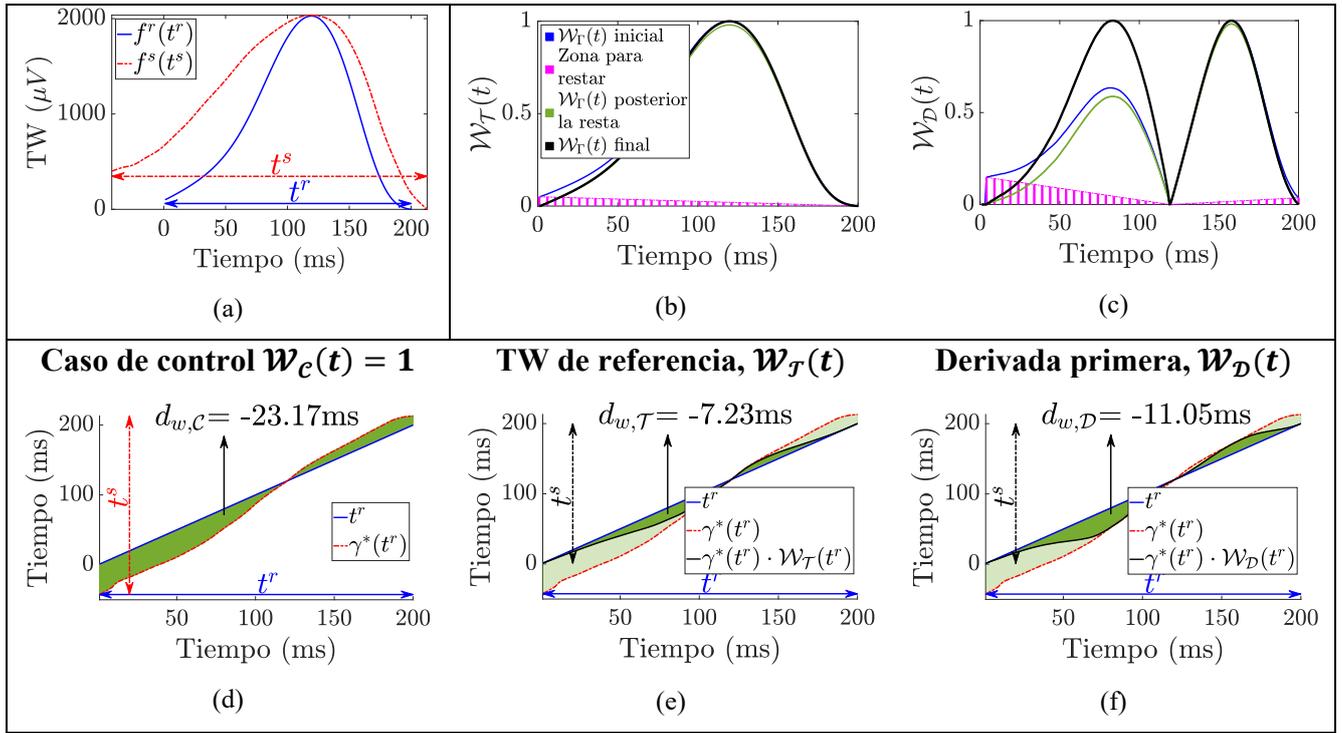


Figura 1. Ejemplo de cálculo de las $\mathcal{W}_\Gamma(t)$, $\Gamma \in \{C, T, D\}$ y su aplicación para la estimación de d_w . El panel (a) muestra la TW de referencia y la que queremos alinear temporalmente (azul y rojo, respectivamente). Los paneles (b) y (c) muestran el proceso para derivar $\mathcal{W}_T(t)$ y $\mathcal{W}_D(t)$, respectivamente, como se detalla en la sección 3. El panel (d) muestra el cálculo de $d_{w,C}$ (área verde oscuro debajo de $\gamma^*(t^r)$), mientras que los paneles (e) y (f) muestran el cálculo de $d_{w,T}$ y $d_{w,D}$, estimados como $\gamma^*(t^r) \cdot \mathcal{W}_T(t^r)$ y $\gamma^*(t^r) \cdot \mathcal{W}_D(t^r)$, respectivamente, obteniendo las zonas verde oscuro.

En este estudio, πC se aplicó para maximizar las componentes periódicas latido a latido de las TW en la primera derivación transformada como en [6].

Funciones de ponderación y cálculo de d_w : Sea $f^s(t^s)$ la TW que queremos alinear temporalmente con respecto a la TW de referencia $f^r(t^r)$ (Fig. 1a en rojo y azul respectivamente), siendo $t^s = [t^s(1), \dots, t^s(N_s)]^T$ y $t^r = [t^r(1), \dots, t^r(N_r)]^T$ sus duraciones y N_s y N_r la duración total de t^s y t^r , respectivamente.

La primera WF, $\mathcal{W}_T(t)$, se tomó como la TW de referencia, normalizando $f^r(t^r)$ por su valor máximo (Fig. 1b, TW azul). Luego, se restó el área residual bajo la TW, definida como la unión de los dos extremos (Fig. 1b, magenta), obteniendo la TW verde (Fig. 1b), que se volvió a normalizar por su valor máximo, resultando en la $\mathcal{W}_T(t^r)$ (Fig. 1b, negro). Así, no se aplica ninguna ponderación a los extremos de la TW, por lo que se reduce el efecto de los errores de delineación de los límites de las TW.

Para calcular la segunda WF, $\mathcal{W}_D(t)$, dividimos en dos mitades el valor absoluto de la primera derivada de la TW de referencia, $f^r(t^r)$, siendo el valor cero del medio (correspondiente al pico de $f^r(t^r)$) el punto de división. Luego, se aplicó individualmente a cada mitad el mismo procedimiento descrito anteriormente. Finalmente, se unieron las dos mitades, obteniendo $\mathcal{W}_D(t^r)$ (Fig. 1c). Luego, se multiplicó la función óptima de warping, $\gamma^*(t^r)$, que relaciona $f^s(t^s)$ y $f^r(t^r)$ ($\gamma^*(t^r)$, sin ponderación, Fig. 1d línea roja) [1] por $\mathcal{W}_T(t^r)$ ($\gamma^*(t^r) \cdot \mathcal{W}_T(t^r)$, Fig. 1e línea negra) y por $\mathcal{W}_D(t^r)$ ($\gamma^*(t^r) \cdot \mathcal{W}_D(t^r)$, Fig. 1f

línea negra), respectivamente. La información de deformación total (área verde oscuro) se cuantificó mediante $d_{w,\Gamma}$:

$$d_{w,\Gamma} = \left(\frac{s_d}{|s_d|} \right) \frac{1}{N_r} \sum_{n=1}^{N_r} |\gamma^*(t^r(n)) \mathcal{W}_\Gamma(t^r(n)) - t^r(n)| \quad (1)$$

$$s_d = \sum_{n \in N_r^u} (\gamma^*(t^r(n)) - t^r(n)) + \sum_{n \in N_r^d} (t^r(n) - \gamma^*(t^r(n))) \quad (2)$$

siendo N_r^u el conjunto de muestras de la pendiente ascendente de la TW. Un signo positivo significa que $f^s(t^s)$ debe ampliarse para que se ajuste a la referencia $f^r(t^r)$, y viceversa para un signo negativo.

Simulación de errores de delineación: A partir de $f^r(t^r)$ en la Fig. 2a, se obtuvieron dos TWs moduladas $f^s(t^s)$, con $s \in \{1, 2\}$ (Fig. 2b y 2c), generando primero la variabilidad no lineal de la amplitud de las TWs como:

$$f_{NL}^s(t^r) = f^r(t^r) + c(s) \cdot \sin\left(2\pi \frac{1}{4N_r} t^r\right) \quad (3)$$

donde $c(1) = 150$ y $c(2) = -150$. Luego, se incluyó la variabilidad lineal de amplitud de la TW lineal como $f^s(t^r) = f_{NL}^s(t^r) \cdot b(s)$, siendo $b(1) = 1.15$ y $b(2) = 0.85$. A continuación, se generó la variabilidad lineal de la duración de la TW como $t_i^s = \gamma_s(t^r)$ donde $\gamma_s(t^r)$ remuestrea t^r a través del factor $k(1) = 0.7$ y $k(2) = 1.3$. Por último, la variabilidad de duración lineal y no lineal de la TW fue generada como:

$$\mathbf{t}^s = \mathbf{t}_i^s + d(s) \frac{N_r}{N_s} \cdot \sin\left(2\pi \frac{1}{N_r} \mathbf{t}_i^s\right) \quad (4)$$

donde $d(1) = -15$, $d(2) = 15$ y N_s es la duración de \mathbf{t}_i^s . Por lo tanto, toda la variabilidad de amplitud y duración de la TW se calculó como:

$$\mathbf{f}^s(\mathbf{t}^s) = \mathbf{f}^s\left(\gamma_s(\mathbf{t}^r) + d(s) \frac{N_r}{N_s} \cdot \sin\left(2\pi \frac{1}{N_r} \gamma_s(\mathbf{t}^r)\right)\right) \quad (5)$$

A continuación, para ambos casos $\mathbf{f}^s(\mathbf{t}^s)$ y $\mathbf{f}^r(\mathbf{t}^r)$, simulamos errores de delineación de los límites de las TW como desplazamientos de los puntos de inicio y de fin, eliminando (agregando) progresivamente m muestras, comenzando por $m = 1$ muestra (1 ms), hasta $m = 25$, como se muestra en la Fig. 2d y la Fig. 2e (TW negros), lo que da como resultado una variabilidad de los límites de las TW, $\sigma = 14,9$ ms, dentro de la tolerancia de las mediciones del final de las TW evaluada manualmente por expertos [7]. A continuación, se calculó d_w a través del análisis time-warping entre cada TW con errores de delineación y la TW de referencia (Fig. 2a) con y sin aplicar las dos WF. Así, para ambos casos, $s \in \{1, 2\}$ (Fig. 2d y 2e, respectivamente) y ponderaciones, $\Gamma \in \{\mathcal{C}, \mathcal{T}, \mathcal{D}\}$, se extrajeron tres series de $d_{w,\Gamma}^s(m)$, $m \in \{1, \dots, 25\}$. Finalmente, se calculó el ratio de variación (\mathcal{R}) como:

$$\mathcal{R}_\Gamma^s = \frac{\sigma(d_{w,\Gamma}^s(m))}{|d_{w,\Gamma}^s(m=0)|} \quad (6)$$

donde $d_{w,\Gamma}^s(m=0)$ es el valor del marcador calculado cuando no se realiza ningún desplazamiento de los límites (Fig. 2d y 2e, en rojo), y $\sigma(d_{w,\Gamma}^s(m))$ es la desviación estándar de la serie.

Pacientes sometidos a tratamiento de HD: Los 29 registros ECGs se filtraron paso bajo (frecuencia de corte 40 Hz) para eliminar el ruido muscular y la interferencia eléctrica sin modificar la morfología de las TW y paso alto (frecuencia de corte 0,5 Hz) para atenuar la desviación de la línea de base. Los complejos QRS se detectaron y delinearon en cada derivación utilizando un método de delineación basado en la wavelet [8]. Luego, se estimó una matriz de transformación $\pi\mathcal{C}$ a partir de un extracto de la señal y se aplicó a las 8 derivaciones independientes [6]. Finalmente, para cada paciente, se delinearon todas las TW en la primera $\pi\mathcal{C}$ [8] y se realizó el análisis time warping como en [5]. Se extrajeron cinco Mean Warped TW (MWTW) en cada hora h_i $i \in \{0, 1, 2, 3, 4\}$ promediando las TW en una ventana de 2 minutos centrada en el minuto 5° de cada h_i , coincidiendo con el tiempo cuando se extrajeron las muestras de sangre para obtener el $[K^+]$ [5]. La MWTW y el valor de $[K^+]$ correspondientes a h_4 fueron, respectivamente, la referencia para el análisis time warping y para la evaluación de las variaciones relativas de $[K^+]$ ($\Delta[K^+]$). Las MWTWs y los valores de $[K^+]$ de las horas h_0 , h_1 , h_2 y h_3 fueron, respectivamente, las cuatro MWTWs para el análisis time warping y los valores de $[K^+]$ para evaluar el $\Delta[K^+] = [K^+]_{h_i} - [K^+]_{h_4}$ siendo $[K^+]_{h_i}$ la concentración de $[K^+]$ en la i -ésima hora. Para

cada paciente, se evaluó la correlación de Pearson (r) entre los cinco valores de $\Delta[K^+]$ y cada valor correspondiente de d_w , derivado con ($d_{w,\mathcal{T}}$ y $d_{w,\mathcal{D}}$) y $\sin(d_{w,\mathcal{C}})$ las WFs.

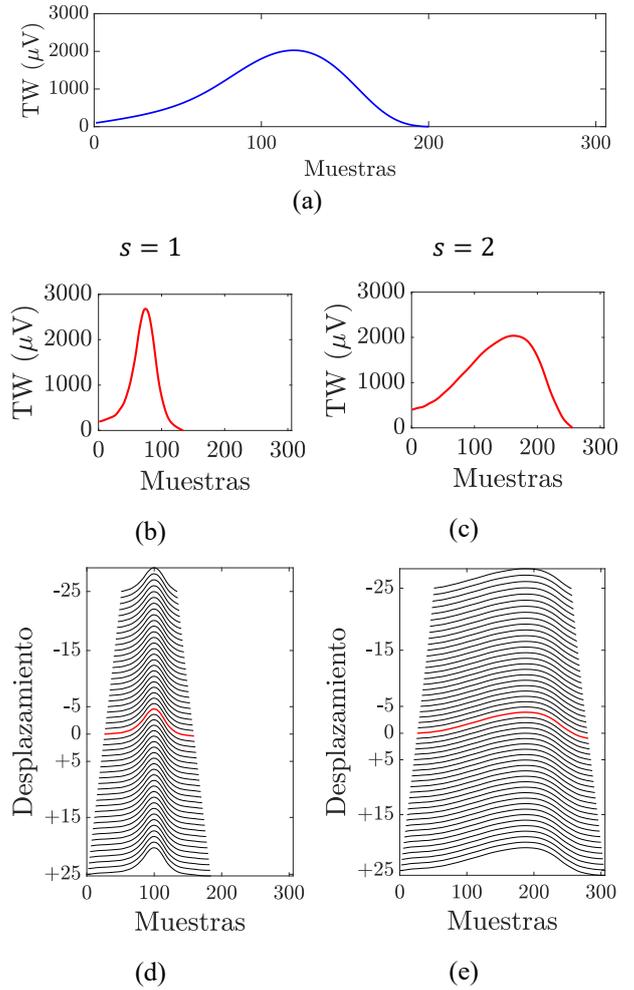


Figura 2. Modulación de la TW simulada y desplazamiento de los límites. El panel (a) muestra la TW de referencia a partir de la cual se obtienen las dos TWs moduladas, $s \in \{1, 2\}$ (paneles (b) y (c), respectivamente) después de sumar la variabilidad lineal y no lineal de duración y amplitud. Los paneles (d) y (e) muestran las TWs resultantes después de simular el desplazamiento de los límites en las dos TWs moduladas (en rojo).

4. Resultados y Discusión

Efectos de la ponderación: Se puede observar una clara reducción del valor de d_w (área verde oscuro) comparando la Fig. 1d, con respecto a las Fig. 1e y 1f, particularmente en las regiones límite (área verde claro) como consecuencia de la ponderación. La WF $\mathcal{W}_D(t)$ se introdujo ya que la función del análisis time warping generalmente se acerca a los valores cero en el pico de las TW (Fig. 1d), así que planteamos la hipótesis de que la derivada enfatizaría los centros de las dos mitades en la Fig. 1d, siendo las partes más ricas en términos de información electrofisiológica [9]. Esto resultó en una función del análisis time warping, Fig. 1f, donde se enfatizó el centro de las dos mitades, mientras que los extremos, más propensos a errores de delineación, se atenuaron.

Simulación de errores de delineación: La Tabla 1 muestra los valores de \mathcal{R} para los dos casos simulados, demostrando que la dispersión relativa con respecto a la referencia (cuando no se realiza ningún desplazamiento), se reduce significativamente cuando se aplican las WF, confirmando así su capacidad de minimizar la efectos de los errores de delineación de los límites de las TW. Además, los valores de $\mathcal{R}_{\mathcal{T}}$ para ambos $s \in \{1, 2\}$ son más pequeños que los de $\mathcal{R}_{\mathcal{D}}$, lo que sugiere un rendimiento ligeramente mejor de $\mathcal{W}_{\mathcal{T}}(t)$ con respecto a $\mathcal{W}_{\mathcal{D}}(t)$. Sin embargo, los resultados siguen siendo muy similares para sacar conclusiones concluyentes sobre cuál es la WF más adecuada para el cálculo de d_w .

s	$\mathcal{R}_{\mathcal{C}}$	$\mathcal{R}_{\mathcal{T}}$	$\mathcal{R}_{\mathcal{D}}$
1	0.17	0.05	0.07
2	0.19	0.08	0.10

Tabla 1. Ratio de variación, \mathcal{R}_{Γ}^s ; $s \in \{1, 2\}$; $\Gamma \in \{\mathcal{C}, \mathcal{T}, \mathcal{D}\}$.

Pacientes sometidos a tratamiento de HD: La Fig. 3 muestra las distribuciones $\Delta[K^+]$ y $d_{w,\Gamma}$ durante la HD en la población de estudio. Todos los marcadores siguen una tendencia decreciente pero el rango de variación disminuye claramente con el uso de la ponderación, como se muestra en la Fig. 3. Se encontraron valores medianos de correlación r similares [rango intercuartílico] para $d_{w,\mathcal{C}}$ (0,90 [0,27]), $d_{w,\mathcal{T}}$ (0,90 [0,25]) y $d_{w,\mathcal{D}}$ (0,90 [0,29]). Esto sugeriría que los efectos de los errores de delineación de los límites de las TW fueron lo suficientemente pequeños como para no afectar la capacidad de seguimiento de $[K^+]$, posiblemente porque esos errores pueden haber sido ya reducidos por el promedio realizado para calcular las MWTW [5].

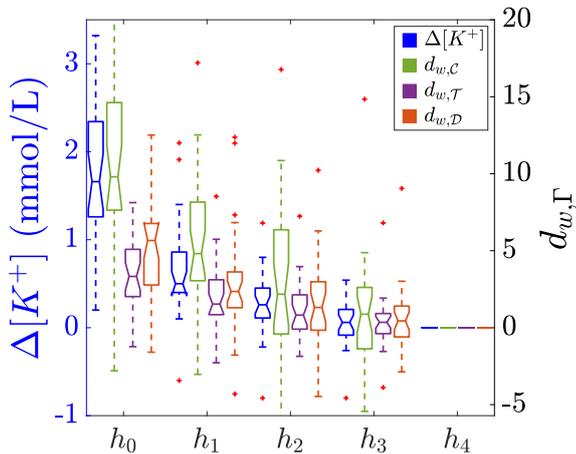


Figura 3. Distribución de $\Delta[K^+]$ (azul, eje izquierdo) y $d_{w,\Gamma}$, $\Gamma \in \{\mathcal{C}, \mathcal{T}, \mathcal{D}\}$ (verde, morado y naranja, respectivamente, eje derecho), durante la sesión de HD. Los valores atípicos se indican en rojo.

Limitaciones y líneas de investigación futuras: Si bien las simulaciones demostraron la utilidad de las WF para reducir los efectos del error de delineación de las TW en el caso del análisis time warping de las TW, no se obtuvo ninguna mejora en la monitorización de $[K^+]$ en pacientes con HD. Nuevos estudios sobre ECG en los que la

influencia de los cambios en la morfología de las TW es mínima, como en la predicción de SCD, determinarán las ventajas de utilizar las WFs propuestas en otras aplicaciones.

5. Conclusiones

Estos resultados muestran la capacidad de la ponderación para reducir los efectos no deseados de los errores de delineación de los límites de TW en el cálculo de biomarcadores como d_w , pero no se traducen en mejoras clínicas, cuando el objetivo es la monitorización de $[K^+]$. El impacto de las WF en otras aplicaciones como el valor de estratificación del riesgo arritmico de d_w queda por evaluarse.

Agradecimientos

Este trabajo fue financiado por Products & Technology S.L. (Castellbisbal) y por la Generalitat de Catalunya (España) DI001-2018 y por el CIBER-BBN ref. “DEKOALE”. El trabajo también fue apoyado por el proyecto PID2019-104881RB-I00, y PID2019-105674RB-I00 financiado por el Ministerio de Ciencia e Innovación de España (MICINN) y FEDER, por el Gobierno de Aragón (Grupo de Referencia (BSICoS) T39-20R) cofinanciado por FEDER 2014-2020 “Building Europe from Aragón”, y por el Consejo Europeo de Investigación (ERC) con el proyecto ERC-StG 638284. J. R. agradece el apoyo del programa *Horizonte 2020*, subvención Marie Skłodowska-Curie No 786833.

Referencias

- [1] Ramírez J. *et al.* Variability of ventricular repolarization dispersion quantified by time-warping the morphology of the T-wave. *IEEE TBME*, vol 64, sup 7, 2017, pp 1619–30.
- [2] Ramírez J. *et al.* T-wave morphology restitution predicts sudden cardiac death in patients with chronic heart failure *JAHA*, vol 6(e005310), 2017, pp. 1–12.
- [3] Madeiro J. P. V. *et al.* New approach for T-wave peak detection and T-wave end location in 12-lead paced ECG signals based on a mathematical model. *Med Eng Physics*, vol. 35, sup. 8, 2013, pp 1105–15.
- [4] García J. *et al.* ECG-based detection of body position changes in ischemia monitoring. *IEEE TBME* vol. 50, sup 6, 2004, pp 677–685.
- [5] Palmieri F. *et al.* Monitoring blood potassium concentration in hemodialysis patients by quantifying T-wave morphology dynamics. *Sci Rep*, vol 11, 2021, Art. no. 3883
- [6] Palmieri F. *et al.* ECG-based monitoring of blood potassium concentration: Periodic versus principal component as lead transformation for biomarker robustness, *Biomedical Signal Processing and Control*, vol. 68. 2021, Art. no. 102719.
- [7] Recommendations for measurement standards in quantitative electrocardiography. *Eur Heart J*, vol 6, sup. 10, 1985, pp. 815–825.
- [8] Martínez J. P. *et al.* A wavelet-based ECG delineator: Evaluation on standard databases. *IEEE TBME*, vol. 4, sup. 51, 2004, pp. 570–581.
- [9] Meijborg V. M. F. *et al.* Electrocardiographic T wave and its relation with ventricular repolarization along major anatomical axes. *Circ*, vol 7, 2014, pp. 524-531.

Contacto: Flavio Palmieri, UPC CREB, Av. Diagonal, 647, 08028, Barcelona (España). flavio.palmieri@upc.edu

Diseño e implementación de un sistema de simulación para el manejo de la fractura de pelvis

B. Larraga García¹, M. Rodríguez Rodríguez¹, J. Rubio Bolívar², M. Quintana Díaz², A. Gutiérrez Martín¹

¹ Departamento de Fotónica y Bioingeniería, ETSI de Telecomunicación, Universidad Politécnica de Madrid, Madrid.

España, {myriam.rodriguez.rodriguez}@alumnos.upm.es, {blanca.larraga, a.gutierrez}@upm.es

² Instituto de Investigación IdiPAZ, Hospital Universitario La Paz, Madrid. España. {javier84.rubio, mquintanadiaz}@gmail.com

Resumen

El trauma es la principal causa de muerte en personas menores de 45 años y una de las principales causas de muerte en el mundo. Así pues, cuando se produce un traumatismo, las maniobras de estabilización inicial son de gran importancia. Dentro de las lesiones de trauma, la fractura de pelvis es una de las lesiones con mayor porcentaje de mortalidad asociada a hemorragias y posibles daños en órganos que pueden comprometer la vida del paciente. Se ha observado en la literatura que el uso de cinturones pélvicos, como método de estabilización inicial, reduce el sangrado producido por el trauma y a su vez, sirve como inmovilizador. Sin embargo, la incorrecta colocación del cinturón pélvico podría disminuir la eficacia de su uso e incluso, empeorar el pronóstico. El objetivo de este trabajo consiste en diseñar un sistema de simulación capaz de reproducir diferentes tipos de fracturas pélvicas y proporcionar información sobre las maniobras de estabilización que se lleven a cabo.

1. Introducción

El traumatismo es un término general que se emplea para designar a todas las lesiones internas o externas provocadas por un impacto exterior. La enfermedad traumática grave (ETG) es un problema de salud mundial y una de las principales causas de muerte y discapacidad, sobre todo en adultos jóvenes [1]. El pronóstico de la ETG depende de varios factores, como la gravedad y la energía del traumatismo, la condición fisiológica del paciente y la calidad y rapidez de la atención prestada.

El traumatismo pélvico presenta en España una incidencia anual de 23 personas por cada 100.000 habitantes. A nivel nacional, las fracturas de pelvis (FP) representan entre un 3% y un 8% de todas las lesiones esqueléticas [2] y las presentan aproximadamente el 20% de los politraumatizados graves [3].

Las causas más comunes de fractura de pelvis son:

- **Impactos intensos:** Traumatismos de alta energía tales como accidentes de tráfico (50 %), atropellos (20 %), precipitaciones o caídas de altura (5-20 %).
- **Caídas moderadas:** Traumatismos de baja energía que normalmente acontecen en personas de edad avanzada o mala calidad ósea por enfermedades como la osteoporosis.

La mortalidad de la FP depende, de forma sustancial, de la gravedad de la fractura y de las lesiones asociadas. Según un estudio de la Corporación Sanitaria Parc Taul en Sabadell, España tiene una mortalidad entre el 19% y el 33%, llegando incluso al 40-60% en pacientes con inestabilidad hemodinámica y hasta un 70% si la fractura es

abierta [2]. Las hemorragias son la principal causa de muerte en el 42% de los traumatismos pélvicos [3].

Por lo tanto, la importancia de estabilizar las fracturas pélvicas radica en el control del shock hemorrágico derivado de la hemorragia que produce la lesión. Los cinturones pélvicos son un sistema de estabilización provisional empleados en el ámbito prehospitalario. Su finalidad es controlar el sangrado atribuible a la inestabilidad pélvica y contenerlo mediante un mecanismo de autotaponamiento [4].

Así pues, el objetivo de este trabajo consiste en diseñar e implementar un sistema de simulación que permita emular diferentes fracturas pélvicas y permita entrenar en la estabilización de las mismas.

El artículo se organiza de la siguiente manera: en la Sección 2 se presenta el diseño hardware realizado. La Sección 3 explica el diseño software. En la Sección 4 se presenta la validación del sistema desarrollado. Por último, la Sección 5 presenta las conclusiones.

2. Diseño hardware

La pelvis o anillo pélvico forma la base del tronco del cuerpo humano, constituyendo el sostén del abdomen y llevando a cabo la unión entre los miembros inferiores y el resto del cuerpo. Se compone de tres piezas óseas: dos huesos coxales y el sacro; y de tres articulaciones: dos sacroilíacas y una sínfisis del pubis. Tal y como se muestra en la Figura 1, cada hueso coxal está compuesto a su vez por tres huesos que son el ilión (destacada la cresta ilíaca en la Figura), el pubis y el isquion. Adicionalmente, el coxis es la parte ósea a continuación del sacro [5].

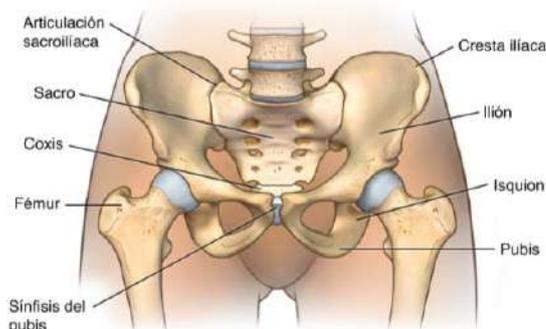


Figura 1. Huesos y articulaciones de la pelvis. [6]

Pueden darse diferentes tipos de fracturas en la pelvis y existen varias clasificaciones de las mismas. Según el sistema de clasificación Tile [7], las fracturas de pelvis se dividen en:

- **Fracturas tipo A:** fracturas estables.
- **Fracturas tipo B:** fracturas que causan inestabilidad rotacional siendo parcialmente inestables.
- **Fracturas tipo C:** fracturas que producen inestabilidad rotacional y vertical. Estas fracturas son inestables.

2.1. Diseño pelvis en 3D

Para el diseño de la estructura pélvica se ha utilizado el software Blender y se ha impreso utilizando una impresora Original Prusa i3 MK3S. Para poder reproducir las diferentes fracturas pélvicas, se ha dividido la pelvis en tres partes que corresponden con las tres partes óseas de la pelvis: los dos huesos coxales y el sacro tal y como se observa en la Figura 2.



Figura 2. Huesos de la pelvis en 3D.

Para poder llevar a cabo el simulador que permita emular los diferentes tipos de fracturas pélvicas y entrenar técnicas de estabilización como la colocación del cinturón pélvico, se incorporarán sensores y actuadores. Estos dispositivos se deberán incorporar en los huesos de la pelvis diseñados en 3D tal y como mostraremos en la sección de validación del sistema.

2.2. Sensores y actuadores

Con el objetivo de reproducir las diferentes fracturas pélvicas y sensorizar lo que ocurre durante la estabilización de las mismas, es necesario un sistema de instrumentación formado por actuadores y sensores.

Los actuadores escogidos son los electroimanes que se muestran en la Figura 3. Estos electroimanes, al hacer circular corriente a través de las espiras conductoras de la bobina, generan un campo magnético en su interior. Si enfrentamos una placa metálica al electroimán, se consigue la unión y desunión de ambas piezas en función de la corriente.



Figura 3. Electroimán circular de 20x15 mm [8].

Cuando se produce una fractura pélvica y se procede a su estabilización mediante la colocación del cinturón pélvico, se pretende evaluar dos parámetros: el posible movimiento de los huesos fracturados tras la colocación del cinturón

pélvico y la presión ejercida. Para medir este movimiento se utilizará el sensor de efecto Hall que se muestra en la Figura 4 junto con un imán de neodimio.

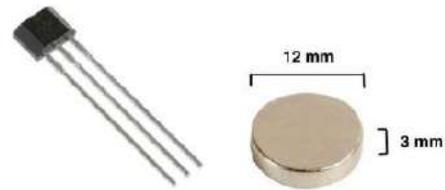


Figura 4. Sensor de efecto Hall SS495A [9] e imán de neodimio N40 de 12 mm de diámetro [10].

Para medir la presión ejercida a través de la colocación del cinturón pélvico se utilizará un sensor resistivo FSR circular como el que se muestra en la Figura 5.

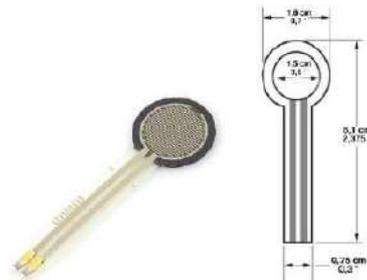


Figura 5. Sensor de fuerza resistivo SEN-09375 [11].

2.3. Microcontrolador

El microcontrolador seleccionado es un Arm Cortex-M0 de 32-bits que puede operar a una frecuencia de 48 MHz y que tiene integrado soporte de comunicación WiFi y Bluetooth. Este microcontrolador permitirá gestionar las señales de los sensores y actuadores así como desarrollar una interfaz gráfica que permita visualizar al usuario lo que está ocurriendo en la simulación.

En la Figura 6 se presenta la conexión de los sensores y actuadores con el microcontrolador.

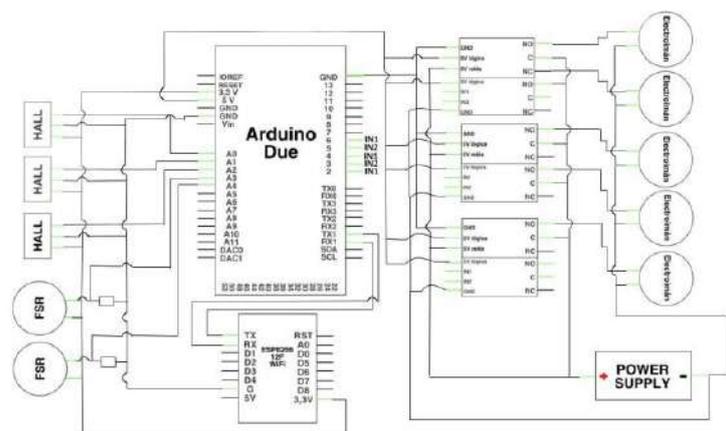


Figura 6. Esquema de conexiones del circuito.

3. Diseño software

A continuación se diseña la arquitectura necesaria para establecer el intercambio de información entre el sistema de sensores y actuadores y la aplicación web que permitirá a los profesionales interactuar con el simulador.

Los requisitos necesarios para desplegar el simulador son:

- Protocolo de comunicación con poca latencia de envío de datos.
- Bidireccionalidad para enviar y recibir información.
- Plataforma de usuario atractiva y que permita visualizar en tiempo real.
- Almacenar los datos para tener un registro para análisis posteriores.

3.1. Plataformas utilizadas

Para poder cumplir los requisitos presentados más arriba, las plataformas utilizadas son las siguientes: Docker, Node-RED, Mosquitto y comunicación MQTT, Thingsboard y MongoDB. En la Figura 7 se muestra la arquitectura del sistema completo.

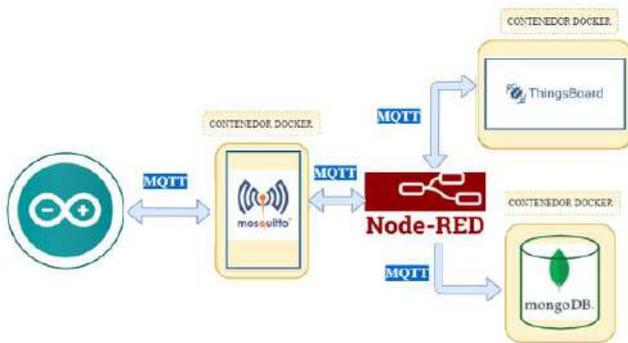


Figura 7. Arquitectura del sistema.

Docker se emplea para contextualizar cada herramienta con sus dependencias dentro de un contenedor y tener el sistema lo más modular posible. Node-RED se utiliza como panel visual de control, para interconectar todo el hardware con el resto de servicios. MQTT es el protocolo de comunicación que cumple con la bidireccionalidad de forma eficiente y es compatible con el microcontrolador utilizado y con Thingsboard (TB). TB es una plataforma IoT de código abierto que permite visualizar y analizar los datos de los dispositivos conectados al microcontrolador, así como realizar acciones sobre dichos dispositivos. La interfaz gráfica se programa con TB. Toda la información referente a los sensores queda registrada en la base de datos MongoDB.

El flujo de datos es bidireccional:

- El microcontrolador se conecta al broker de Mosquitto y publica la información de cada sensor en un topic (tema) diferente. Node-RED se suscribe a dicho servidor y filtra la información por cada topic en un nodo diferente para cada sensor. A continuación, Node-RED redirecciona los datos tanto a MongoDB, para tener almacenados todos los datos; como a ThingsBoard con el objetivo de visualizarlos.
- El flujo de recepción de datos comienza en el broker interno de ThingsBoard. Node-RED se suscribe a dicho broker, y envía las órdenes al broker de Mosquitto. Una vez los mensajes llegan

a este servidor, el cliente del microcontrolador se suscribe a los diferentes topics para recibir los mensajes.

3.2. Interfaz gráfica

La interfaz gráfica desarrollada presentará dos opciones en función del perfil de la persona que acceda a este simulador. Por un lado puede acceder el entrenador, que será quién establezca el escenario de simulación y por otro lado, el estudiante que será quién realice la estabilización de la fractura.

Si accedemos como entrenador, la pantalla que veremos es la que se muestra en la Figura 8.



Figura 8. Interfaz gráfica desarrollada para el entrenador.

En esta pantalla se pueden observar los diferentes tipos de fracturas que se pueden generar. Adicionalmente, se puede comprobar si el cinturón se ha colocado o no por parte del estudiante y, en la parte superior derecha, aparece un círculo en gris. En caso de que el cinturón esté colocado, este círculo aparecerá en verde mientras que si el cinturón no está colocado, el círculo aparecerá en rojo. Finalmente, en la parte inferior aparece la información recogida de los diferentes sensores; los de presión en la parte izquierda y los sensores de efecto Hall en la parte derecha.

Si accedemos como estudiante, la pantalla que veremos es la que se muestra en la Figura 9.



Figura 9. Interfaz gráfica desarrollada para el estudiante.

En esta pantalla se verá, una vez el instructor haya generado un tipo de fractura, si ya se ha generado o no la fractura para comenzar a tratar la lesión. Posteriormente, se detectará si el cinturón se ha colocado y se mostrará en rojo o verde en función de si se ha puesto el cinturón pélvico o

no. Finalmente, aparecerá información acerca de la presión ejercida en la maniobra.

4. Validación del sistema

Tras realizar el diseño y la implementación de todos los componentes hardware, la implementación de toda la arquitectura y el diseño de la interfaz gráfica, se procede a validar el sistema.

En la Figura 10 se muestra la pelvis impresa en 3D y se indica donde están insertados los cinco electroimanes así como los relés que se encuentran numerados con el mismo dígito del electroimán sobre el que actúa. Con estos electroimanes se generan roturas tipo B y tipo C según la clasificación de Tile. Cada tipo de rotura se subdivide en 1, 2 ó 3 por lo que se podrán seleccionar fracturas tipo B1, B2, B3 ó C1, C2, C3 en función de la inestabilidad de las mismas.



Figura 10. Conjunto formado por los huesos de la pelvis incluyendo electroimanes y relés.

Una fractura tipo B1 que es parcialmente inestable, conlleva una separación de la articulación sacroilíaca y de la sínfisis del pubis. Para ello, según los actuadores que se ven en la Figura 10, se cesa la alimentación de los electroimanes E3 y E5. En la Figura 11 podemos ver que, efectivamente se produce la separación entre los huesos correspondientes una vez se selecciona el tipo de fractura en la interfaz gráfica y que la información llega al simulador.

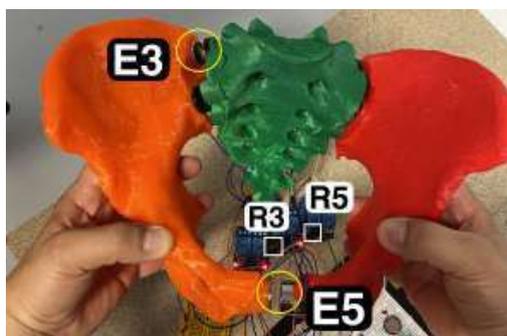


Figura 11. Fractura tipo B1 generada.

Adicionalmente, se valida que se recibe información de los sensores de presión y de los sensores de efecto Hall en la interfaz gráfica. Para ello, se ha probado a ejercer presión directamente sobre los sensores y a mover los huesos de la pelvis, recibiendo información de los sensores tal y como se muestra en la Figura 12.



Figura 12. Interfaz gráfica donde se puede ver la fractura generada así como la información recibida por los sensores.

5. Conclusiones

El principal objetivo de este trabajo ha sido el diseño e implementación de un sistema de simulación que permita crear diferentes fracturas pélvicas para el entrenamiento de habilidades y técnicas necesarias para atender a pacientes de trauma y recabe información acerca de las actuaciones de los clínicos en formación. Como líneas futuras, se deberá incorporar este simulador en una estructura que permita colocar un cinturón pélvico y realizar la maniobra completa de estabilización.

Referencias

- [1] M Chico-Fernández y col. *Epidemiology of severe trauma in Spain. Registry of trauma in the ICU (RETRAUCI). Pilot phase.* Medicina Intensiva (English Edition) 40.6 (2016), págs. 327-347.
- [2] Fernando Martínez y col. *Pelvic fracture in the patient with multiple injuries: factors and lesions associated with mortality.* Emergencias 30.2 (2018), págs. 91-97.
- [3] Juan María Sánchez-Tocino y col. *Fracturas de pelvis graves, lesiones asociadas e inestabilidad hemodinámica: incidencia, manejo y pronóstico en nuestro medio.* Cirugía Española 81.6 (2007), págs. 316-323.
- [4] Michael Bottlang y col. *Emergent management of pelvic ring fractures with use of circumferential compression.* En: JBJS 84.suppl 2 (2002), S43-S47.
- [5] M Schünke, E Schulte y U Schumacher. *Prometheus: Texto y atlas de anatomía. Tercera Edición.* En: Madrid: Editorial Médica Panamericana (2014), págs. 440-527.
- [6] Datos acerca de la espina dorsal, el hombro y la pelvis. URL: <https://www.stanfordchildrens.org/es/topic/default?id=facts-aboutthepineshoulderandpelvis-85-P04460>; 2021.
- [7] Medigraphic. Clasificación de las fracturas de pelvis. PDF: <https://www.medigraphic.com/pdfs/orthotips/ot-2008/ot084e.pdf>, 2008.
- [8] Electroimán de succión circular 20x15mm. url: https://www.amazon.es/gp/product/B08FBCTF49/ref=ppx_yo_dt_b_asin_title_o01_s00?ie=UTF8&psc=1.
- [9] Sensor Hall SS495A Honeywell. url: <https://es.rs-online.com/web/p/sensores-de-efecto-hall/>
- [10] Imán de Neodimio N40. url: <https://www.electronicaembajadores.com/es/Productos/Detalle/IM50N06/iman-y-electroimanes/iman-neodimio-convencional/iman-de-neodimio-disco-o12-x-3-mm-n40>.
- [11] Sensor FSR SEN-09375 Sparkfun. url: <https://www.sparkfun.com/products/9375>.

Cambio de FiO_2 como alternativa al gadolinio para estudios de perfusión cerebral por resonancia magnética

Moreno-Castán A.¹, Calle D.², Desco M.^{1,2,3,4}, Cussó L.^{1,2,3,4}

¹ Universidad Carlos III de Madrid, Madrid, España

² Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC); Madrid, España

³ Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, España (dcalle@hggm.es, desco@hggm.es, lcusso@hggm.es)

⁴ Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Madrid, España

Resumen

El contraste dinámico por susceptibilidad (DSC) es la técnica de valoración de la perfusión cerebral por resonancia magnética más extendida actualmente y requiere la administración de gadolinio, que se sabe puede desencadenar efectos secundarios nocivos. Este trabajo desarrolla una nueva técnica que permite sustituir el gadolinio por cambios de la FiO_2 mediante la inhalación de nitrógeno. Para validar este método, se realizó un estudio de imagen en roedores con modelo de ictus y se procesaron las imágenes adquiridas. Los resultados confirmaron que la nueva técnica es una alternativa viable a la evaluación estándar con gadolinio.

1. Introducción

La perfusión por resonancia magnética (PRM) permite el estudio del flujo sanguíneo en los capilares cerebrales. Esta técnica tiene valiosas aplicaciones en la práctica clínica, entre las que destacan el diagnóstico de ictus, tumores cerebrales y trastornos neurológicos como el Alzheimer [1]. Existen diversos métodos de PRM, siendo uno de los más comunes el de contraste dinámico por susceptibilidad (DSC). Este método requiere la inyección de un agente externo de contraste cuya susceptibilidad magnética altere el valor de campo produciendo una disminución de la señal de RM [2]. Gracias a sus propiedades paramagnéticas, los quelatos de gadolinio son los agentes de contraste más usados en PRM [3]. Así, mediante la adquisición rápida de imágenes durante el paso de gadolinio a través de los capilares cerebrales se pueden apreciar cambios en la intensidad de la señal en función del tiempo (figura 1), dando lugar a curvas a partir de las cuales se extraen diversos parámetros de perfusión.

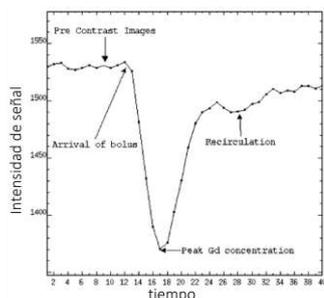


Figura 1. Curva de intensidad de RM característica de DSC [4]

Sin embargo, es bien conocido que la administración de gadolinio en pacientes puede producir efectos secundarios leves tales como náuseas, mareos, cefaleas, erupciones cutáneas, etc. [5,6]. En algunas ocasiones, los efectos secundarios pueden ser de gravedad, como es en caso de alergia o en pacientes con insuficiencia renal, en los que la administración de gadolinio se identifica como desencadenante de la fibrosis sistémica nefrogénica [7]. Por otra parte, en investigación preclínica –especialmente en el trabajo con roedores–, la administración intravenosa de gadolinio es compleja técnicamente y debe ser realizada por personal especializado [8]. La alternativa propuesta consiste en utilizar variaciones en la fracción inspirada de oxígeno (FiO_2) para sustituir la acción del contraste. Una disminución de la FiO_2 se traduce en un menor nivel de oxígeno en sangre, que conlleva un aumento de la proporción de desoxihemoglobina (paramagnética) respecto a la de oxihemoglobina (diamagnética) [9]. Este aumento tendrá un efecto similar a la inyección de gadolinio, es decir, una disminución de la señal de RM. El cambio de FiO_2 se obtiene ventilando al sujeto con 100% de nitrógeno durante un tiempo, para después volver rápidamente a una FiO_2 de 100%, lo cual genera un perfil de intensidad de RM análogo al de la figura 1.

El objetivo de este trabajo fue desarrollar esta nueva técnica de DSC, sustituyendo el contraste exógeno por una reducción artificial temporal de la FiO_2 .

2. Material y Métodos

El trabajo se organiza en dos partes: desarrollo de una herramienta de cuantificación y validación de la técnica FiO_2 mediante estudios animales in vivo.

2.1. Desarrollo de la herramienta de cuantificación

La perfusión no se puede evaluar con las imágenes en crudo, por lo que se desarrolló una herramienta en MATLAB que implementa el procesamiento necesario para la obtención de mapas de perfusión, a través de varios pasos.

En primer lugar, se realiza un alineamiento espacial de las imágenes utilizando un registro espacial basado en intensidad utilizando el error cuadrático medio como función de coste. A continuación, se elimina el fondo de la imagen con una segmentación del cerebro mediante el

método Otsu [10]. El resultado de la segmentación se mejora utilizando morfología matemática (cierre + apertura). Finalmente, se aplica un filtro de media móvil para obtener un suavizado temporal de las señales de los distintos píxeles y mejorar la relación señal-ruido.

El siguiente paso es la selección de la función de entrada arterial (AIF, del inglés “arterial input function”). La AIF se define como la concentración de agente de contraste en la arteria de entrada y es necesaria para calcular varios parámetros de perfusión [11]. Para ello, se seleccionan los píxeles de la imagen que cumplen determinados criterios: (1) mínimo más temprano; (2) mayor altura; (3) anchura a altura media (FWHM, del inglés “full width at half maximum”) más similar a la curva de saturación de oxígeno; y (4) menor nivel de ruido.

A continuación, es necesario implementar un método que permita deconvolucionar la AIF de cada una de las curvas para obtener la función residual, que describe la retención del agente de contraste en cada zona del tejido y está directamente relacionada con los parámetros de perfusión cerebral [11,12]. A diferencia del procesamiento convencional de estudios con gadolinio, este paso es imprescindible debido a la elevada anchura de la función de entrada que se obtiene con el nuevo método. Esta deconvolución se traduce en resolver un problema inverso mal condicionado, para lo cual se utiliza una descomposición en valores singulares (SVD) con regularización de Tikhonov [13]. Para seleccionar el parámetro de regularización se implementa la curva L con el método desarrollado en [14].

El último paso es la extracción de los diferentes parámetros cuantitativos para generar los mapas de perfusión. En este caso, se calcularon el área bajo la curva (AUC), la altura y la FWHM, parámetros relacionados con el flujo de sangre en el cerebro y el tiempo de tránsito medio de la sangre en los capilares cerebrales.

2.2. Estudio de imagen in vivo y validación

Se emplearon 4 ratas Wistar hembras adultas a las cuales se les realizó un estudio de PRM basal y otro tras la inducción de un ictus en el hemisferio izquierdo del cerebro mediante la oclusión de la arteria cerebral media [15] (procedimientos ajustados a la Directiva de la UE 2010/63EU y a la normativa nacional RD 53/2013, PROEX 241-19). Las imágenes fueron adquiridas con un escáner preclínico de 7 Tesla (BioSPec 70/20 USR, Bruker, Alemania) y se emplearon dos secuencias distintas: EPI (TR/TE: 1000/7.710 ms, tiempo de muestreo 1 s/rep, 20 cortes de 1 mm de grosor, n° repeticiones 140, FOV 30x25 mm y 92x50 píxeles) y True FISP (TR/TE: 2.550/1.275 ms, tiempo de muestreo 1.058 s/rep, 7 cortes de 1 mm de grosor, n° repeticiones 140, FOV 30x25 mm y 92x50 píxeles).

Durante las adquisiciones, la FiO_2 se modificaba usando un mezclador de gases GSM-3 (CWE Inc., Estados Unidos). Inicialmente, el animal inspiraba 100% oxígeno y tras 20 segundos de adquisición se modificaba el mezclador a 100% nitrógeno (ambas mezclas a un flujo de 1000 mL/min y al 2% de sevoflurano). Después de 60

segundos de nitrógeno, se cambiaba de nuevo a 100% oxígeno. Todos estos cambios se reflejaban en la saturación de oxígeno, monitorizada en todo momento usando un pulsioxímetro magneto-compatible para animal pequeño (Small Animal Instruments Inc., Estados Unidos). Además de las adquisiciones con FiO_2 , se adquirió una secuencia EPI con gadolinio (0.3 mL ProHance inyectado por la vena de la cola a 2.4 mL/min) como técnica estándar de control.

Las imágenes adquiridas durante el estudio in vivo fueron procesadas con la herramienta desarrollada. Los mapas obtenidos con MATLAB fueron analizados dibujando una región de interés tanto de la zona sana como la infartada con Image J (1.53e). Se calculó la ratio de los valores medios de cada hemisferio y se realizó una prueba t de Student con los datos obtenidos.

3. Resultados

Observando la evolución de las imágenes en crudo obtenidas con FiO_2 (Figura 2), se puede apreciar que su intensidades medias disminuyen para después volver a su valor original, tal y como ocurre con el gadolinio. Como se esperaba, la curva de FiO_2 es más ancha y presenta una menor caída de señal RM que la curva de gadolinio. Sin embargo, tiene la ventaja de que la señal con FiO_2 vuelve a su valor original, mientras que la de gadolinio está dominada por la recirculación y no recupera su intensidad inicial. Además, se aprecian visualmente diferencias entre las curvas del hemisferio infartado y el sano, siendo la caída de intensidad en el primero menor que en el segundo.

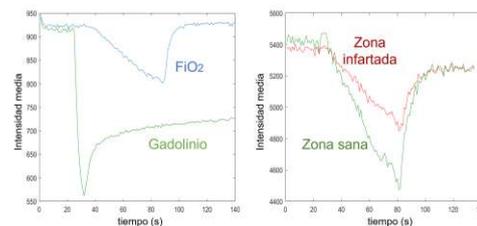


Figura 2. Intensidad media de la imagen de FiO_2 vs gadolinio (izquierda) y de la zona sana vs infartada con FiO_2 (derecha).

También es notable la diferencia de resolución espacial entre las imágenes obtenidas con las secuencias EPI y True FISP (Figura 3). Las dos son secuencias rápidas por lo que la resolución espacial es limitada con ambas, pero True FISP ofrece mejor resolución con un tiempo de adquisición similar.



Figura 3. Vista axial del cerebro del mismo animal adquirida con EPI (izquierda) y True FISP (derecha).

La figura 4 muestra los mapas de perfusión obtenidos tras el procesamiento de las imágenes. Se muestran los mapas de AUC, altura y FWHM para uno de los cortes de un animal durante la oclusión. Tanto con EPI como con True

FISP, se puede apreciar una disminución de los valores en el hemisferio izquierdo –zona infartada– similar a la que se aprecia con gadolinio. En general, los mapas generados con True FISP delimitan la extensión y localización del ictus de forma más clara y similar a la adquisición estándar con gadolinio que los generados con EPI. Los mapas de AUC y altura muestran con detalle las distintas zonas del cerebro, mientras que los de FWHM son ruidosos y presentan mala resolución espacial.

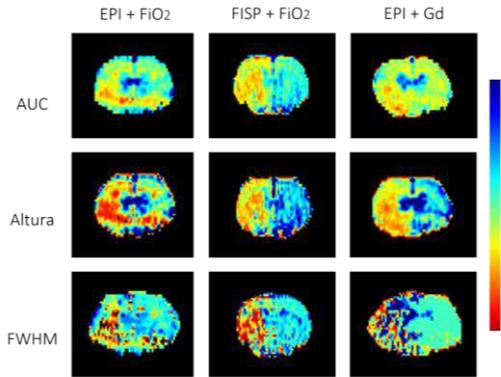


Figura 4. Selección de los mapas de perfusión durante la oclusión. Las distintas filas muestran los mapas de AUC, altura y FWHM. Cada una de las columnas corresponde a una técnica de adquisición. Los valores más bajos se representan en rojo.

En la figura 5 se muestra la ratio del valor medio del mapa en la zona infartada respecto al de la zona sana para cada PRM. Los valores de AUC y altura son similares tanto si se emplea la FiO₂ como el gadolinio, lo cual no ocurre con el FWHM, donde se aprecia una diferencia entre ambos contrastes. En general, las ratios son menores que 1 lo que implica valores disminuidos en la zona infartada respecto a la zona sana. No hay diferencias significativas ($p < 0.05$) entre los valores de AUC y altura entre las secuencias con FiO₂ y las de gadolinio, mientras que sí observamos diferencia en el valor de FWHM.

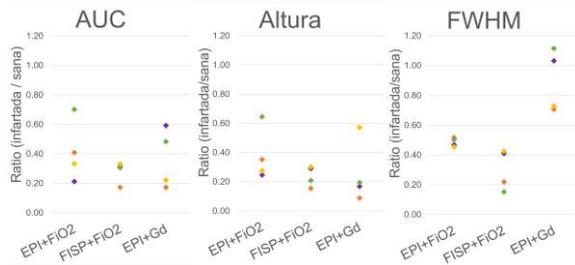


Figura 5. Ratios de valor medio para los distintos parámetros y adquisiciones (EPI y True FISP, FiO₂ o gadolinio). Cada punto representa un animal. Cada animal se representa con un color.

4. Discusión

El objetivo de este trabajo era desarrollar una nueva técnica de DSC que permitiese sustituir un contraste exógeno como el gadolinio por una reducción artificial de la FiO₂ mediante la inhalación de nitrógeno. Hemos demostrado que, aunque las formas de las señales obtenidas con ambas técnicas no son idénticas (Figura 2), la reducción de la FiO₂ analizada mediante deconvolución permite detectar el cambio de intensidad de la señal con

precisión suficiente para diferenciar las áreas sanas de las áreas infartadas, cumpliendo además los requisitos necesarios para extraer de ellas los parámetros de perfusión.

Uno de los pasos más cruciales para obtener los parámetros de perfusión fue la deconvolución de la señal. Este proceso es muy sensible al ruido de los datos por lo que en el desarrollo de la herramienta de postproceso se dedicó especial atención al suavizado de la señal, así como a la optimización del criterio de selección de la AIF. A pesar del procesado previo de la señal, no se consiguió eliminar completamente el ruido, por lo que se generó una inestabilidad en el resultado de la deconvolución dando lugar a una serie de oscilaciones. El método de deconvolución elegido –Tikhonov con la curva L– fue capaz de reducir el nivel de estas oscilaciones sin alterar la exactitud de los resultados.

Los mapas de perfusión generados a partir de las señales producto de la deconvolución (figura 4) muestran que el método Tikhonov es adecuado para deconvolucionar la señal. Mientras que en los mapas basales los dos hemisferios muestran valores de perfusión similares, en todos los mapas con ictus el área infartada es claramente detectable como una zona de perfusión reducida, tanto en las adquisiciones con EPI como con True FISP. Por tanto, ambas secuencias fueron capaces de proporcionar los resultados deseados. Sin embargo, cabe destacar que los mapas generados con la secuencia True FISP se asemejan más a los obtenidos con la técnica estándar de gadolinio que los generados con EPI. En general, True FISP ofrece mayor resolución espacial (figura 3), dando lugar a mapas más fácilmente interpretables que delimitan mejor el área infartada. Por tanto, esta secuencia parece preferible sobre la secuencia EPI para la técnica con FiO₂, teniendo en cuenta que el tiempo de adquisición de ambas es similar. Estas observaciones cualitativas se refieren a los mapas de altura y AUC, puesto que los de FWHM no ofrecieron resultados concluyentes. Los resultados cuantitativos y el análisis estadístico apoyan la valoración cualitativa anteriormente presentada.

Así, este estudio presenta resultados positivos sobre el uso de la FiO₂ como contraste para estudios de PRM. La relación entre los cambios en la saturación de oxígeno y la señal de RM (técnica BOLD, “blood oxygenation level-dependent”) se ha estudiado en numerosas ocasiones con el objetivo de evaluar la actividad cerebral o el consumo de oxígeno [16]. Sin embargo, la aplicación de esta técnica al estudio de la perfusión cerebral es un tema menos explorado hasta el momento en la literatura. Algunos autores [17-19] han relacionado las alteraciones de la FiO₂ y la consecuente variación de la señal de RM con eventos hemodinámicos y parámetros como el volumen de sangre cerebral. MacDonald M.E. et al. [20] y Vu C. et al. [21] utilizaron el modelo DSC junto con cambios en la saturación de oxígeno para generar mapas de perfusión a partir de datos obtenidos de pacientes sanos y con anemia crónica. En estos trabajos, las imágenes se adquirieron con la secuencia EPI y se procesaron ajustando las señales a una combinación de modelos lineales y exponenciales. Nuestro trabajo

representa un avance respecto a los anteriores constituyendo la primera validación en un modelo animal de ictus e introduce una nueva secuencia, True FISP, con resultados satisfactorios que no requieren ajuste de las señales a modelos cinéticos para generar mapas de perfusión significativos. Sin embargo, este es un estudio exploratorio, por lo que aumentar el número de animales o examinar otros modelos cardiovasculares será necesario para reforzar nuestros hallazgos.

5. Conclusión

La técnica de FiO_2 propuesta en este trabajo ha sido validada como una alternativa viable al método de DSC utilizado actualmente, eliminando la necesidad de la inyección de gadolinio y los inconvenientes derivados de esta. Además, la herramienta de procesamiento desarrollada en este trabajo permite obtener mapas de perfusión precisos con la técnica de FiO_2 , comparables a los obtenidos con la técnica estándar de gadolinio, con valores significativos de perfusión capaces de detectar una zona afectada por ictus. Estos mapas presentan una mayor calidad para las imágenes adquiridas con la secuencia True FISP que para la secuencia EPI.

Agradecimientos

Agradecemos a Alexandra de Francisco y María de la Jara Felipe su excelente trabajo en la preparación de los animales y la obtención de imágenes. Este trabajo ha sido parcialmente subvencionado por el ISCIII (PT20/00044), cofinanciado por el Fondo Europeo de Desarrollo Regional, “Una manera de hacer Europa” y por la Comunidad de Madrid a través de la línea “Estímulo a la Investigación de Jóvenes Doctores” del Convenio Plurianual con la UC3M (CSFFLOW-CM-UC3M), en el marco del V PRICIT (V Plan Regional de Investigación Científica e Innovación Tecnológica).

Referencias

- [1] J.R. Petrella, J.M. Provenzale. MR Perfusion Imaging of the Brain: Techniques and Applications, *American Journal of Roentgenology*, vol.175, pp.207-219, 2000
- [2] G.H. Jahng, K.L. Li, L. Ostergaard, F. Calamante. Perfusion Magnetic Resonance Imaging: A Comprehensive Update on Principles and Techniques, *Korean Journal of Radiology*, vol. 15, no. 5, p. 554, 2014
- [3] Val M. Runge, Jeffrey A. Clanton, Ann C. Price, Chris J. Wehr, William A. Herzer, C. Leon Partain, A. Everette James, The use of GD DTPA as a perfusion agent and marker of blood-brain barrier disruption. *Magnetic Resonance Imaging*, vol.3, no.1, pp.43-55, 1985
- [4] P. Keston, A. Murray, A. Jackson. Cerebral perfusion imaging using contrast-enhanced MRI, *Clinical Radiology*, vol. 58, no. 7, pp. 505–513, 2003
- [5] A. F. Delgado et al., Diagnostic value of alternative techniques to gadolinium-based contrast agents in MR neuroimaging – a comprehensive overview, *Insights into Imaging*, vol. 10, no. 1, pp. 1–15, 2019
- [6] M. Llamas. Gadolinium side effects: Toxicity; nephrogenic systemic fibrosis. Nov 2020. <https://www.drugwatch.com/>
- [7] Y. Wang et al., Incidence of nephrogenic systemic fibrosis after adoption of restrictive gadolinium-based contrast agent guidelines, *Radiology*, vol. 260, no. 1, pp. 105–111, 2011
- [8] M. Resch, T. Neels, A. Tichy, R. Palme, T. Rüllicke, Impact assessment of tail-vein injection in mice using a modified anaesthesia induction chamber versus a common restrainer without anaesthesia, *Laboratory animals*, vol. 53, no. 2, pp. 190–201, 2019
- [9] R.B. Buxton. The physics of functional Magnetic Resonance Imaging (fMRI), *Reports on Progress in Physics*, vol. 76, no. 9, 096601, 2013
- [10] N. Otsu, A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979
- [11] A. Fieselmann, M. Kowarschik, A. Ganguly, J. Hornegger, R. Fahrig. Deconvolution-based CT and MR brain perfusion measurement: Theoretical model revisited and practical implementation details, *International Journal of Biomedical Imaging*, vol.2011, 2011
- [12] L. Østergaard, Principles of cerebral perfusion imaging by bolus tracking, *Journal of Magnetic Resonance Imaging*, vol.22, no.6, pp.710–717, 2005
- [13] M. Manhart, A. Maier, J. Hornegger, A. Doerfler, Fast adaptive regularization for perfusion parameter computation, in *Bildverarbeitung für die Medizin 2015*, pp. 311–316, 2015
- [14] P.C.Hansen, The L-curve and its use in the numerical treatment of inverse problems, *Computational Inverse Problems in Electrocardiology*, vol. 4, pp. 119-142, 2001
- [15] Y. Sierra-Palomares, A. de Francisco, M. Felipe, D. Calle, L. Cussó, M. Desco. Optimización del modelo quirúrgico de isquemia reperfusión en rata y caracterización mediante imagen biomédica (RM y PET), *XIV Congreso Nacional de la Sociedad Española para las Ciencias del Animal de Laboratorio (SECAL)*, 2017
- [16] S.Ogawa, T.M. Lee, A.R. Kay, D.W. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 87, 1990
- [17] P.C. Van Zijl et al., Quantitative assessment of blood flow, blood volume and blood oxygenation effects in functional magnetic resonance imaging, *Nature medicine*, vol. 4, no. 2, pp. 159–167, 1998
- [18] Y. Ginosar et al., BOLD-MRI demonstrates acute placental and fetal organ hypoperfusion with fetal brain sparing during hypercapnia, *Placenta*, vol.63, pp.53–60, 2018
- [19] D. Bulte, P. Chiarelli, R. Wise, P. Jezard, Measurement of cerebral blood volume in humans using hyperoxic MRI contrast, *Journal of Magnetic Resonance Imaging*, vol.26, no.4, pp.894–899, 2007
- [20] M. E. MacDonald, A. J. Berman, E. L. Mazerolle, R. J. Williams, G. B. Pike, Modeling hyperoxia-induced BOLD signal dynamics to estimate cerebral blood flow, volume and mean transit time, *NeuroImage*, vol. 178, pp. 461–474, 2018
- [21] C. Vu et al., Quantitative perfusion mapping with induced transient hypoxia using bold MRI, *Magnetic Resonance in Medicine*, vol.85, no.1, pp.168–181, 2021

Análisis de OCTA (angiografía por tomografía de coherencia óptica) mediante redes neuronales convolucionales para el diagnóstico de esclerosis múltiple

R. Gutiérrez¹, R. Barea¹, M. Ortiz², E. Garcia-Martin³, A. López Dorado¹, C. Cavaliere¹, L. Boquete¹

¹ Grupo de Ingeniería Biomédica, Departamento de Electrónica, Universidad de Alcalá, España

² Interdisciplinary Center for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg.

³ Servicio de Oftalmología del Hospital Universitario Miguel Servet de Zaragoza. Universidad de Zaragoza. Instituto de Investigación Sanitaria de Aragón

Resumen

El objetivo de este trabajo es investigar la capacidad diagnóstica de la enfermedad de esclerosis múltiple (EM) utilizando datos de los espesores de las venas de los ojos, obtenidos mediante Angiografía por Tomografía de Coherencia Óptica, preprocesados utilizando segmentadores basados en redes neuronales y clasificados mediante una red neuronal convolucional (RNC). La angiografía por OCTA permite obtener información angiográfica utilizando múltiples algoritmos para la detección de flujo sanguíneo en la zona de la retina. Se dispone de imágenes de OCTA de 15 sujetos de control y 15 pacientes EM (recurrente-remitente) de diagnóstico reciente. Mediante el uso de técnicas de aprendizaje profundo se realiza una segmentación de las venas en la imagen OCTA. Se implementa una RNC formada por un módulo de extracción de características y un módulo de clasificación. Se utilizan como entradas a la RNC las imágenes segmentadas. El entrenamiento y test se realiza mediante validación cruzada dejando uno fuera. Utilizando las mismas en una RNC con un solo módulo de extracción de características (número de filtros: 64, dimensiones de cada filtro: 7x7, función de transformación: $f(x)=\max(0,x)$, reducción de datos en ventanas 2x2), se obtiene una precisión en la clasificación del 90%. Estos resultados sugieren que la utilización de técnicas de aprendizaje profundo para la clasificación de espesores de las venas de la retina, obtenidas por OCTA, puede contribuir al diagnóstico de EM.

1. Motivación

La esclerosis múltiple (EM) es una enfermedad crónica, neurodegenerativa y desmielinizante que afecta al sistema nervioso central y provoca en los pacientes un elevado número de síntomas, siendo frecuentes las afectaciones al sistema cognitivo, motor y visual [1]. En el año 2015 el número de afectados a nivel mundial fue de 2,5 millones de personas y en un elevado porcentaje de casos se empieza a desarrollar la enfermedad entre los 20-40 años; tres de cada cuatro pacientes son mujeres.

El diagnóstico de EM se realiza mediante los criterios de McDonald [2], que principalmente se basan en el análisis de imágenes de resonancia magnética (MRI) y en la situación clínica del paciente. Sin embargo, su efectividad en las fases iniciales de la enfermedad no es plena y, en consecuencia, es conveniente investigar en otros posibles biomarcadores.

La Tomografía de Coherencia Óptica (OCT) permite obtener in vivo, de forma no invasiva y con gran precisión espesores de las diferentes capas de la retina. Mediante un sistema de interferometría óptica de baja coherencia se puede medir la distancia a partir de la luz reflejada por diferentes tejidos y con ello, se obtiene la morfología tisular de capas con diferentes índices de reflectividad. Numerosos estudios evalúan las ventajas de la técnica OCT como biomarcador en el diagnóstico de EM [3][4][5][6] y su capacidad diagnóstica sigue siendo objeto de investigación.

Por otro lado, la angiografía por OCT (OCT-A), es una nueva técnica no invasiva emergente que permite la visualización en alta resolución de imágenes angiográficas volumétricas de la retina que se está utilizando ampliamente para estudiar la vascularización de la retina [7][8]. Numerosos estudios describen la evaluación de patologías como la degeneración macular relacionada con la edad (DMRE), las oclusiones vasculares retinianas (OVR), retinopatía diabética (RD), glaucoma y otras alteraciones vasculares [1-8, finales].

Las técnicas de aprendizaje profundo se utilizan ampliamente para el análisis de información en medicina, como pueden ser señales unidimensionales, por ejemplo electroencefalogramas [9]; sin embargo son más numerosas las aplicaciones dedicadas al análisis de imágenes [10]; en el campo de los sistemas OCT, se utilizan para la segmentación de las diferentes capas de la retina [11].

La estructura de las redes neuronales convolucionales (RNC) se pueden dividir en dos módulos conectados en cascada: módulo de extracción de características y módulo de clasificación [12]. El primero obtiene las características de la imagen mediante convoluciones con máscaras de diferentes coeficientes y dimensiones, que destacan los bordes, gradientes de intensidad, etc. de la imagen original. El módulo de clasificación tiene como entradas características obtenidas en el módulo previo y puede ser implementado con una red neuronal con conexiones hacia delante (perceptrón multicapa, red de base radial, etc); su entrenamiento puede realizarse con algoritmos basados en descenso por el gradiente, entre otras opciones.

1.1. Objetivo

La OCTA permite visualizar en alta resolución las tres dimensiones de la vasculatura retiniana, incluyendo pequeños vasos. De este modo se pueden obtener imágenes de alta resolución de 4 capas vasculares: a) Red vascular superficial (capa de células ganglionares y de fibras nerviosas). b) Red vascular profunda (capilares en medio de las capas plexiforme interna y externa). c) Retina externa (fotorreceptores, no tiene vasos, pero hay índices de perfusión). d) Coriocapilaris. Aunque normalmente se trabaja con las capas vascular superficial (Superficial vascular complexes –SVC-) y profunda (Deep vascular complexes –DVC-).

Por lo tanto, el objetivo de este trabajo es evaluar un sistema de ayuda al diagnóstico de EM mediante la segmentación y el análisis de imágenes de OCTA con redes neuronales convolucionales. Así como realizar una comparación entre los resultados obtenidos utilizando la imagen en color sin procesamiento y los resultados obtenidos utilizando la imagen segmentada. En ambos casos, estos resultados se obtienen utilizando la misma red neuronal convolucional.

2. Métodos

El estudio ha sido aprobado por el Comité Ético del Hospital Universitario Miguel Servet (HUMS), se han seguido los protocolos recogidos en la declaración de Helsinki y los participantes ratificaron su consentimiento por escrito.

Se dispone de una base de datos consistente en 15 registros OCTA de ambos ojos de pacientes con diagnóstico precoz de EM (recurrente-remitente) sin neuritis óptica previa ($46,64 \pm 12,51$ años) y 15 registros de control ($47,91 \pm 15,12$ años). Para cada ojo se dispone de imágenes correspondientes al espesor de los vasos sanguíneos.

El protocolo de exploración obtiene una imagen de 320×320 puntos de medida para cada ojo. Se selecciona aleatoriamente un ojo de cada sujeto para incluirlo en el análisis.

Para la clasificación de las imágenes se realiza un preprocesamiento dividido en varios pasos (Fig. 1).

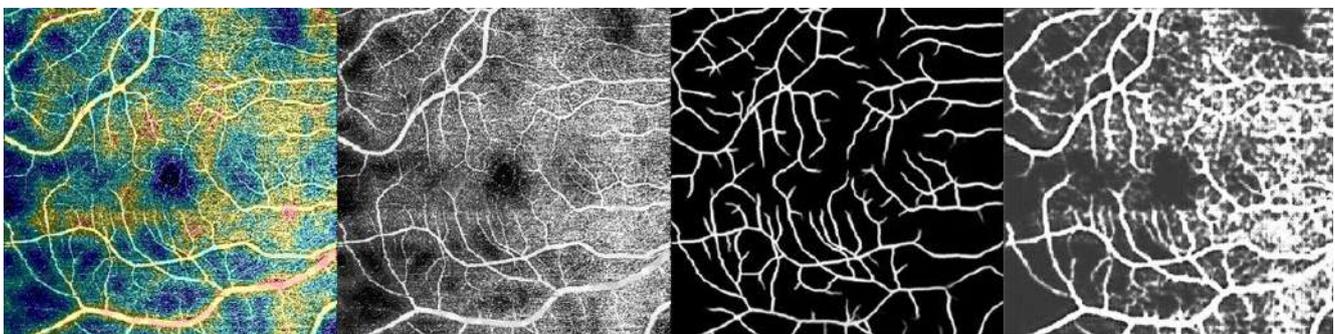
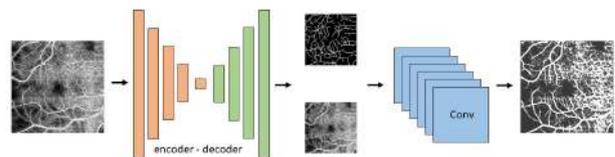


Figura 1. Ilustración que contiene las distintas imágenes obtenidas en cada proceso. De izquierda a derecha: imagen original OCTA, imagen centrada y de un solo canal, imagen segmentada (groundtruth) e imagen segmentada como entrada al clasificador.

- Reducción y centrado de la imagen: se toma como el punto central la mácula, generando la imagen a clasificar. Además, las imágenes obtenidas son de tres canales, en este primer procesamiento se convierten en imágenes de un solo canal.
- Generación de una imagen “groundtruth”: se utiliza el segmentador BOE2020-OCTA-vessel-segmentation [11][12]. Se realiza una delineación del Plexo Vascular Superficial en una proyección 2D de máxima intensidad utilizando redes neuronales convolucionales que de una forma iterativa refinan la calidad de la segmentación de las venas.
- Extracción de características: Para ello se modifica la arquitectura propuesta por iMED-Lab [12]. Como entradas a la red se utilizan tanto la imagen a clasificar como la imagen “groundtruth”. La salida de esta red es una imagen segmentada de tal forma que la clasificación futura sea más sencilla.

2.1. Arquitectura de modelo de segmentación.

Se divide la segmentación en dos capas. La primera capa se ha diseñado para producir dos mapas preliminares segmentando tanto a nivel de píxel como a nivel de la línea central de las venas. Para ello, se propone un encoder-decoder basado en la ResNet50. Una vez que se tiene esta segmentación, las imágenes son concatenadas y pasan por varias capas convolucionales. Se realiza una fusión obteniendo la segmentación refinada final.



En la fig. 2 se muestra un esquema de la arquitectura de la segmentación, además de las imágenes que se van obteniendo en cada una de las fases de esta segmentación.

2.2. Arquitectura de la RNC

Se muestra en la Fig. 3. la estructura de la RNC

implementada mediante Toolbox Deep Learning (Matlab). La entrada son las imágenes segmentadas. El módulo de extracción de características está formado por una unidad de convolución con N_F filtros de dimensiones $d \times d$. Los resultados de la convolución son transformados mediante la función $f(x) = \max(0, x)$ para mejorar el proceso de entrenamiento. El último paso dentro de este módulo consiste en la reducción de las dimensiones (reducción por muestreo), también para disminuir el número de pesos ajustables en la red neuronal y, en consecuencia, beneficiar el entrenamiento.



Figura 3. Arquitectura de la red neuronal convolucional implementada.

El módulo de clasificación está formado por una red totalmente conectada (FCL), constituida por la capa de entrada y la capa de salida (2 neuronal), estando todas las neuronas de entrada conectadas a las dos neuronas de salida con un peso y un bias, ambos ajustables. Las salidas de FCL se normalizan mediante una función exponencial normalizada (softmax):

$$p(i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

que asigna un valor de probabilidad a la salida de la FCL. La capa de clasificación (CL) asigna el valor de salida de la clasificación (control o EM).

Los resultados del clasificador se obtienen mediante validación cruzada dejando uno fuera [12]: de los 30 datos de sujetos disponibles, se utilizan 29 en el entrenamiento y los datos del sujeto restante se utilizan en la prueba. El proceso se repite 30 veces, de tal forma, que todos los sujetos son testeados. Los resultados se muestran mediante una matriz de confusión.

3. Resultados

La base de datos está formada por imágenes de OCTA provenientes de 15 pacientes (hombres/mujeres: 9/39) con EM (43.79 ± 8.41 años) y 15 (hombres/mujeres: 10:38) sujetos de control (44.44 ± 7.18 años). No existe diferencia significativa en las medias de edad entre ambos grupos ($p=0.107$) ni en la distribución entre sexos ($p=0.451$). Los pacientes son de reciente diagnóstico (7.35 ± 1.95 meses).

Los parámetros que definen la arquitectura de la RNC que permite obtener los mejores resultados en la clasificación son: número de filtros $N_F=64$, dimensiones de los filtros= 7×7 . Por lo tanto, se obtienen un total de 64 imágenes que representan los rasgos o características de las imágenes OCT. En la capa de reducción se definen ventanas de dimensiones 2×2 , sin solapamiento entre ventanas y se selecciona el píxel de mayor valor, destacando de este modo las características más abruptas de los mapas de características.

El ajuste del clasificador se realiza mediante descenso de gradiente estocástico [12].

En la tabla 1 se muestra la matriz de confusión con los resultados de la clasificación de las imágenes sin procesar de los 30 sujetos participantes mediante la RNC.

	Casos Reales			Precisión
		Controles	EM	
Predicción RNC	Controles	10	7	60%
	EM	5	8	

Tabla 1. Matriz de confusión

En la tabla 2 se muestra la matriz de confusión con los resultados de la clasificación de las imágenes tras el procesamiento de los 30 sujetos participantes mediante la RNC.

	Casos Reales			Precisión
		Controles	EM	
Predicción RNC	Controles	13	2	90%
	EM	1	12	

Tabla 2. Matriz de confusión

4. Discusión

En este trabajo se ha propuesto un sistema de ayuda al diagnóstico de EM mediante el análisis de imágenes de OCTA mediante redes neuronales convolucionales, obteniendo unos resultados prometedores en la clasificación (precisión=90%).

Los resultados obtenidos en el presente trabajo suponen una mejora en relación a un trabajo previo, en el que se utilizaron redes neuronales modulares [13] y la misma base de datos y permiten abrir nuevas posibilidades en el diagnóstico temprano de pacientes con EM.

Una de las ventajas de las redes neuronales convolucionales es que pueden considerarse como una caja negra, en la que se establecen correspondencias entre sus entradas y salidas, no siendo necesario en muchos casos una extracción explícita de características. En el presente trabajo se han utilizado las imágenes completas obtenidas mediante OCTA. Sin embargo, como líneas futuras de investigación se considera la posibilidad de seleccionar las zonas con mayor capacidad discriminante de la imagen, antes de ser utilizadas como entradas a la RNC. Además, sería conveniente ampliar la base de datos, de tal modo que se incluyan datos de sujetos obtenidos en diferentes centros y también utilizando equipos OCTA con diferentes tecnologías. También sería deseable complementar la información obtenida con OCTA con otros tipos de imágenes médicas, que han demostrado eficacia en el diagnóstico de EM.

Si los trabajos futuros confirman unos resultados satisfactorios, sería posible implementar un sistema de cribado para la detección de EM de fácil uso por parte del personal médico, no invasivo y de coste razonable.

Agradecimientos

Al Ministerio de Ciencia, Innovación y Universidades mediante el proyecto: DPI2017-88438-R, “Investigación de la técnica de potenciales evocados visuales multifocales. Aplicación en estudios de evolución de esclerosis múltiple y evaluación de medicamentos”.

Al Instituto de Salud Carlos III, por su financiación al proyecto PI17/01726 RETICS: “Enfermedades oculares: prevención, detección precoz y tratamiento de la patología ocular prevalente degenerativa y crónica: OFTARED”. (EGM: RD16/0008/029) (LB: RD16/0008/0020).

A la Dirección General de Investigación e Innovación Tecnológica, Comunidad de Madrid, por su apoyo a través del Programa de Empleo de Garantía Juvenil PEJ-2019-AI/SAL-13374.

Referencias

- [1] C. A. Dendrou, L. Fugger, and M. A. Friese, “Immunopathology of multiple sclerosis,” *Nat. Rev. Immunol.*, vol. 15, no. 9, pp. 545–558, Sep. 2015.
- [2] A. J. Thompson *et al.*, “Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria,” *Lancet Neurol.*, vol. 17, no. 2, pp. 162–173, Feb. 2018.
- [3] A. Petzold *et al.*, “Retinal layer segmentation in multiple sclerosis: a systematic review and meta-analysis,” *Lancet. Neurol.*, vol. 16, no. 10, pp. 797–812, 2017.
- [4] J. Britze, G. Pihl-Jensen, and J. L. Frederiksen, “Retinal ganglion cell analysis in multiple sclerosis and optic neuritis: a systematic review and meta-analysis,” *J. Neurol.*, vol. 264, no. 9, pp. 1837–1853, Sep. 2017.
- [5] R. Alonso, D. Gonzalez-Moron, and O. Garcea, “Optical coherence tomography as a biomarker of neurodegeneration in multiple sclerosis: A review,” *Mult. Scler. Relat. Disord.*, vol. 22, pp. 77–82, May 2018.
- [6] C. Cavaliere *et al.*, “Computer-Aided Diagnosis of Multiple Sclerosis Using a Support Vector Machine and Optical Coherence Tomography Features,” *Sensors*, vol. 19, no. 23, p. 5323, Dec. 2019.
- [7] R. A. Leitgeb, “En face optical coherence tomography: a technology review,” *Biomed. Opt. Express*, vol. 10, no. 5, pp. 2177–2201, 2019.
- [8] Y. Jia, O. Tan, J. Tokayer, B. Potsaid, Y. Wang, J. J. Liuet *al.*, “Split-spectrum amplitude-decorrelation angiography with optical coherence tomography,” *Opt. Express*, vol. 20, no. 4, pp. 4710–4725, 2012.
- [9] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, “Deep learning-based electroencephalography analysis: a systematic review,” *J. Neural Eng.*, vol. 16, no. 5, p. 051001, Aug. 2019.
- [10] J. Ker, L. Wang, J. Rao, and T. Lim, “Deep Learning Applications in Medical Image Analysis,” *IEEE Access*, vol. 6, pp. 9375–9389, 2018.
- [11] M. Pekala, N. Joshi, T. Y. A. Liu, N. M. Bressler, D. C. DeBuc, and P. Burlina, “Deep learning based retinal OCT segmentation,” *Comput. Biol. Med.*, vol. 114, p. 103445, Nov. 2019.
- [12] Y. Ma, H. Hao, J. Xie, H. Fu, J. Zhang, J. Yang, Z. Wang, J. Liu, Y. Zheng, Y. Zhao. ROSE: A Retinal OCT-Angiography Vessel Segmentation Dataset and New Model. *IEEE Trans Med Imaging*. 2021 Mar;40(3):928-939. doi: 10.1109/TMI.2020.3042802. Epub 2021 Mar 2. PMID: 33284751.
- [13] A. Shrestha and A. Mahmood, “Review of Deep Learning Algorithms and Architectures,” *IEEE Access*, vol. 7, pp. 53040–53065, 2019.
- [14] Theodoros Pissas and Edward Bloch, “Deep iterative vessel segmentation in OCT angiography,” *Biomed. Opt. Express*, pp 2490–2510, 2020.
- [15] Ma, Yuhui and Hao, Huaying and Xie, Jianyang and Fu, Huazhu and Zhang, Jiong and Yang, Jianlong and Wang, Zhen and Liu, Jiang and Zheng, Yalin and Zhao, Yitian. “ROSE: a retinal OCT-angiography vessel segmentation dataset and new model”, *IEEE Transactions on Medical Imaging*, pp 928–939. 2021
- [16] L. Bottou, “Large-Scale Machine Learning with Stochastic Gradient Descent,” in *Proceedings of COMPSTAT’2010*, Heidelberg: Physica-Verlag HD, 2010, pp. 177–186.
- [17] R. Barea *et al.*, “Detección de pacientes con esclerosis múltiple mediante procesamiento de imágenes de tomografía de coherencia óptica utilizando redes neuronales,” in *Actas del 27 Seminario Anual de Automática, Electrónica Industrial e Instrumentación (SAAEI 2020)*, Ciudad Real, (ISBN: 978-84-122260-2-7), 2020, pp. 432–436.

Desarrollo y validación de los bucles de control para las interfaces hápticas de un simulador de silla de ruedas manual

A. Pizarro-Chong¹, F. Chénier²

¹ Department of Systems Engineering, École de Technologie Supérieure, Montreal, Canadá, ary.pizarro@mail.mcgill.ca

² Department of Physical Activity Science, Université du Québec à Montréal, Montreal, Canadá, chenier.felix@uqam.ca

Resumen

Alrededor del 50% de los usuarios de silla de ruedas manual (SRM) empiezan a padecer lesiones en los miembros superiores en los cinco primeros años de uso. La investigación es necesaria para entender mejor esta patología. Los simuladores de SRM son una modalidad de investigación para la cual el realismo es de suma importancia en cuanto a su alta validez ecológica. Se buscó mejorar el realismo del simulador de SRM de nuestro laboratorio. Se reemplazó el transductor de fuerza actual (ruedas instrumentadas) por una célula de fuerza que pudiese tomar en cuenta las fuerzas inerciales debidas al movimiento del tronco. Se utilizaron parámetros biomecánicos pertinentes para evaluar el nivel de realismo del simulador que incorpora la célula de fuerza. Se comparó su nivel al mismo proporcionado por ruedas instrumentadas y también se comparó a la propulsión en el suelo. Se construyó un bucle de control de admitancia estable, que incorporaba esta célula de fuerza, para controlar la interacción momento-velocidad de las ruedas de la SRM. En pruebas con un sólo sujeto sano, el nivel de realismo del simulador con célula de fuerza fue un tanto mayor al nivel proporcionado por las Smartwheels. La célula de fuerza puede captar los efectos de la dinámica debida al movimiento del tronco, algo que no pueden hacer las ruedas instrumentadas. Esto podría explicar el ligero incremento en el realismo.

1. Introducción

Hay más o menos 10 millones de usuarios de sillas de ruedas manuales (SRM) en los países desarrollados [1]. Para las personas que padecen trastornos severos de los miembros inferiores, las SRM son una ayuda de movilidad necesaria que incrementa su autonomía y participación en la sociedad (actividades diarias, trabajo, educación, vida social) [2]. Sin embargo, las SRM requieren la realización de tareas que pueden ser perjudiciales para la salud de sus usuarios. Dos problemas que han salido a la luz con respecto al uso de las SRM son: 1) lesiones de sobreutilización y 2) falta de formación. Con respecto a las lesiones, ciertos estudios han informado que el 50 a 75% de los usuarios de SRM empiezan a padecer de dolores de los miembros superiores y trastornos músculo-esqueléticos debidos a la propulsión en SRM en los cinco primeros años de uso [3]. Con respecto a la formación, estudios informan que el nivel de formación en habilidades y técnicas de SRM es del 11 al 55% en países tales como Canadá, EE.UU. y el Reino Unido [4]. Nuestro laboratorio de investigación, el LIO (Laboratoire de recherche en imagerie et orthopédie), ubicado dentro del Centro de Investigación del Centro Hospitalario de la Universidad de Montreal (CRCHUM), mantiene un programa de investigación dedicado a los

problemas que han surgido debido al uso de las SRM. El simulador de propulsión en SRM del LIO fue construido como parte de este programa (Figura 1). Tiene características avanzadas: ejecuta un bucle de control háptico, proporciona feedback háptico y ejecuta un modelo bidimensional de una SRM en el suelo. Dos ruedas instrumentadas (Smartwheels), que reemplazan las ruedas originales de la SRM, se usan como sensores de fuerza. Esto crea dos inconvenientes: en primer lugar, las Smartwheels no miden la dinámica debida al movimiento del tronco. En segundo lugar, al reemplazar las ruedas originales de la SRM, no permiten el uso completo de la SRM del usuario. Su rigidez y geometría podrían ser diferentes de las de las ruedas originales. Estas deficiencias podrían afectar el realismo del simulador.



Figura 1. El simulador de propulsión en SRM del LIO

La meta de nuestro trabajo fue mejorar la validez ecológica del Simulador del LIO y se buscó incrementar su nivel de realismo para realizarlo. Nuestro objetivo fue reemplazar las Smartwheels por un transductor de fuerza que pudiese resolver los inconvenientes mencionados previamente. Se propusieron dos hipótesis de investigación: H1) Era posible desarrollar un bucle de control estable utilizando una célula de fuerza para controlar el Simulador del LIO; H2) Este bucle de control podría proporcionar un nivel de realismo superior a ese proporcionado por las Smartwheels.

2. Método

2.1. Modelo matemático

Investigar el realismo relacionado con la reproducción de la dinámica de una SRM en el suelo implica el análisis del modelo matemático que se va a implementar en el simulador. Si estudiamos el movimiento del cuerpo del usuario durante la propulsión en SRM, percatamos que su tronco experimenta un movimiento hacia atrás y hacia adelante, de una posición recta a una inclinada con respecto al suelo. Este movimiento tiene un efecto cuantificable

sobre la dinámica del sistema SRM-usuario. Par tomar en cuenta el movimiento del tronco, este se reemplaza en el modelo por la fuerza que el tronco aplica en el marco de la SRM. Esta fuerza, que se denomina F_{tronco} , es igual a la masa del tronco multiplicado por la aceleración del tronco relativa al marco de la SRM.

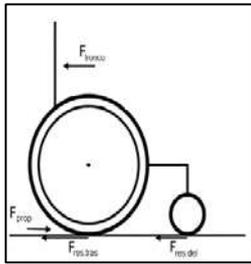


Figura 2. Fuerzas aplicadas en una silla de ruedas manual

Se aplica la Segunda Ley del Movimiento de Newton a nuestro sistema, y se obtiene la ecuación mostrada más abajo que es sólo en la dirección horizontal (Figura 2):

$$F_{prop} - F_{rod,tras} - F_{rod,del} - m_{tronco} a_{tronco/srm,marco} = m_{srm,marco} a_{srm,marco}$$

Al lado izquierdo, las fuerzas son la fuerza de propulsión, la resistencia de rodadura en las ruedas traseras y en las ruedas delanteras y la fuerza del tronco. Al lado derecho está la masa del marco de la SRM multiplicada por la aceleración del mismo.

2.2. Bucle de control

En base a los previos trabajos de investigación [5-7], se decidió implementar el bucle de control de admitancia que se muestra en el diagrama conceptual de la Figura 3. En la esquina superior izquierda, se observa que tanto las fuerzas propulsivas como esas inerciales debidas al movimiento del tronco se miden mediante la célula de fuerza. Se envía el resultado de esta medición al sistema de control y al modelo matemático ejecutado por una computadora en tiempo real.

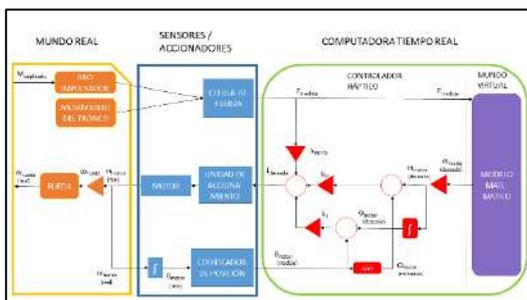


Figura 3. Diagrama conceptual del bucle de control de admitancia del simulador de SRM del LIO

2.3. Hipótesis H1

La primera hipótesis plantea la cuestión si se puede desarrollar un bucle de control estable que use el transductor de fuerza elegido. El bucle de control propuesto fue un bucle Proporcional-Integral con control anticipativo. Se obtuvo un bucle de control funcional ajustando tres ganancias de control k_P , k_I et k_{FF} , por ensayo y error. Los valores iniciales de las tres ganancias fueron

esas obtenidas por Chénier [8] durante el desarrollo del Simulator del LIO, para el cual se usó un modelo bidimensional de la propulsión en SRM. El ajuste de las ganancias se hizo realizando un experimento en que se propulsó la SRM del Simulator del LIO desde cero manteniendo ambas ruedas a la velocidad de 1 m/s, durante 20 segundos. Se trazaron las curvas de la velocidad deseada (la salida del modelo matemático) y la velocidad real (estimada mediante la medición hecha por un codificador óptico). Se buscó minimizar la diferencia entre las velocidades deseadas y reales mediante una inspección visual.

2.4. Hipótesis H2

Para cuantificar el nivel de realismo del simulador impulsado por la célula de fuerza, Se compararon la propulsión en SRM entre tres condiciones. La primera condición fue la propulsión sobre el suelo (se denominó condición SS); la segunda condición fue la propulsión en el simulador con las Smartwheels (condición SW); la tercera condición fue la propulsión en el simulador con la célula de fuerza (condición CF). Los parámetros biomecánicos usados para la comparación surgieron naturalmente del patrón de propulsión utilizado por los individuos. El ciclo de propulsión empieza cuando el usuario coge el aro impulsador en el punto CM (contacto de la mano). Aplica una fuerza F_{TOT} durante cierto tiempo, que se llama fase de empuje. Esta fuerza genera un momento M_Z alrededor del eje de la rueda. El usuario luego suelta la rueda en el punto LM (liberación de la mano). La fase entre los puntos CM y LM se llama fase de recuperación y varía según los individuos. Se focalizó en los cuatro siguientes parámetros biomecánicos: la fuerza máxima F_{TOTMAX} aplicada durante la fase de empuje, medida en Newtons; el momento máximo M_{ZMAX} aplicado durante la fase de empuje, medido en Newton-metros; la duración de la fase de empuje P_T , medida en segundos y la duración de la fase de recuperación R_T , medida en segundos.

3. Resultados

3.1. Hipótesis H1

Se obtuvieron los valores de $k_P = 2$, $k_I = 6$ y $k_{FF} = 0.25$ como valores óptimos de las ganancias de control para la célula de fuerza. En la Figura 4, se muestra un gráfico de la velocidad angular deseada vs. real, para ambas ruedas, cuando el bucle de control fue impulsado por la célula de fuerza. Se percata que la velocidad medida de la rueda, trazada en azul, sigue la velocidad deseada trazada en rojo, con una oscilación de alta frecuencia.

3.2. Hipótesis H2

Se presenta en la Figura 5 la evaluación de H2, hecha mediante la comparación de los parámetros biomecánicos, entre las condiciones mencionadas previamente.

En la esquina superior derecha, con respecto a la fuerza máxima total aplicada F_{TOTMAX} , se observa que las diferencias entre las condiciones SS y CF son grandes (-19, -20) y que las diferencias entre las condiciones SS y SW son igual de grandes (-23, -27).

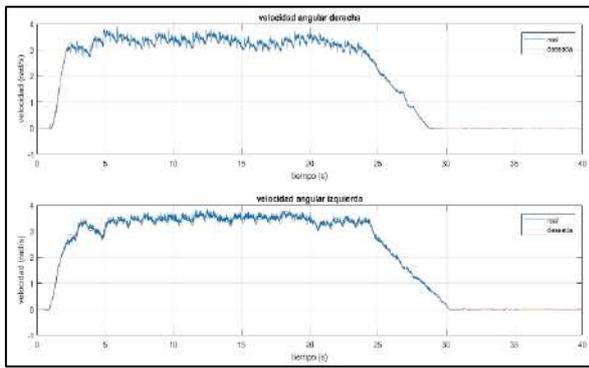


Figura 4. Célula de fuerza: velocidades angulares medida vs. deseada, ruedas izquierda y derecha

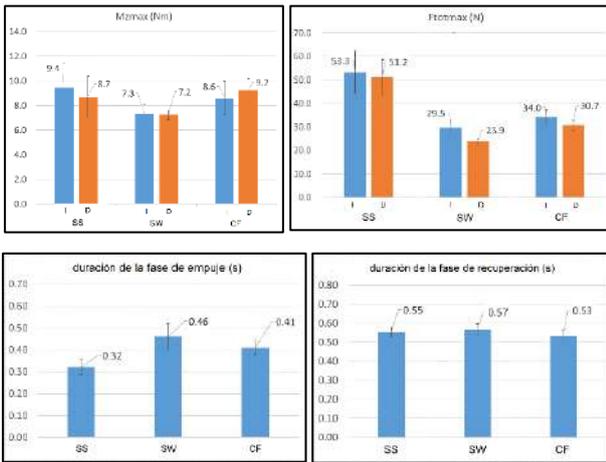


Figura 5. Promedio sobre 10 ciclos de propulsión para 1 sujeto sin discapacidad

Con respecto al tiempo de empuje, la diferencia entre SS y CF es grande (0.09), y la diferencia entre SS y SW es igual de grande (0.14). Con respecto al tiempo de recuperación, la diferencia entre SS y CF es pequeña (0.02), y la diferencia entre SS y SW es igual de pequeña (0.02). Los datos para M_{ZMAX} están la esquina superior izquierda. Se percata que las diferencias entre SS y CF (-0.8, 0.5) son menores que las diferencias entre SS y SW (-2.1, -1.4). Esto significa que, con respecto a esos parámetros, la condición CF está más cercana de la propulsión real (condición) que la condición SW.

4. Discusión

4.1. Bucle de control

Se confirmó la Hipótesis H1, que establece que se puede desarrollar un bucle de control estable que utiliza una célula de fuerza para controlar la interacción momento-velocidad del Simulador del LIO. Sin embargo, se realizó el ajuste de las ganancias de control ejecutando experimentos sólo a la velocidad de 1 m/s. Por lo tanto, no se sabe si, en condiciones diferentes, por ejemplo, a la velocidad de 0.5, 0.8, 1.5 o 2 m/s, los bucles de control mantienen su estabilidad; esto queda por averiguar. Así mismo, se evaluaron los diferentes gráficos mediante una inspección visual, pero, en el futuro, se podrá cuantificar la diferencia entre las velocidades medida y deseada mediante un cálculo del error rms entre las curvas. Además, se obtuvieron las ganancias de control óptimas

para la velocidad autoseleccionada del usuario pasando por un rango de valores de los parámetros k_P , k_I , k_{FF} . En el futuro, se podrán determinar usando algoritmos de optimización ya incluidos en MATLAB, o mediante algoritmos ya existentes en la literatura científica. Otra observación que se hizo es que el resultado del control con la Smartwheel era mejor que con la célula de fuerza: la amplitud de oscilación de la velocidad real alrededor de la velocidad deseada era mucha menor. La diferencia se debe al control por anticipación. Se trata de controlar la velocidad angular del motor, y esta velocidad está sometida a una perturbación que es el momento aplicado en los aros impulsadores. Cuando se usan las Smartwheels como transductores de fuerza, se mide directamente esta perturbación, lo que da un mejor resultado. Cuando se usa la célula de fuerza como transductor de fuerza, sólo se tiene un estimado de esta perturbación, obtenido multiplicando las fuerzas equivalentes en la izquierda y la derecha por el radio de los aros impulsadores, y por lo tanto esto produce un resultado de control menos preciso.

En cuanto a la Hipótesis H2, con respecto a las tres variables fuerza total máxima, tiempo de empuje y tiempo de recuperación, no se pudo llegar a una conclusión en base a los resultados que mostrados aquí porque las dos condiciones CF y SW dieron resultados similares. Sin embargo, cuando se analizaron los resultados del parámetro M_{ZMAX} , se pudo concluir que el realismo proporcionado por el bucle de control impulsado por la célula de fuerza es ligeramente más realista que ese proporcionado por las Smartwheels. Por lo tanto, a nuestro juicio, se confirma la hipótesis H2. No obstante, tendríamos que realizar un análisis estadístico basado en datos colectados sobre por lo menos 15 sujetos sanos para establecer con certidumbre que la diferencia observada con el único sujeto sano es estadísticamente significativa. Más aún, se puede realizar una validación adicional del bucle de control, incluyendo a 15 sujetos discapacitados. Como los usuarios principiantes impulsan diferentemente de los usuarios experimentados, sería beneficioso repetir la investigación realizada aquí. Luego, se realizaría un análisis estadístico de tipo anova para investigar si las diferencias podrían ser significativas.

4.2. Célula de fuerza

Se realizaron ciertas observaciones interesantes como que resultaron de los experimentos que se ejecutaron. Se comparó la señal de entrada al modelo matemático obtenido de la célula de fuerza con eso obtenido de las Smartwheels. La célula de fuerza mide la fuerza anteroposterior aplicada por el usuario en el resto de la estructura del simulador. Las Smartwheels miden los momentos aplicados en los aros impulsadores. La fuerza tangencial propulsa la SRM hacia adelante; se calculó esta fuerza dividiendo el momento aplicado por el radio del aro impulsador. Se realizó la comparación trazando en la Figura 6, para el mismo experimento, la fuerza medida por la célula de fuerza y la fuerza tangencial aplicada mediante las Smartwheels.

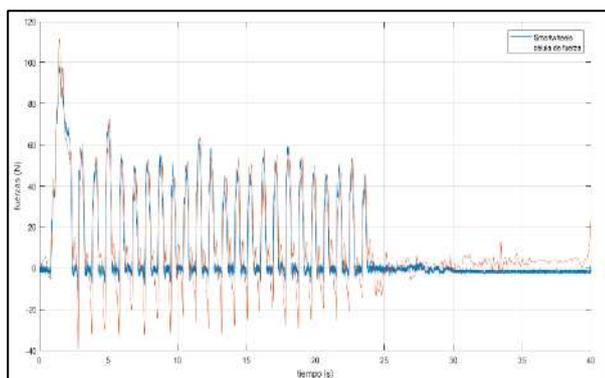


Figura 6. Fuerza medida por la célula de fuerza (rojo) y por las Smartwheels (azul)

Se percató que hay una diferencia leve entre las dos curvas. Durante la fase de empuje, se observó que la fuerza medida por la célula de fuerza era ligeramente mayor a esa medida por las Smartwheels. La diferencia se debe posiblemente al movimiento del tronco. Justo al principio de la fase de recuperación, la fuerza de la célula de fuerza sube un poco, luego baja, antes de subir otra vez cuando el usuario impulsa el aro impulsador. Esta diferencia en la fuerza se debe probablemente al cambio en momento lineal debido al movimiento del tronco del usuario.

4.3. Límites del trabajo de investigación

El estudio presentado aquí fue el primer intento de hacer funcionar el simulador de SRM del LIO con un control impulsado por una célula de fuerza. Se hicieron varias suposiciones, que limitan la validez de los resultados. Se pueden enumerar las siguientes limitaciones:

- Se supuso que la resistencia de rodadura era constante, y no variaba con el movimiento del usuario y de su tronco
- Con respecto al análisis de parámetros biomecánicos, si se hace funcionar el simulador del LIO sólo con la célula de fuerza y no se reemplazan las ruedas originales por Smartwheels, no se mide directamente el momento aplicado en el aro impulsador. Este sólo se puede estimar mediante la medición hecha por la célula de fuerza.

5. Conclusión

Dos de los temas biomédicos que han salido a la luz con respecto al uso de las SRM son las lesiones de sobreutilización de los miembros superiores y la falta de formación. Un simulador de SRM ofrece varias posibilidades interesantes para abordar ambos temas. La validez ecológica de los experimentos de investigación y de la formación realizada en un simulador de SRM es un aspecto importante de los simuladores de SRM. Estudios de investigación indican que un alto nivel de realismo mejora la validez ecológica de los simuladores.

El Simulador del LIO ubicado en el CRCHUM tiene características avanzadas, pero no cumple con todos los requisitos de un simulador de alto realismo debido a su uso de las ruedas instrumentadas. Para abordar este tema, se instaló una célula de fuerza que podía tomar en cuenta las fuerzas de propulsión y las fuerzas inerciales debidas a la

aceleración del tronco. Se construyó un bucle de control de admitancia estable y comparamos cuatro parámetros biomecánicos entre tres condiciones. Así mismo, se pudo concluir que el realismo proporcionado por el bucle de control impulsado por la célula de fuerza era ligeramente más realista que el realismo proporcionado por las Smartwheels.

Agradecimientos

Se agradece al profesor Hironao Yamada (Universidad de Gifu, Gifu, Japón) y la organización Mitacs por el apoyo que nos brindaron mediante la beca de investigación JSPS-MITACS 2018. También se agradece al Natural Sciences and Engineering Research Council of Canada (NSERC) por su apoyo financiero.

Referencias

- [1] New Disability. (2002). *Wheelchair Statistics: How Many Wheelchair Users Are There?* <http://www.newdisability.com/wheelchairstatistics.htm> (Consultada: Mayo 2020)
- [2] Rushton, P. W., Miller, W. C., Mortenson, W. B., & Garden, J. (2010). Satisfaction with participation using a manual wheelchair among individuals with spinal cord injury. *Spinal Cord*, 48(9), 691–696. <https://doi.org/10.1038/sc.2009.197>
- [3] Paralyzed Veterans of America, C. for S. C. M. (2005). Preservation of Upper Limb Function Following Spinal Cord Injury: A Clinical Practice Guideline for Health-Care Professionals. *The Journal of Spinal Cord Medicine*, 28(5), 434–470. <https://doi.org/10.1080/10790268.2005.11753844>
- [4] Kirby, R. L., Miller, W. C., Routhier, F., Demers, L., Mihailidis, A., Polgar, J. M., Rushton, P. W., Titus, L., Smith, C., McAllister, M., Theriault, C., Thompson, K., & Sawatzky, B. (2015). Effectiveness of a Wheelchair Skills Training Program for Powered Wheelchair Users: A Randomized Controlled Trial. *Archives of Physical Medicine and Rehabilitation*, 96(11), 2017–2026.e3. <https://doi.org/10.1016/j.apmr.2015.07.009>
- [5] Bentaleb, T., Nguyen, V. T., Sentouh, C., Conreur, G., Poulain, T., & Pudlo, P. (2019). A Real-Time Multi-Objective Predictive Control Strategy for Wheelchair Ergometer Platform. *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2397–2404. <https://doi.org/10.1109/SMC.2019.8914220>
- [6] de Klerk, R., Vegter, R. J. K., Veeger, H. E. J., & Van der Woude, L. H. V. (2020). Technical note: A novel servo-driven dual-roller handrim wheelchair ergometer. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(4), 953–960. <https://doi.org/10.1109/TNSRE.2020.2965281>
- [7] Klaesner, J., Morgan, K. A., & Gray, D. B. (2014). The Development of an Instrumented Wheelchair Propulsion Testing and Training Device. *Assistive Technology*, 26(1), 24–32. <https://doi.org/10.1080/10400435.2013.792020>
- [8] Chénier, F. (2012). Développement D'un Simulateur De Propulsion En Fauteuil Roulant Manuel Avec Biofeedback Haptique [Ph.D. thesis]. École de Technologie Supérieure.

Modificación de un modelo de miocito ventricular humano para representar el papel de los canales SK en insuficiencia cardíaca

M.Gómez¹, J. Carro¹, E. Pueyo^{3,4}, V. Monasterio¹

¹ Universidad San Jorge, Villanueva de Gállego, España {magomez,jcarro,vmonasterio}@usj.es

² Aragón Institute for Engineering Research, University of Zaragoza, IIS Aragón, Spain, epueyo@unizar.es

³ CIBER in Bioengineering, Biomaterials & Nanomedicine (CIBER-BBN), Spain, epueyo@unizar.es

Resumen

Los canales SK son canales de potasio activados por calcio de conductancia individual alrededor de 10 pS. En los últimos años han cobrado importancia, ya que, aunque no tienen un papel determinante en la electrofisiología ventricular en corazones sanos, se ha comprobado que en condiciones de insuficiencia cardíaca aparecen sobreexpresados, lo que puede tener efectos relevantes en la repolarización ventricular.

En este trabajo se propone la ampliación de un modelo electrofisiológico de miocito ventricular humano mediante la introducción de las ecuaciones que representan la actividad de los canales SK, con el fin de obtener una herramienta que nos permita estudiar el papel de estos canales en la repolarización ventricular en condiciones de insuficiencia cardíaca. Para ello, se parte de un modelo ventricular capaz de representar las condiciones de insuficiencia cardíaca. Dicho modelo se amplía añadiendo una formulación para la corriente generada por estos canales I_{SK} . Para el ajuste de la conductancia máxima, G_{SK} , se realizan simulaciones que se comparan con datos experimentales de duración del potencial de acción obtenidos de la literatura. Los resultados obtenidos con el nuevo modelo son coherentes con la prolongación del potencial de acción producida por el bloqueo farmacológico de los canales SK que se observa experimentalmente en miocitos humanos con insuficiencia cardíaca.

1. Introducción

El desarrollo de modelos matemáticos que describan la actividad eléctrica del corazón y ayuden a comprender mejor la electrofisiología cardíaca es importante ya que, hoy en día, las enfermedades cardiovasculares son una de las principales causas de mortalidad en el mundo [1]. La insuficiencia cardíaca (IC) es una afección crónica y progresiva en la que el músculo cardíaco no puede bombear suficiente sangre para satisfacer las necesidades corporales de sangre y oxígeno. Puede ocurrir porque el corazón se ha vuelto demasiado débil o rígido. Además, suele aparecer asociada a otras patologías metabólicas y/o cardíacas como la fibrilación auricular (FA), que es la arritmia más común en la práctica clínica, caracterizada por una activación rápida y descoordinada de las aurículas. Avances recientes en la búsqueda de terapias contra la FA apuntan a que la modulación de los canales SK, canales de potasio activados por calcio de conductancia individual pequeña (del orden de 10 pS), podría tener efectos beneficiosos en estos pacientes. Es por eso por lo que, en los últimos años, los canales de

potasio activados por calcio han cobrado importancia en el campo de la electrofisiología cardíaca. Estos canales regulan el potencial de membrana y el tono vascular en los vasos sanguíneos. De acuerdo con su conductancia (pequeña, media o alta), son clasificados de tres formas diferentes. En concreto, los canales SK (conductancia pequeña) carecen del sensor de voltaje común de otros canales iónicos y, por lo tanto, están activados puramente por calcio. Cuando el calcio intracelular aumenta durante la descarga repetida del potencial de acción (AP), la activación de los canales SK genera una hiperpolarización de larga duración, denominada “post-hiperpolarización lenta”, que reduce la excitabilidad. Por lo tanto, los canales SK proporcionan un vínculo crítico entre el calcio intracelular y el potencial de membrana [2,3]. Se ha comprobado que los canales SK contribuyen notablemente a la fase de repolarización tardía del AP auricular, pero hasta hace poco se consideraba que no desempeñaban un papel relevante en la repolarización ventricular. Sin embargo, estudios recientes en miocitos y tejidos ventriculares han demostrado que los canales SK responden de forma diferente en ciertos estados patológicos, destacando estos canales como un posible mecanismo de respaldo en circunstancias patológicas [4].

En concreto, se ha observado experimentalmente que el bloqueo de los canales SK produce una prolongación de la duración del potencial de acción (APD) ventricular en condiciones de IC, pero no tiene efecto en condiciones fisiológicas [4]. Estos resultados sugieren, que la sobreexpresión de los canales SK en condiciones patológicas podría ser una respuesta fisiológica adaptativa para acortar el APD bajo una reserva de repolarización reducida [5]. Sin embargo, las condiciones bajo las cuales la acción de estos canales puede considerarse proarrítmica o antiarrítmica aún no se comprenden bien.

El objetivo de este trabajo es proponer un modelo matemático de electrofisiología celular para investigar los posibles efectos arrítmicos ventriculares de los canales SK en condiciones de IC. Se ha partido de un modelo ventricular capaz de representar condiciones de IC y se ha ampliado añadiendo una formulación para la corriente generada por los canales I_{SK} basada en datos disponibles en la literatura. Los resultados del modelo se han comparado con datos experimentales para comprobar su validez.

2. Métodos

2.1. Modelo ventricular humano

Se utilizó como punto de partida el modelo de miocito ventricular humano O'Hara-Rudy (ORd) [6]. Es el modelo ventricular humano más ampliamente utilizado y validado en la literatura. Se modificó para representar las condiciones de IC según [7]. También se sustituyó la formulación original de I_{Na} por la formulación de tenTuscher et al. [8] para representar mejor la pendiente máxima del inicio del AP, de manera similar al método utilizado en [9].

Para incluir la formulación de la corriente I_{SK} , se utilizaron las ecuaciones propuestas en [10], es decir:

$$\begin{cases} I_{SK} = G_{SK} x_{SK} (V - E_K) \\ \frac{dx_{SK}}{dt} = \frac{x_{SK,\infty} - x_{SK}}{\tau_{SK}} \\ x_{SK,\infty} = \frac{1}{1 + (K_d / cass)^n} \\ \tau_{SK} = \tau_0 + \frac{\tau_1}{1 + (cass / 0.1)} \end{cases}$$

donde G_{SK} representa la conductancia de la corriente, x_{SK} la compuerta de activación por calcio, τ_{SK} la constante de tiempo de dicha compuerta, $x_{SK,\infty}$ el valor de la compuerta en estado estacionario y $cass$ la concentración de calcio en la zona en la que se encuentran los canales SK.

Se tomaron como referencia datos experimentales publicados en la literatura [10,11] para obtener el valor de los distintos parámetros de los canales SK: $\tau_0 = 4$ ms, $\tau_1 = 20$ ms, $n = 3.14$, $K_d = 0.000345$ mM.

En la literatura no se encontraron valores experimentales para el valor de la conductancia G_{SK} en miocitos ventriculares humanos. Por lo tanto, se decidió realizar un ajuste para estimar un valor óptimo en función de datos experimentales.

2.2. Valores experimentales de referencia

Para llevar a cabo el ajuste de la conductancia, se consideraron los datos experimentales publicados en [4], donde se estudiaron células endocárdicas de 7 pacientes con IC y se midieron los marcadores APD₅₀ (duración al 50% de la repolarización) y APD₉₀ (duración al 90% de la repolarización) en condiciones con y sin bloqueo de la corriente I_{SK} , para tres frecuencias de estimulación distintas: 0.5, 1 y 2 Hz (ver Figura 1). El bloqueo se llevó a cabo con 100 nM de apamina.

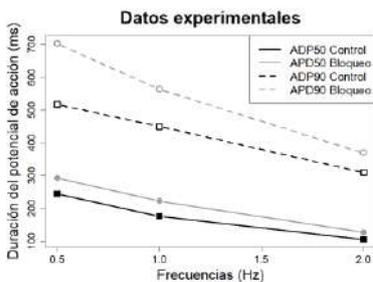


Figura 1. Valores de APD₅₀ y APD₉₀ en condiciones de bloqueo de los canales SK frente a condiciones de control. Figura replicada de [4].

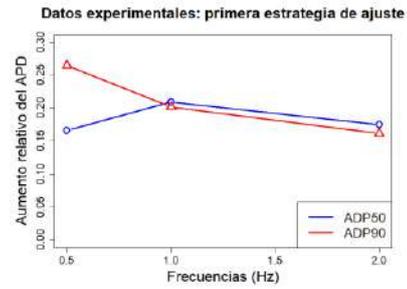


Figura 2. Aumento relativo en APD₅₀ y APD₉₀ sin I_{SK} frente a condiciones con I_{SK} .

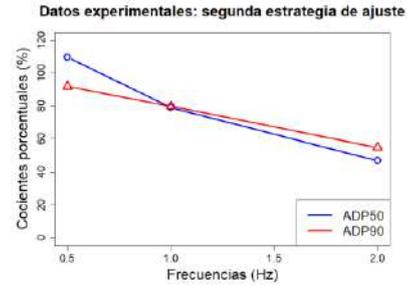


Figura 3. Cocientes porcentuales del APD₅₀ y el APD₉₀ sin I_{SK} frente a condiciones con I_{SK} .

A partir de los datos experimentales se estudió cómo variaba el APD en presencia de los canales SK. Se siguieron dos estrategias diferentes para estudiar esta variación, en función de qué valor se tomaba como referencia.

Para la primera estrategia, se calculó el aumento relativo (A) producido por el bloqueo de los canales, en APD₅₀ y APD₉₀ para cada una de las frecuencias de estimulación, según la siguiente ecuación:

$$A = \frac{APD_{\sin I_{SK}} - APD_{\text{con } I_{SK}}}{APD_{\sin I_{SK}}}$$

Los datos de referencia de esta primera estrategia se presentan en la Figura 2.

Como referencia para la segunda estrategia, se tomaron los cocientes porcentuales (C) de los datos en control, respecto de los valores de referencia de APD₅₀ y APD₉₀ a 1 Hz en condiciones de bloqueo de los canales SK. Se tomó esa frecuencia de referencia al considerarse situación de control (a nivel de frecuencia). Es decir:

$$C = \frac{APD_{i-\text{con } I_{SK}}(j \text{ Hz})}{APD_{i-\text{sin } I_{SK}}(1 \text{ Hz})} \cdot 100 \quad i = 50, 90 \text{ y } j = 0.5, 1, 2$$

Los datos de referencia de esta segunda estrategia se muestran en la Figura 3.

2.3. Ajuste de la conductancia G_{SK}

Para hallar el valor de G_{SK} , se realizó una minimización por superficies de respuesta de la siguiente función de error que representa el error cuadrático medio de la

diferencia entre los valores experimentales y los resultados del modelo [13]:

$$f(G_{SK}) = \sum_{i=1}^M (err_{m_i}(G_{SK}))^2 = \sum_{i=1}^M (\mu_i - m_i(G_{SK}))^2$$

donde μ_i representa cada uno de los $M = 6$ valores obtenidos experimentalmente (los valores de A o C representados en las Figuras 2 y 3 según la estrategia de ajuste) y $m_i(G_{SK})$ representa los valores simulados con el modelo computacional para cada valor de conductancia G_{SK} .

Siguiendo la metodología de [13], para no tener que simular el modelo y calcular todos los parámetros en cada iteración de la búsqueda del mínimo, se realizaron 20 simulaciones para diferentes valores de G_{SK} y con esos resultados se aproximaron los valores de A y C mediante polinomios de grado 2. Con estos polinomios y la función de error indicada anteriormente, se realizó la búsqueda del mínimo mediante el método de *Brent* [14], tomando como punto inicial para el algoritmo el valor de G_{SK} asociado con menor error de entre los 20 valores simulados. Una vez obtenido el mínimo, se comprobó que las simulaciones del modelo con ese valor de G_{SK} devolvían los valores esperados.

Las simulaciones se llevaron a cabo utilizando DENIS [15], un simulador de electrofisiología cardiaca basado en el estándar CellML, que utiliza una red de computación voluntaria. Se utilizó *Forward Euler* (FE) para la integración con un paso temporal de 0.01 ms. Se utilizó una estimulación monofásica, para la cual se estableció un valor de $-80 \mu\text{A}/\mu\text{F}$, correspondiente a dos veces el umbral diastólico. Tras una exploración inicial considerando valores de conductancia entre 0 y 0.5, según referencias de la literatura [10], se estableció un rango de búsqueda para G_{SK} óptimo de 0 a $0.006 \mu\text{S}$. Para cada valor de G_{SK} , el modelo se estimuló con frecuencias de 0.5 Hz, 1 Hz y 2 Hz, durante 20 ciclos cada vez bajo condiciones de control y con bloqueo por apamina. El bloqueo experimental con apamina se simuló como un bloqueo total de la corriente I_{SK} , es decir, $G_{SK} = 0$, ya que la concentración utilizada (100 nM) es un orden de magnitud mayor que la IC_{50} reportada en la literatura [16].

3. Resultados

3.1. Resultados de la primera estrategia de ajuste

Como resultado de la primera estrategia de ajuste, se obtuvo un error mínimo para un valor de conductancia de $G_{SK} = 0.004356 \mu\text{S}$. La Figura 4 muestra gráficamente la comparativa de los valores obtenidos.

3.2. Resultados de la segunda estrategia de ajuste

Como resultado de la segunda estrategia de ajuste, se obtuvo un error mínimo para un valor de conductancia de $G_{SK} = 0.00507 \mu\text{S}$. Los valores obtenidos se muestran en la Figura 5.

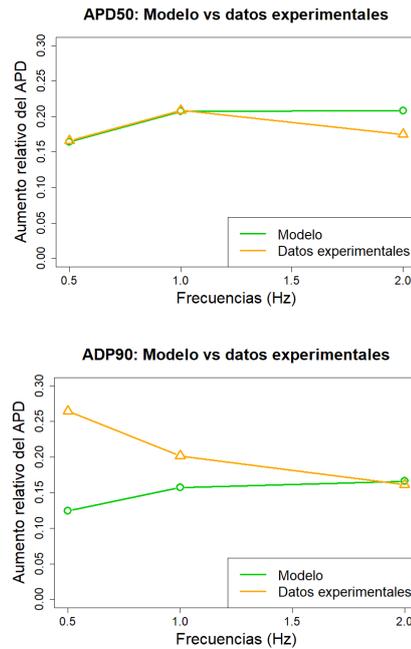


Figura 4. Comparativa de los resultados entre el modelo y los datos experimentales del APD₅₀ y APD₉₀.

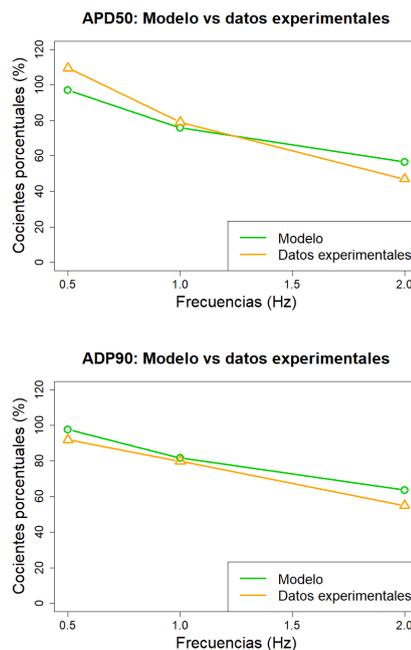


Figura 5. Comparativa de los resultados entre el modelo y los datos experimentales del APD₅₀ y APD₉₀.

4. Discusión y conclusiones

En este trabajo se propone un modelo matemático celular para investigar el efecto de los canales SK en ventrículo humano en condiciones de insuficiencia cardiaca. Para ello, se amplió un modelo electrofisiológico existente añadiendo las ecuaciones para la corriente I_{SK} . Para esta corriente se eligió una formulación sin rectificación, ya que formulaciones alternativas con rectificación [12] requerirían estimar un mayor número de parámetros y,

dada la falta de datos experimentales, podría introducirse un grado de incertidumbre elevado en el modelo. En la formulación elegida, los valores para las constantes de tiempo y para la constante de disociación se extrajeron de la literatura. No fue posible encontrar un valor para la conductancia G_{SK} , por lo que se realizó un ajuste mediante dos estrategias diferentes.

En primer lugar, se intentó replicar la prolongación de APD_{50} y APD_{90} producida por el bloqueo de los canales SK para cada frecuencia de estimulación. Los resultados del modelo se acercaron más a los experimentales en el caso de APD_{50} que en el caso de APD_{90} (Figura 4). La variación con la frecuencia de APD_{90} simulado no siguió la tendencia descendente de los resultados experimentales. Cabe destacar que dicha tendencia descendente puede no ser significativa, ya que en el estudio original no se comparan estadísticamente las diferencias en la prolongación del APD en función de la frecuencia.

A la vista de estos resultados, se decidió adoptar una segunda estrategia tomando como referencia la variación de los marcadores con respecto a la situación de bloqueo con estimulación a 1 Hz.

Los resultados obtenidos con la segunda estrategia basada en cocientes porcentuales muestran que los marcadores simulados tienen un comportamiento similar al de los marcadores experimentales (Figura 5).

Se encontró que las dos estrategias para la determinación de G_{SK} produjeron valores muy similares, aportando fiabilidad al valor propuesto. Por lo tanto, se considera que el ajuste puede ser aceptable para estudios posteriores, y que un valor fisiológicamente razonable podría estar bastante cercano al que se ha encontrado en este estudio. No obstante, el modelo resultante deberá ser validado en simulaciones 1D y 2D para comprobar si permite reproducir propiedades observadas experimentalmente en tejido.

Agradecimientos

Este trabajo está financiado por los proyectos ERC-StG 638284 (ERC) y PID2019-105674RB-I00 (Ministerio de Ciencia e Innovación) y por el Fondo Social Europeo (UE) y el Gobierno de Aragón a través del grupo BSICoS T39_20R. Los cálculos se realizaron utilizando la ICTS NANBIOSIS (HPC Unit, Univ. Zaragoza).

Referencias

- [1] Gómez, Luis Alberto "Las enfermedades cardiovasculares: un problema de salud pública y un reto global". *Biomédica*, vol. 31, no.4, 2011.
- [2] Skibbye L, Poulet C, Diness JG, Bentzen BH, Yuan L, Kappert U, Matschke K, Wettwer E, Ravens U, Grunnet M, Christ T, Jespersen T. "Small-conductance calcium-activated potassium (SK) channels contribute to action potential repolarization in human atria". *Cardiovasc Res.*, vol.103, no.1, pp. 156-167. 2014.
- [3] C. Celotto, C. Sánchez, P. Laguna and E. Pueyo, "Calcium-Activated Potassium Channels Inhibition in Autonomically Stimulated Human Atrial Myocytes," *Computing in Cardiology (CinC)*, Singapur, 2019, pp. 1- 4.
- [4] Bonilla, I. M., Long, V. P., 3rd, Vargas-Pinto, P., Wright, P., Belevych, A., Lou, Q., Mowrey, K., Yoo, J., Binkley, P. F., Fedorov, V. V., Györke, S., Janssen, P. M., Kilic, A., Mohler, P. J., & Carnes, C. A. "Calcium-activated potassium current modulates ventricular repolarization in chronic heart failure". *PLoS one*, vol.9, no.10, 2014.
- [5] Gu, M., Zhu, Y., Yin, X., & Zhang, D. M. "Small-conductance Ca^{2+} -activated K^{+} channels: insights into their roles in cardiovascular disease". *Experimental & molecular medicine*, vol.50, no.4, pp. 1-7, 2018.
- [6] O'Hara, T., Virág, L., Varró, A., & Rudy, Y. "Simulation of the undiseased human cardiac ventricular action potential: model formulation and experimental validation". *PLoS computational biology*, vol.7, no.5, 2011.
- [7] Gomez, J. F., Cardona, K., Romero, L., Ferrero, J. M., Jr, & Trenor, B. "Electrophysiological and structural remodeling in heart failure modulate arrhythmogenesis. 1D simulation study". *PLoS one*, vol.9, no.9, 2014.
- [8] ten Tusscher, K. H., Noble, D., Noble, P. J., & Panfilov, A. V. "A model for human ventricular tissue". *American journal of physiology. Heart and circulatory physiology*, vol.286, no.4, pp.73-89, 2004.
- [9] Mora, M. T., Ferrero, J. M., Romero, L., & Trenor, B. "Sensitivity analysis revealing the effect of modulating ionic mechanisms on calcium dynamics in simulated human heart failure". *PLoS one*, vol.12, no.11, 2017.
- [10] Landaw, J., Zhang, Z., Song, Z., Liu, M. B., Olcese, R., Chen, P. S., Weiss, J. N., & Qu, Z. "Small-conductance Ca^{2+} -activated K^{+} channels promote J-wave syndrome and phase 2 reentry". *Heart rhythm*, vol.17, no.9, pp. 82-90, 2020.
- [11] Chang, P. C., Turker, I., Lopshire, J. C., Masroor, S., Nguyen, B. L., Tao, W., Rubart, M., Chen, P. S., Chen, Z., & Ai, T. "Heterogeneous upregulation of apamin-sensitive potassium currents in failing human ventricles". *Journal of the American Heart Association*, vol.2, no.1, 2013.
- [12] Bronk, P., Kim, T. Y., Polina, I., Hamilton, S., Terentyeva, R., Roder, K., Koren, G., Terentyev, D., & Choi, B. R. "Impact of Isk Voltage and Ca^{2+}/Mg^{2+} -Dependent Rectification on Cardiac Repolarization". *Biophysical journal*, vol.119, no.3, pp. 690-704, 2020.
- [13] Carro, J., Pueyo, E., & Rodríguez Matas, J. F. "A response surface optimization approach to adjust ionic current conductances of cardiac electrophysiological models. Application to the study of potassium level changes". *PLoS one*, vol.13, no.10, 2018.
- [14] Brent, R. "Algorithms for Minimization without Derivatives". Englewood Cliffs N.J.: Prentice-Hall, 2002.
- [15] Monasterio, V., Castro-Mur, J., & Carro, J. "DENIS: Solving cardiac electrophysiological simulations with volunteer computing". *PLoS one*, vol 13, no.10, 2018.
- [16] Benton, D. C. H., Garbarg, M., & Moss, G. W. J. "The Relationship between Functional Inhibition and Binding for KCa_2 Channel Blockers". *PLoS one*, vol.8, no.9, 2013.

Análisis del impacto combinado del grupo sanguíneo y los antígenos leucocitarios humanos en la supervivencia del trasplante renal en Andalucía

A. Talaminos Barroso¹, J. Reina Tosina^{1,2}, L. M. Roa Romero¹, J. Calvillo Arbizu^{1,3}, M. A. Pérez Valdivia⁴, R. Medina López⁴, P. Castro de la Nuez⁵

¹ Grupo de Ingeniería Biomédica, Universidad de Sevilla, Sevilla, España {atalaminos, jreina, lroa, jcalvillo}@us.es

² Departamento de Teoría de la Señal y Comunicaciones, Universidad de Sevilla, Sevilla, España

³ Departamento de Ingeniería Telemática, Universidad de Sevilla, Sevilla, España

⁴ Unidad de Urología y Nefrología, Hospital Universitario Virgen del Rocío, Sevilla, España, {miguel.perez, rantonio.medina}.sspa@juntadeandalucia.es

⁵ Coordinación Autonómica de Trasplante de Andalucía, Sevilla, España, pablo.castro.sspa@juntadeandalucia.es

Resumen

El trasplante es el tratamiento de elección para pacientes que padecen enfermedad renal crónica, una de las causas de muerte con mayor prevalencia en todo el mundo. El éxito de un trasplante a corto y largo plazo depende de multitud de factores, entre los que se encuentran principalmente la edad y el tipo de muerte del donante o las incompatibilidades de antígenos leucocitarios humanos (HLA) entre donante y receptor. En este trabajo se presenta un estudio de la influencia del grupo sanguíneo y las incompatibilidades de HLA en la supervivencia renal de trasplantes en Andalucía utilizando el método de Kaplan-Meier conjuntamente con diferentes métricas de significación estadística. Los resultados ponen de manifiesto que el grupo sanguíneo no tiene prácticamente impacto en la supervivencia del injerto renal, mientras que la influencia de los HLA es mayor. Por otra parte, cuando ambos factores son analizados conjuntamente, sus efectos sobre la viabilidad del trasplante a corto y largo plazo son significativos.

1. Introducción

El trasplante es el tratamiento más adecuado para la enfermedad renal en etapa terminal [1]. La baja disponibilidad de órganos produce largas listas de espera, lo que ha acentuado el turismo de trasplantes en algunos países. En contraste con esta tendencia, las donaciones de órganos en España son voluntarias, gratuitas y anónimas. Actualmente existe la necesidad de desarrollar políticas adaptadas a las normas culturales de las diferentes sociedades y principios éticos universalmente aceptados para afrontar este problema de salud pública. En este sentido, España es líder mundial en número de donaciones de órganos por millón de habitantes [2].

En relación con el trasplante renal y, al igual que sucede con otros tipos de órganos, la cuidadosa selección del donante y su compatibilidad con el receptor, es un proceso fundamental para garantizar el éxito de la supervivencia del injerto a corto y largo plazo [3]. En general, está prácticamente aceptado que la edad del donante es uno de los factores individuales con mayor impacto en la supervivencia del injerto renal [4]. Otros factores que tienen influencia en menor medida son el tipo de muerte

del donante, la edad del receptor, o la presencia de diabetes, hepatitis C o citomegalovirus en el donante. Otras características asociadas a la compatibilidad entre donante y receptor también son evaluadas, entre las que se incluye el grupo sanguíneo y las incompatibilidades de los antígenos leucocitarios humanos (HLA) [5].

Con respecto al grupo sanguíneo, en la gran mayoría de trasplantes renales que se realizan existe coincidencia entre los grupos del donante y receptor. En este sentido, trabajos recientes han puesto de manifiesto que el grupo sanguíneo no tiene influencia en la supervivencia del injerto a corto o largo plazo [6,7].

En lo que respecta a los HLA, existen dos moléculas: las de clase I y II. Los antígenos HLA-A y HLA-B, pertenecientes a la clase I, y el antígeno HLA-DR, de la clase II, son generalmente considerados para el estudio de la supervivencia renal. En particular, un mayor grado de incompatibilidad de HLA se asocia con una mayor frecuencia de fracasos [8].

En este trabajo se presenta, en primer lugar, un análisis de los efectos del grupo sanguíneo y de las incompatibilidades HLA, considerados independientemente, sobre la probabilidad de supervivencia renal en datos de trasplantes de la población andaluza. A continuación, se define un índice que tiene en cuenta el grado de incompatibilidad acumulada de las distintas moléculas HLA, y se realiza el estudio de este factor combinado con el grupo sanguíneo. Los estudios de esta naturaleza existentes en la literatura se han caracterizado por el análisis independiente de estos dos factores y, hasta el conocimiento de los autores, no existe ningún estudio que los combine, ni tampoco que incluya la suma combinada de los efectos de las incompatibilidades HLA.

2. Materiales y métodos

2.1. Descripción de los datos

Los datos de trasplantes renales que han sido analizados en este trabajo provienen del Sistema de Información de la Coordinación Autonómica de Trasplantes de Andalucía

(SICATA), en un periodo de tiempo comprendido entre el 1 de enero de 2006 y el 31 de diciembre de 2019. Los datos se refieren a trasplantes de un único riñón de donantes con edades comprendidas entre 3 y 90 años, y receptores mayores de 18 años.

En la Tabla 1 se presenta un resumen de las características más relevantes de los datos utilizados. El número de casos es 5078, con un total de 558 fracasos, lo que supone un 11% aproximadamente del total de datos considerados. El tiempo medio de seguimiento para todos los casos de trasplantes ha sido de 5.8 años.

Característica	n	%	Media (std)
Edad del donante			
<20 años	171	3.37	14.46 (4.38)
[20-40) años	587	11.57	30.38 (5.94)
[40-60) años	2319	45.66	51.06 (5.41)
>60 años	2001	39.4	67.64 (5.52)
Grupo sanguíneo del donante			
A	2318	45.65	-
B	495	9.75	-
AB	181	3.56	-
0	2084	41.04	-
Edad del receptor			
<20 años	12	0.24	38.75 (13.61)
[20-40) años	808	15.91	38.97 (13.79)
[40-60) años	2443	48.1	51.7 (12.84)
>60 años	1815	35.75	63.8 (10.73)
Grupo sanguíneo del receptor			
A	2333	45.95	-
B	491	9.67	-
AB	222	4.36	-
0	2032	40.02	-

Tabla 1. Características principales de los datos: donante y receptor

En la Tabla 2 se resume el número de casos para cada nivel de incompatibilidad HLA (HLA-A, HLA-B y HLA-DR). Los valores de estas incompatibilidades están cuantificados como 0, 1 y 2, donde 0 representa alto grado de compatibilidad entre donante y receptor, y 2 indica bajo grado. En la tabla se observa que existe un número significativo de casos para los que no existen registros de esta variable, por lo que estos trasplantes han sido descartados en el estudio.

	0	1	2	Perdidos
HLA-A	568	1761	1151	1598
HLA-B	247	1410	1823	1598
HLA-DR	647	1965	866	1600

Tabla 2. Número de casos considerando el tipo de incompatibilidad HLA y valores perdidos. La columna izquierda representa el tipo y la fila superior el número de incompatibilidades

2.2. Métodos

Para la realización de este estudio se ha seguido la metodología general para este tipo de trabajos:

1. Búsqueda bibliográfica de estudios de análisis de supervivencia que consideren el grupo sanguíneo y las incompatibilidades HLA como factores que tienen impacto en la supervivencia del trasplante renal.
2. Proceso de limpieza y mejora de calidad de datos, incluyendo normalización de las variables de tipo fecha, aplicación de un criterio único para identificar valores nulos, eliminación de variables intrascendentes, cambio de nombre de algunas variables con objeto de describir mejor su contenido, normalización de valores mediante la técnica del escalado y conversión de variables categóricas mediante la técnica del One Hot Encoding.
3. Análisis de supervivencia del injerto renal considerando cada una de las características más relevantes del donante, receptor e injerto mediante Kaplan-Meier [9], incluyendo un análisis de las diferencias mediante métricas de significación estadística como el logrank y sus distintas variantes de ponderación, entre las que se ha utilizado Fleming-Harrington [10]. Otras técnicas más avanzadas como las redes neuronales han sido descartadas dado que el entrenamiento de modelos con este tipo de algoritmos requiere de una cantidad de datos mayor a la manejada en este trabajo.
4. Investigación, diseño y desarrollo de un algoritmo de exploración automatizada que permita la comparación de datos tomados dos a dos del donante, receptor e injerto y las diferencias de comportamiento temporales tempranas, medias y tardías mediante el método Fleming-Harrington.
5. Representación gráfica de las probabilidades de supervivencia obtenidas y discusión de los resultados.

3. Resultados

3.1. Análisis del grupo sanguíneo

En la Tabla 3 se presentan el número de casos de trasplantes considerando el grupo sanguíneo del donante y del receptor. La incompatibilidad de los grupos sanguíneos es poco frecuente, como puede observarse, por lo que estos casos han sido descartados en el análisis de datos. En consecuencia, se han considerado cuatro grupos de trasplantes: A-A, B-B, AB-AB y 0-0, si bien el número de trasplantes AB-AB son poco numerosos en comparación con el resto. El número total de casos para los cuatro grupos es 4977.

Donante/ Receptor	A	B	AB	0
A	2287	0	26	5
B	0	483	11	1
AB	0	0	181	0
0	46	8	4	2026

Tabla 3. Número de trasplantes considerando el grupo sanguíneo del donante (filas) y el grupo sanguíneo del receptor (columnas)

En la Figura 1 se presenta la probabilidad de supervivencia del injerto renal para cada uno de los grupos sanguíneos considerados. A partir de las curvas de Kaplan-Meier obtenidas, se ha comparado cada grupo poblacional con respecto al resto de grupos (dos a dos). El nivel de significación estadística p -valor para todas las posibles combinaciones fue siempre mayor que 0.08, lo que pone de manifiesto que, en general, no hay diferencias de impacto entre grupos sanguíneos sobre la probabilidad de supervivencia del injerto renal. Este resultado está en concordancia con trabajos de otros autores [6,7]. No obstante, puede observarse en la Figura 1 que, a partir del sexto año, existe cierta variación de comportamientos entre los grupos 0-0, con respecto a los grupos A-A y grupos B-B. Esta variabilidad puede ser identificada mediante el test de Fleming-Harrington para $p = 1$ y $q = 1$, obteniendo un p -valor de 0.05 a partir de la mayor asignación de peso a instantes intermedios en el tiempo.

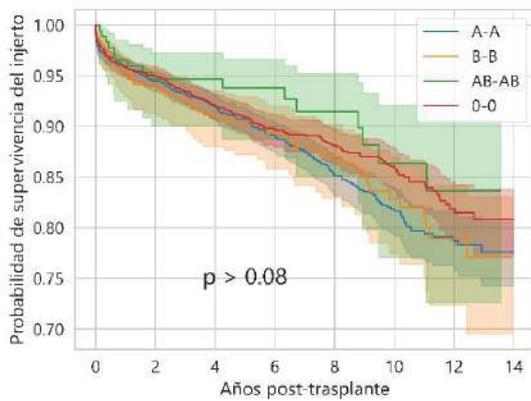


Figura 1. Probabilidad de supervivencia por grupo sanguíneo del donante y del receptor

3.2. Análisis de las incompatibilidades HLA

En la Figura 2 se presenta la probabilidad de supervivencia renal para los tres valores posibles de HLA-DR, observando un solapamiento de comportamientos en las tres probabilidades prácticamente todo el tiempo del estudio (p -valor > 0.36). Al contrario de lo esperado, el valor de incompatibilidad 0, asociado a una mejor compatibilidad entre donante y receptor, presenta la menor probabilidad de supervivencia en algunos periodos de tiempo.

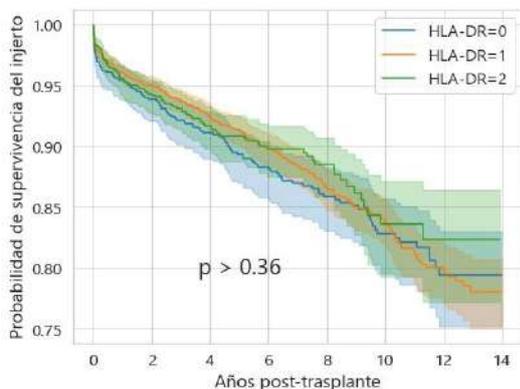


Figura 2. Probabilidad de supervivencia para diferentes valores de incompatibilidad HLA-DR

A pesar de que la incompatibilidad HLA-DR es la que parece tener mayor influencia en la probabilidad de supervivencia renal [11], en nuestros datos la incompatibilidad HLA-B es la que posee mayor impacto (p -valor > 0.05). En la Figura 3 se presenta la probabilidad de supervivencia para los tres valores de HLA-B, observándose que la probabilidad es mayor cuando HLA-B=0 durante todo el tiempo del estudio, lo que está en concordancia con otros estudios de estas características [12]. Con respecto a las incompatibilidades HLA-B=1 y HLA-B=2, las diferencias obtenidas son poco significativas.

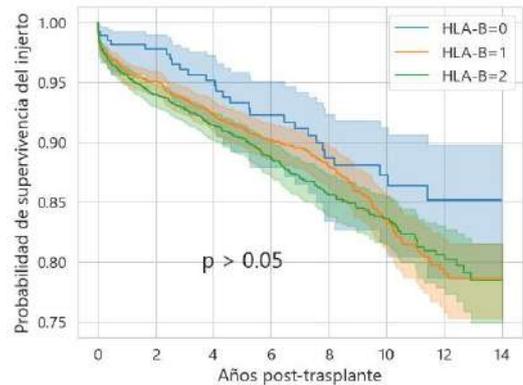


Figura 3. Probabilidad de supervivencia para diferentes valores de incompatibilidad HLA-B

3.3. Análisis combinado del grupo sanguíneo y las incompatibilidades HLA

Para amplificar el efecto de las incompatibilidades HLA se ha inferido un nuevo factor a partir de la suma de las tres incompatibilidades disponibles: HLA-A, HLA-B, HLA-DR. Los valores de este nuevo factor oscilan entre 0 y 6, donde 0 supone un mayor grado de compatibilidad entre donante y receptor, y 6 el menor. En la Figura 4 se muestra la distribución de valores para todos los casos considerados en el análisis.

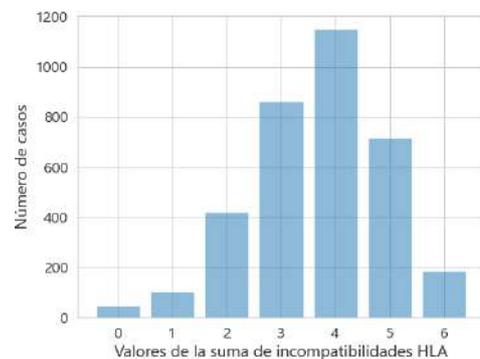


Figura 4. Distribución de casos para cada uno de los valores de la suma de incompatibilidades HLA

En la Figura 5 se presentan los resultados del análisis de supervivencia renal combinando el grupo sanguíneo y la suma de incompatibilidades HLA. En particular, las diferencias entre grupos sanguíneos no son significativas cuando la suma de incompatibilidades HLA toma un valor mayor de tres (p -valor > 0.72), pero son relevantes cuando toma un valor de tres o inferior. En este sentido, se han considerado todas las posibles agrupaciones con uno, dos

y tres valores, no hallando diferencias significativas entre grupos, excepto para el caso anterior comentado.

Cuando la suma de incompatibilidades es menor de tres, las diferencias de comportamiento se evidencian en todos los grupos sanguíneos y desde prácticamente la realización del trasplante (p -valor <0.05 para A-A y 0-0, y B-B y 0-0). El grupo B-B presenta la menor probabilidad de supervivencia en los primeros años después del trasplante, aunque posteriormente lo es el grupo A-A. El grupo 0-0 se mantiene con la mayor probabilidad de supervivencia desde prácticamente el inicio, mientras que con respecto al grupo AB los resultados no son concluyentes debido a la escasez de casos.

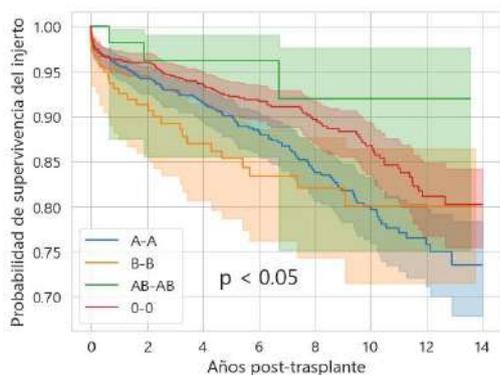


Figura 5. Probabilidad de supervivencia por grupos sanguíneos considerando una suma de incompatibilidades de tres o menor

4. Conclusiones

En este trabajo se ha presentado el análisis de la influencia del grupo sanguíneo y de las incompatibilidades HLA sobre la supervivencia del injerto renal, tanto de forma aislada como conjunta. Se ha definido también un índice que tiene en cuenta la influencia acumulada de las incompatibilidades HLA. Los resultados muestran que los casos de trasplantes para el grupo sanguíneo B-B poseen una probabilidad de supervivencia más baja en nuestros datos durante los primeros años de estudio. A partir del sexto año, los datos en relación a este grupo sanguíneo son escasos, por lo que la probabilidad se mantiene prácticamente constante. Los casos 0-0 y AB-AB presentan una probabilidad más alta durante todo el periodo del estudio, si bien para este último grupo sanguíneo los resultados no son tan definitivos.

El trabajo pone de manifiesto que, si bien algunos factores no tienen relevancia en la supervivencia renal si son analizados de forma aislada, la combinación de varios de ellos, o la consideración de otros nuevos a partir de los existentes, pueden ofrecer resultados de interés. La construcción de algoritmos de exploración automatizada (como el utilizado en este trabajo), que combine múltiples factores conjuntamente, puede ayudar a identificar efectos que se encuentran anulados o enmascarados por la influencia contraria que ejercen diferentes grupos poblacionales. Los hallazgos obtenidos en este trabajo podrían ayudar a mejorar los protocolos de decisión pretrasplante para la selección del receptor más adecuado.

Agradecimientos

Este trabajo ha sido financiado por la Fundación “Fundación Mutua Madrileña” a través del proyecto VÍA-RENAL.

Referencias

- [1] Bastani B. The Present and Future of Transplant Organ Shortage: Some Potential Remedies. *J Nephrol*, vol. 33, no 3, 2020, pp. 277–288 (ISSN: 1724-6059)
- [2] Saez-Giménez B, Berastegui C, Barrercheguren M, Revilla-López E, Arcos I, Alonso R, Aguilar M, et al. COVID-19 in lung transplant recipients: A multicenter study. *American Journal of Transplantation*, vol. 21, no 5, 2021, pp. 1816-1824 (ISSN: 1600-6135)
- [3] Galvan NTN, Moreno NF, Garza, JE, Bourgeois S, Hemmersbach-Miller M, Murthy B, Timmins K, O’Mahony CA, Anton J, Civitello A, et al. Donor and Transplant Candidate Selection for Solid Organ Transplantation during the COVID-19 Pandemic. *American Journal of Transplantation*, vol 20, no 11, 2020, pp. 3113-3122 (ISSN: 2090-0007)
- [4] Terasaki PI, Gjertson DW, Cecka JM, Takemoto S, Cho YW. Significance of the Donor Age Effect on Kidney Transplants. *Clin Transplant*, vol 11, no 5, 1997, pp. 366–372 (ISSN: 1399-0012).
- [5] Motter JD, Jackson KR, Long JJ, Waldram MM, Orandi BJ, Montgomery RA, Stegall MD, Jordan SC, Benedetti E, Dunn TB, et al. Delayed Graft Function and Acute Rejection Following HLA-Incompatible Living Donor Kidney Transplantation. *American Journal of Transplantation*, vol 21, no 4, 2021, pp. 1612-1621 (ISSN: 1600-6135).
- [6] Kachouie NN, Hamad F. Potential Impact of OPTN Plan in Association with Donors’ Factors on Survival Times of Transplanted Kidneys. *Transplantation Reports*, vol 5, no 2, 2020, pp. 100042 (ISSN: 2451-9596).
- [7] Ng MSY, Ullah S, Wilson G, McDonald S, Sypek M, Mallett AJ. ABO Blood Group Relationships to Kidney Transplant Recipient and Graft Outcomes. *PLOS ONE* 2020, vol 15, no 7, 2020, pp. e0236396 (ISSN: 1932-6203).
- [8] Echterdiek F, Latus J, Döhler B, Schwenger V, Süsal C. Impact of HLA Compatibility in Recipients of Kidneys from Expanded Criteria Donors: A Collaborative Transplant Study Report. *International Journal of Immunogenetics*, vol 48, no 2, 2021, pp. 201-210 (ISSN: 1744-313X).
- [9] Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, vol 53, no 282, 1958, pp. 457-481 (ISSN: 0162-1459).
- [10] Lin RS, León LF Estimation of Treatment Effects in Weighted Log-Rank Tests. *Contemporary Clinical Trials Communications*, vol 8, 2017, pp. 147-155 (ISSN: 2451-8654).
- [11] Sureshkumar KK, Chopra B. Induction Type and Outcomes in HLA-DR Mismatch Kidney Transplantation. *Transplantation Proceedings*, vol 51, no 6, 2019, pp. 1796-1800 (ISSN: 0041-1345).
- [12] Rodrigues S, Escoli R, Eusébio C, Dias L, Almeida M, Martins LS, Pedroso S, Henriques AC, Cabrita AA. Survival Analysis of Living Donor Kidney Transplant. *Transplant Transplantation Proceedings*, vol 51, no 5, 2019, pp. 1575-1578 (ISSN: 1873-2623).

Machine learning algorithms to predict adherence to treatment pathways in secondary prevention of patients with acute coronary syndrome

V. Martínez Seijas¹, G. Moreno Muñoz^{2,3}, M. E. Hernando^{1,4}, G. García Saez^{1,4}, H. Bueno^{2,5}

¹ Grupo de Bioingeniería y Telemedicina, ETSI de Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain
{victoria.martinez.seijas@alumnos.upm.es, mariaelena.hernando@upm.es, gema.garcia.saez@upm.es}

²Servicio de Cardiología, Hospital 12 de Octubre, Madrid, Spain {guimoren@ucm.es, hbueno@cnic.es}

³Facultad de Enfermería, Fisioterapia y Podología, Universidad Complutense de Madrid, Spain.

⁴ Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina, Zaragoza, Spain.

⁵Centro Nacional de Investigaciones Cardiovasculares (CNIC), Centro de Investigación Biomédica en Red en enfermedades Cardiovasculares (CIBERCV), Madrid, Spain.

Abstract

This work presents a system that applies data mining techniques to a dataset of patients admitted for Acute Coronary Syndrome (ACS) to detect the risk of low adherence at six months.

To this end, the development consists of several stages, following the CRISP-DM methodology, including pre-processing and filtering to reduce the dimensionality. Second, different subsets of the data were generated for each of the three adherence variables (to medication, to the Mediterranean diet and physical activity). These subsets are created after statistical analysis and feature selection processes. The training and validation sets followed different alternatives. Third, once all subsets were prepared, different machine learning classifiers were tested, and their performance was studied. As a result, models were generated obtaining a success rate ranging from 70 to 85%.

Knowing these predictions, the professional in charge will be able to make timely interventions to improve the clinical prognosis in non-adherent patients, adapting the recommendations and using additional motivational strategies in the case of non-adherent patients.

1. Introduction

As an introduction to the clinical problem, the term **acute coronary syndrome** (ACS) is defined within the universal definition of acute myocardial infarction as type II and caused by atheroma plaque rupture [1]. The death rate from ACS has declined over the last decades in developed countries. This is due to the increased spread of primary prevention measures, the ageing of the population, and the use of more sensitive analytical tests. Despite this reduction, it still accounts for approximately one-third of all deaths in people aged > 35 years old [2].

On the positive side, 80% of ACS and premature cardiovascular accidents are preventable. A healthy diet, regular physical activity, and smoking cessation are essential. Checking and controlling risk factors for cardiovascular diseases and heart attacks, such as hypertension, high cholesterol levels, and high blood sugar or diabetes, is also essential [3]. Therefore, after an ACS, a series of standardized **guidelines** are proposed by cardiologists to facilitate the decision-making process, which can be pharmacological and non-pharmacological recommendations.

Identifying **precursors** to the occurrence of secondary cardiovascular events could help individualise clinical

follow-up and assess the effectiveness of treatment, as well as optimise resource allocation. Thus, for decreasing the risk of secondary cardiovascular events, a prognostic factor is **adherence** [4] to the recommendations. Some factors are related to non-adherence such as psychological, social, biological aspects, comorbidities, and disease perception [5] [6]. However, it is still unknown if the combination of these factors can predict non-adherence. Creating a model which combines measures of the aforementioned factors could be established to early detect non-adherent patients. Thus, it could help to improve adherence to secondary prevention measures and therefore improve outcomes and prevent secondary events in ACS patients. In this line, the application of data techniques such as machine learning (ML) may help to identify the hidden patterns necessary to create this model to predict the risk of non-adherence.

2. Objectives

The aim is to design and implement a system to predict the risk of low adherence to treatment recommendations in secondary prevention in ACS patients.

3. Methodology

As this present work is a data analysis project, the **CRISP-DM** (Cross Industry Standard Process for Data Mining) methodology [7] has been applied, which is widely used in these cases. It provides a standardised description of the cycle of a data analysis project. Cyclically has different phases:

3.1. Business understanding

This case corresponded to understanding the clinical problem, the real problem that involves the **low adherence** to post-ACS recommendations. The data was extracted from the PLATON study, a research project being carried out at the Cardiology Service of Hospital 12 de Octubre. In the database of this study the information of all psychological variables was collected through questionnaires carried out when the patient is enrolled and during the follow-up visits along with biological, demographical, and socioeconomic information of the patient. Besides, their medical records are also incorporated.

3.2. Data understanding

The cohort was formed by 315 patients participating in the cardiac rehabilitation programme in the study. There were in total 579 cases which include the initial visit at the moment of admission, the first follow-up visits at six months and in some cases the corresponding follow up visit at 12 months. The study is still ongoing so this work is focused on predicting non-adherence after six months.

To describe the sample, concerning sex, only 17.1% of the population are women, the average age is 57.17 for men and 60.92 years old for women, with the highest percentage of patients in the 56-65 age range (39%), the majority (95%) are of Spanish nationality and only 65.9% have ACS with ST-segment elevation.

All the variables collected through answers to the questionnaires indicated or through inferences from the clinical history were **classified** into 5 dimensions: interpersonal, intrapersonal, clinical, and logistic factors and the adherence variables. The latter are three:

- **MEDAS:** questionnaire to evaluate adherence to the Mediterranean diet. It consists of 14 questions to learn about the eating habits of each patient. Each of the questions has been established when it adds up to 1 or 0 (if a series of requirements are met in each item). Subsequently, the final score is calculated, and a cut-off point of 9 is established to separate between patients with low or good adherence [8].
- **MMAS:** the Morisky- Green scale consists of an 8-item questionnaire to determine adherence to the drug regimen. To determine good adherence, a score of 8 is established as a criterion for determining good adherence [9].
- **IPAQ:** the International Physical Activity Questionnaire consists of seven questions referring to the time spent being active in the last week. The number of metabolic equivalents of task determines whether the physical activity is light, moderate, or vigorous [10].

The **non-adherence rates** observed in the sample after answering the questionnaires are collected in *Table 1*.

MEDAS	MMAS	IPAQ
70%	55.1%	18.8%

Table 1: Non-adherence rates.

3.3. Data preparation: processing and filtering.

It includes all the pre-processing steps to organize the data. The SPSS software platform was used. Firstly, a **reduction of the dimensionality** of the characteristics was carried, eliminating those values that were affected by the presence of missing values, i.e., an exclusion criterion was applied for those attributes that were incomplete, answered or collected in less than 30%. Also, identifying the outliers and dealing with them. The data were then standardised with a z-score. Thus, the dataset is reduced from the initial 650 attributes to 95, which include the final scores of the questionnaires in levels after thresholding. Only patients with a visit after six months (168) were considered.

For the following steps, Waikato Environment for Knowledge Analysis (WEKA) v3.8.5 workbench was used with its different panels: Explorer, Experimenter and Knowledge Flow. Moreover, before entering the data into WEKA, different **subsets** are created on which to train.

All variables were stratified according to each of the three variables to be predicted. This **stratification** of the data was done to avoid the class imbalance that occurs when the number of instances of each class (adherent or non-adherent) is very different, thus minimising the classification error. Adherence at six months was characterised in three different domains: adherence to medication (MMAS), to the Mediterranean diet (MEDAS) and physical activity (IPAQ).

Therefore, it was necessary to prepare the subsets for testing the algorithm. First, three subsets were created with the variables that were **statistically correlated** with the class to be predicted (MMAS, MEDAS, and IPAQ). With this aim, a study of the correlation with Spearman's rho and t-Student was performed.

Second, **automatic feature selection** techniques were applied for each of the stickiness variables, creating further subsets. This can be performed with scheme-independent methods (*CfsSubsetEval*) or scheme-specific selection (*Wrapper*). All attribute selection methods perform some **search algorithm** to find the best attributes within the set of available attributes. Some of them are Random Search: (*GeneticSearch*); Sequential search methods: “forward” *GreedyStepwise* or “backwards” *Best First*.

3.4. Modelling

Using WEKA, different modelling techniques were used, the type of algorithms to use was based on the **inductive** learning method because they find a concept description that fits the data, namely: J48, Support Vector Machine, Random Forest, Multilayer perceptron, Naïve Bayes, Logistic and K-nearest neighbours' classifier (IBk).

3.5. Evaluation

5-fold repeated stratified cross-validation was used to evaluate the performance of each ML algorithm with each data set. Also, the confusion matrix was obtained in each case. To understand how predictive the model is, some performance measures are carried out for the classification problem. Among them, they are worth mentioning precision, recall, specificity, F1-score, accuracy, Mathew's correlation coefficient and the receiver operating characteristic (ROC) curves.

4. Results

Once all the data subsets were prepared, they were introduced into each of the classifiers. Different results were obtained after the parameter tuning for each data set.

4.1. Feature extraction

Feature extraction was done from Weka's Explorer option in the attribute selection panel. The feature selection algorithms used were Wrapper (with J48) and CSF Subset

Eval. For each of them the search algorithms Best First, Genetic Search or Greedy Stepwise were used. The number of folds (from 0 to 10) in which the feature was selected during the 10-fold cross-validation is collected. Those that are chosen in at least some folds were selected.

The number of features contained in each of the subsets can be seen in *Table 2*. In this way, it can be seen how the data sets have been reduced from the initial number of variables.

	MMAS	MEDAS	IPAQ
1. Statistically significant	18	14	37
2. CSF with Best First	4	9	9
3. CSF with Genetic Search	21	20	35
4: Wrapper with Best First	16	13	8
5: Wrapper Greedy Stepwise	13	9	7

Table 2. The number of features in each subset.

4.2. Performance

The performance of each classifier for all models of 5-fold cross-validation was evaluated for the three adherence variables, MEDAS, MMAS and IPAQ. The area under the ROC curve (AUC) for the best methods for each of the adherence variables are reported here (and in *Table 3*) due to space constraints:

- For MEDAS adherence variable: IBk classifier in the Wrapper (J48) with Greedy Stepwise subset AUC = 0.6998.
- For MMAS adherence variable: MLP classifier in the Wrapper (J48) with Greedy Stepwise subset AUC = 0.6693.
- For IPAQ adherence variable: Naïve Bayes classifier in the Wrapper (J48) with Best First subset AUC = 0.8672.

Adherence Variable	Best model	Accuracy (%)	Precision	Recall	F score	MCC
MEDAS	IBk	65.56	0.66	0.66	0.65	0.31
MMAS	MLP	66.00	0.66	0.66	0.66	0.32
IPAQ	Naïve Bayes	85.00	0.85	0.85	0.85	0.70

Table 3: Performance of the best models for each adherence variable.

5. Discussion

After applying the classifiers and assessing their performance, none of the prediction methods had consistently superior performance across all the metrics and different subsets. However, Naïve Bayes tended to have slightly superior performance compared to the others. A potential explanation for this is that maybe they are not strong predictors, and most attributes are independent of one from another, considering the small sample size (150 instances for MMAS cases, 90 instances for MEDAS and

60 instances for the IPAQ subsets) and the statistical power.

5.1. The clinical meaning of the variables

Once the results were obtained, it was important to make clinical sense of them. Therefore, the most relevant of the selected characteristics in each of the final best models for each adherence variable was analysed and framed in a clinical context.

MEDAS: The contained features were *NEOFFI_NEURO* (the NEOFFI questionnaire evaluates 5 personality traits), *hypertension*, *alcoholism*, *caregivers' presence*.

The patient's predisposition to neuroticism is a psychological factor that combined with hypertension and tendencies to alcoholism are related to poorer adherence to MEDAS. The absence of caregivers increases nonadherence to the Mediterranean diet (they do not have others to support them in following healthy eating patterns).

MMAS: The 37 characteristics contained in this subset were the following: *weight*, *height*, *obesity*, *pad*, *HDL*, *hba1c*, *smoker*, *ex-smoker*, *Medlife level*, *hypertension*, *atrial fibrillation*, *ventricular fibrillation*, *coronary heart disease*, *sex*, *age*, *NEOFFI_OPENNESS*, *NEOFFI_FRIENDLY*, *BIPQ score*, *PSS score* (Perceived Stress Scale, to measure the degree of stressful situations lived during the last month), *HADS anxiety level*, *negative PANAS scale score* (measures the negative affects), *ethnicity*, *the continent of origin*, *income*, *health coverage*, *employment status*, *type of contract*, *educational level*, *presence of caregivers*, *daily intake*, *background difficulties*, *BMQ score*, *disease information*, *therapy information*, *hospitalisation days*, *hospitalisation satisfaction*, *cognitive coping*.

An important risk factor in ACS is metabolic syndrome. This includes overweight, hypertension and insulin resistance. In this subset of characteristics, increased glycated haemoglobin, overweight (obesity) and decreased HDL make up the profile of the least adherent patient. This profile is associated with patients whose lifestyle habits are far from the Mediterranean dietary guidelines (according to Medlife scores which are lower in fewer adherent patients). In addition, current and past smoking habits are also a risk factor for low adherence. For age and sex, men aged 45-65 years are the least adherent.

About the NEOFFI questionnaire variables, not being open to new experiences and not being friendly are found to be related to lower medication adherence.

Concerning the PSS, more stressed patients are less adherent to medication. The same is true for HADS anxiety level and PANAS negative scale score. Furthermore, the presence of caregivers reveals that family interest in the patient provides care that leads to greater adherence.

IPAQ: the best results were obtained with 8 characteristics: *cholesterol*, *LDL*, *hypertension*, *psychiatric comorbidities*, *Charlson level* (it assesses the life expectancy at 10 years depending on the age and

present comorbidities), *Barthel level* (analyses the degree of dependence in performing daily activities), *health coverage recommended exercise type*.

Patients with low adherence to physical activity have higher serum concentrations of total and LDL cholesterol and higher blood pressure values, consistent with international recommendations for the prevention of atherogenesis.

The Charlson comorbidity index has been shown to hinder adherence to exercise. Thus, the greater the comorbidity, the more this multi-pathology leads to lower adherence to physical activity.

It should be noted that **hypertension** is a characteristic that has been extracted in all three subsets. Hypertension is a classic disease associated with lower adherence, as it can develop prematurely, patients are asymptomatic, and it is necessary to take medication all their lives without perceiving the benefits. This can lead to loss of motivation to continue treatment and poorer adherence. These acquired patterns seem to be repeated in adherence to ACS, i.e., patients maintain these habits after infarction.

To conclude this clinical side, the predictor value is found in the set of variables because none of them exceeds 60% weight. Some complement each other. It is also necessary to consider that the measurements are subjective to the patient's interpretation. A more objective, direct, and cross-sectional approach, which relates to secondary prevention, would be adequate as well.

A strength of this work is that the data set is a real one, therefore, associations between predictors and outcomes are realistic, conferring a diagnostic power that would not be present if the data were automatically generated or invented. The work was fostered by a multidisciplinary group that brings together medical, nursing, psychological and engineering expertise.

6. Conclusions and future lines

The focus of this work was on generating models for early detection of non-adherent patients automatically, based on baseline conditions at the time of occurrence of a cardiovascular event. The proposed models detect behavioural, personality, clinical and social patterns of non-adherent patients concerning three different aspects: medication, nutrition, and physical activity. The data from adherence to medication, diet, and physical activity along with patients' diagnosis, demographics, intra and inter-personal factors were utilized to predict rates of nonadherence along the cohort using machine learning-based classification models. In this case, adherence was studied six months after the hospital admission. The classification models predicted nonadherence with varying performance depending on the variable to predict.

This work has shown the importance of attribute selection and how an appropriate selection can greatly modify classification accuracy. As observed in the results, the best performances are obtained with wrapper methods. In this

way, it has been possible to reduce the algorithmic complexity after the initial pre-processing of the attributes.

Based on this work, several **lines of future work** are open. Firstly, to generalise the results obtained to clinical practice, it would be necessary to increase the sample size to reinforce the relationships obtained. Moreover, the accuracy could be improved by reducing false-positive rates.

On the other hand, it will be necessary to verify the effectiveness of the selection of attributes and classifiers with the best performance in terms of adherence at twelve months.

7. References

- [1] K. Thygesen, J. S. Alpert, A. S. Jaffe, B. R. Chaitman, J. J. Bax, D. A. Morrow y H. D. White, «Consenso ESC 2018 sobre la cuarta definición universal del infarto de miocardio.» *Rev Esp Cardiol*, vol. 72, nº 1, pp. 72 e1-e27, 2019.
- [2] D. Lloyd-Jones, R. J. Adams y e. al., «Heart Disease and Stroke Statistics.» American Heart Association Statistics Committee and Stroke Statistics Subcommittee, 23 Feb 2010.
- [3] WHO, «Cardiovascular diseases» WHO, May 2017. [Online]. [2021 Sept 17].
- [4] E. D. Peterson and A. M. Navar, ““Sticky” issues for adherence in secondary prevention,” *Journal of the American College of Cardiology*, 70(13), Sept 2017.
- [5] Pfizer, “How to measure adherence to treatment,” Pfizer professionals, 01 09 2019.[Accessed 14 04 2021].
- [6] B. Gracey, C. Jones, D. Cho, S. Conner and E. Greene, “Improving Medication Adherence By Better Targeting Interventions Using Artificial Intelligence - A Randomized Control Study,” *Value in Health*, vol. 21, no. Supplement 1, May 2018.
- [7] J. Villena Román, “CRISP-DM: La metodología para poner orden en los proyectos.,” *Singular*, 2 08 2016. [Online]. [Accessed 22 03 2021].
- [8] Schröder, H.; Fitó, M.; Estruch, R.; et al. A short screener is valid for assessing Mediterranean diet adherence among older Spanish men and women. *J. Nutr.* 2011, 141, 1140–1145.
- [9] X. Tan, I. Patel and J. Chang, “Review of the four item Morisky Medication Adherence Scale and eighth item Morisky Medication Adherence Scale,” *Innov Pharm*, vol. 5, no. 3, p. 165, 2014.
- [10] Craig CL, Marshall AL, Oja P. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc.* 2003 Aug;35(8):1381-95.

Deep Learning-Based Segmentation of Head and Neck Organs on CT for Radiotherapy Treatment: Lessons Learned with Clinical Data

L. Cubero¹, E. Cabezas¹, A. Simon², J. Castelli², R. de Crevoisier², J. Serrano³, O. Acosta², J. Pascau^{1,4}

¹ Departamento de Bioingeniería e Ingeniería Aeroespacial, Universidad Carlos III de Madrid, Madrid, España

² CLCC Eugene Marquis, INSERM, LTSI-UMR 1099, 35000 Rennes, France

³ Clínica Universidad de Navarra, Madrid, España

⁴ Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, España

Abstract

Every year, around 500,000 people are diagnosed with head and neck cancer worldwide. Currently, radiotherapy plays a crucial role in the treatment of this type of cancer. Its main drawback, however, is that the healthy tissues surrounding the tumor are also unavoidably affected by the radiation. To minimize this toxicity, it is fundamental to delineate the organs at risk and design a personalized radiotherapy treatment. These segmentations are still mainly performed manually, being both a very time-consuming and subjective procedure. Deep learning algorithms have recently achieved excellent results in the automatic segmentation of different organs on CT images. In this context, we present a deep learning-based workflow to automatically segment four organs at risk in head and neck CT images that could be easily transferred into a clinical context. Moreover, we evaluate and compare the performance of this deep learning model when trained both with a curated database, provided by Hospital Gregorio Marañón, and with original clinical data supplied by CLCC Eugene Marquis of Rennes.

1. Motivation

Every year, more than half a million people worldwide are diagnosed with head and neck cancer [1]. Currently, one of the most effective treatments to deal with this cancer is radiotherapy. This procedure manipulates high-energy radiation from x-rays, gamma rays or protons, to damage the DNA of cancerous cells and prevent cellular growth. Despite its significant advantages, the principal limitation of radiation therapy is that the healthy tissues surrounding the tumor also suffer partial irradiation, which may lead to several complications. It is crucial to accurately define the organs at risk (OAR) and design a personalized radiotherapy treatment to minimize radiation exposure.

With this objective, one of the first steps in any radiotherapy treatment planning is to delineate or segment the tumor and the OAR. This step allows estimating the absorbed dose in each organ and evaluating the treatment's safety and expected efficacy [1]. Nowadays, segmentations in radiotherapy are performed manually, which is a very time-consuming practice limited by extensive inter- and intra-practitioner variabilities [1-2]. Therefore, it is of great interest to automatize the segmentation process to overcome this subjectivity and design more robust radiotherapy treatments focused on minimizing the toxicity and enhancing the benefits for the patient.

Deep learning-based algorithms, particularly convolutional neural networks (CNN), have recently

proven to be one of the most powerful tools to automatically segment different OAR in Computed Tomography (CT) [1-2]. Nonetheless, integrating these algorithms in the radiotherapy workflow is still limited mainly because there is not a unique segmentation protocol, so independent ground truth segmentations of the same organ usually differ significantly. Consequently, medical datasets are commonly not robust enough to train deep learning algorithms. The solution has often been to ask an experienced radiation oncologist or radiologist to resegment the images before training the network, to obtain a homogeneous and consistent database [1].

This work is part of a European agreement between Universidad Carlos III de Madrid and CLCC Eugene Marquis of Rennes. The challenge proposed was to design and implement a deep learning-based workflow to automatically segment the brainstem, spinal cord and parotids from head and neck CTs, that could be used in a clinical context. To accomplish this goal, we trained a deep learning network known as DenseVNet. This CNN has achieved outstanding results in the segmentation of abdominal and chest organs on CT and presents certain advantages compared to other alternatives. DenseVNet has a significantly low number of parameters and includes a feature reuse mechanism and various memory efficiency tools, which allow training the network with small datasets. Furthermore, prior research from our group demonstrated that this network obtains good accuracy segmenting head and neck CT [3], which motivated the choice of the algorithm. In our previous work, the medical images provided by Hospital Gregorio Marañón (HGM) were resegmented with a homogeneous protocol before training to remove any disagreement in the data. On the contrary, the data provided by CLCC Eugene Marquis of Rennes includes clinical images and contours obtained in standard radiotherapy treatment. In this context, the second objective of this study was to evaluate the performance of DenseVNet when trained with non-curated clinical data.

2. Materials and Methods

2.1. Dataset

134 anonymized CTs and their manual segmentations for the brainstem, cord, left and right parotids performed by expert radiotherapists were provided by CLCC Eugene Marquis of Rennes. These images belong to 21 patients with diverse head and neck cancers. For each patient, CTs

were acquired at consecutive weeks along the radiotherapy treatment. The number of CTs available per patient is heterogeneous, ranging from 3 to 6 samples. Images have a fixed matrix size of 512 by 512 voxels in the anterior-posterior (AP) and left-right (LR) axes, while the inferior-superior (IS) dimension ranges from 104 to 225 voxels. This implies that the field of view, which is the size of the real space captured by the acquisition, is different in each image. The data were randomly divided into train, validation and test. Instead of directly separating the CTs, the patients were divided into those three sets to avoid similar images (from the same patient) being both in the train and test sets, which would yield biased results. Finally, the train set included 11 patients and 68 images, the test set 4 patients and 28 images, and the validation one 6 patients and 38 images.

The segmentations were provided as 3D binary masks. No information was given about the number of practitioners who segmented the images. The congruence of the ground truth contours was assessed for validity to identify possible inaccuracies or observer-dependent subjectivities. With this goal, each patient's CTs and segmentations were registered with an affine transformation using the first image acquired as reference. As the segmentations depict anatomical volumes, the shape and size of the OAR of a particular patient were expected to be almost identical among that patient's images. Then, a comparative assessment was made by computing the Dice Score Coefficient (DSC), which is the relative spatial overlap between two binary volumes [2], [4]; the 95% Hausdorff distance (95% HD), that corresponds to the 95-th percentile of the Euclidean distance between the boundary voxels of two binary masks [2], [4]; and the Percentage Volume Difference (PVD), which indicates how similar the volumes of two segmentations are. Finally, a descriptive statistical analysis was carried out for each metric, including minimums and maximums, mean values and standard deviations.

Before training, we applied data augmentation to the images and their respective segmentations. This technique involves slight modifications to the existing data to obtain new images and expand the dataset. For this study, a sequence of augmentations was applied to random batches of data, including elastic deformations, scalings, rotations, mirroring and brightness transformations, nonlinear intensity gamma transformations, and Gaussian noise. Overall, the training dataset size increased by around 40%.

2.2. Deep learning-based workflow

A six-step workflow including two different deep learning algorithms was designed and integrated into an interface. For each CT image, the steps are the following:

1. Preprocessing: First, the CT is downsampled to a lower resolution of 144x144x144 due to memory constraints in the GPU. Then, the image is cropped to the region of interest to prevent information loss, ensuring that the patient's head is included entirely through thresholding and morphological processing.

2. Landmark Network or LandmarkNet: The downsampled CT is the input to LandmarkNet, a simple CNN trained to predict the central point's coordinates of each OAR. This network is built by seven subsequent convolutional layers and two dense layers. Each convolutional layer is a composite function of a 3D convolution, a batch normalization and a rectified linear unit. The network was trained for 10000 epochs with an Adam optimizer of learning rate 0.0001. The loss function used was the mean square error, and the metric to evaluate the performance in the validation set was the mean absolute error. Therefore, LandmarkNet takes as input a CT and outputs a set of coordinates corresponding to the OAR's central points.

3. Bounding boxes: Then, a bounding box is cropped for each CT and OAR, centered in the predicted coordinates for that OAR, and then resized again to 144x144x144. The bounding boxes have a fixed side size of 55 millimeters for the brainstem and parotids. For the spinal cord, the size is fixed to 90 millimeters in the AP and LR dimensions but ranges from 90 to 170 millimeters in the IS axis, depending on the cord's length. Consequently, for each image in the workflow, four different bounding boxes are constructed.

4. Dense-V Network or DenseV-Net: The bounding boxes are the input to DenseVNet, a fully CNN developed by Eli Gibson [2], [5] (Figure 1). Each convolutional unit consists of a 3D convolution with a learned kernel, a batch normalization and a rectified linear unit.

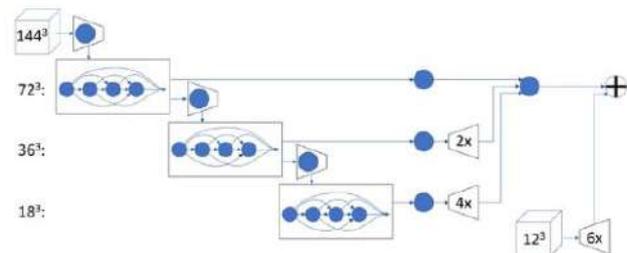


Figure 1. DenseVNet architecture

This architecture presents five advantageous key components. First, it uses batch-wise spatial dropout, a regularization technique that prevents overfitting and improves computational and memory efficiency [2]. Secondly, dense feature stacks are introduced, which are a string of composite functions in which the input of each function is the concatenated outputs of the previous ones. This feature ensures maximum information flow between the layers and significantly reduces the network's number of parameters. Thirdly, the architecture has a V-shape comprising a downsampling and an upsampling path with skip connections, allowing higher resolution information to be propagated to the final output. Fourthly, dilated convolutions are used to represent functions with large receptive fields but few trainable parameters. Finally, an explicit spatial prior is added to the output to give context to the segmentations.

After various trials, the best results were obtained when a single-class DenseVNet was trained for each OAR. The network's input is a single bounding box, while the output corresponds to the OAR binary segmentation. Therefore, four independent networks were trained using the weighted sum of an L2 regularization with label smoothed

probabilistic Dice score as the loss function [2], [5], and an Adam optimizer with varying learning rate. The initial number of epochs was 10000, although every training converged earlier, taking approximately only 2 hours to train each instance on an Nvidia Titan V GPU.

5. Postprocessing and evaluation: Next, the four output binary segmentation are upsampled back to their original CT resolution. For every image in the test set, the predicted segmentations and the ground truth volumes were compared by computing the DSC, the 95% HD and the PVD (explained in section 2.1).

6. 3D Slicer Module: Finally, the previous steps were integrated into a 3D Slicer interface, which allows the user to visualize a CT and segment the brainstem, cord and parotids. The segmentations are obtained in less than 30 seconds on a CPU, a very good figure compared to the hours it takes a radiotherapist to perform this manually. This module depicts a possibility to easily integrate deep learning algorithms into the radiotherapy workflow.

2.3. Experiments

Only the two most relevant experiments performed will be presented. On the one hand, LandmarkNet and DenseVNet were trained with the medical imaging database provided by CLCC Eugene Marquis of Rennes. The trained models were then evaluated in the test set by measuring the DSC, 95% HD and PVD. On the other hand, the models that our group had previously trained and tested in the external dataset HGM [3] were used to predict the segmentations of the images in Rennes' test set. In this way, networks trained with completely different databases were used to evaluate the same test images. It must be highlighted that whereas the data from Rennes was unaltered and presented various clinical variabilities, the images from HGM had been resegmented before training with a consistent and homogeneous approach.

3. Results and Discussion

3.1. Ground truth segmentations evaluation

The evaluation of the provided ground truth segmentations depicted substantial variabilities among the data. Overall, the comparison yielded a DSC of 0.86, a 95% HD of 6.01 mm, and a PVD of 11.66%. These differences were more significant in the cord's volumes, for which the DSC was around 0.75 and the 95% HD 14 mm. An example of these disagreements is shown in Figure 2. Moreover, the CTs belonging to the test set comprise the most conflictive ground truth volumes, particularly for the cord and the parotids. This suggest that different users probably

performed the segmentations. Unfortunately, these errors will likely limit and bias DenseVNet's predictions.

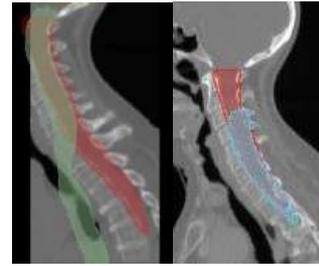


Figure 2. Discrepancies in the cord's ground truth segmentations in Rennes' database

3.2. Evaluation and comparison of the models trained with Rennes and HGM databases

On the one hand, the models trained and tested with Rennes' database images were evaluated. First, the predicted coordinates by LandmarkNet were compared to the actual central points by measuring the Euclidean distances between them. Overall, the predictions showed mean distance errors of less than 1 centimeter, which is an acceptable result as these coordinates were only used to build the bounding boxes. The most significant errors were identified in the prediction of the cord. Then, the predicted segmentations of DenseVNet and the ground truth volumes were compared, yielding an overall DSC of 0.56, a 95% HD of 21.88 mm and a PVD of 51.95% in the test set (Table 1). The worst results corresponded to the right parotid's segmentations, with meaningful differences in accuracy between the two parotids, a surprising result since similar values were expected.

On the other hand, the performance of the pre-trained models provided by HGM and tested with Rennes' database images was evaluated. The segmentations' predictions in the test set yielded an overall DSC of 0.60, a 95% HD of 19.23 mm and a PVD of 46.99% (Table 1). The worst segmentations corresponded to the cord. The resulting accuracies are worse compared to the reported model's performance when tested in their database, for which the DSC values were 0.81 for the brainstem, 0.81 for the cord, 0.73 for the left parotid and 0.75 for the right one [3]. This downgrade was probably primarily influenced by the variances in the ground truth segmentations of Rennes' database, more profound in the case of the cord. Nonetheless, considering that these models were trained with a different dataset, the results are still quite optimistic as they show the ability of the trained network to predict external data.

		Brainstem	Cord	Left Parotid	Right Parotid
Rennes	DSC	0.62 ± 0.15	0.57 ± 0.18	0.55 ± 0.24	0.52 ± 0.15
	95% HD (mm)	18.01 ± 14.83	26.42 ± 13.83	14.01 ± 10.69	29.90 ± 19.47
	PVD (%)	63.61 ± 30.57	50.24 ± 31.84	50.30 ± 39.15	43.67 ± 41.31
HGM	DSC	0.67 ± 0.14	0.51 ± 0.26	0.56 ± 0.22	0.65 ± 0.10
	95% HD (mm)	8.57 ± 5.49	42.72 ± 47.35	13.28 ± 12.09	12.36 ± 12.10
	PVD (%)	49.37 ± 40.63	80.40 ± 67.28	37.35 ± 42.62	20.87 ± 18.43

Table 1. Evaluation of the predicted segmentations obtained with Rennes and HGM algorithms in the test set

Furthermore, the results from the trained Rennes and pre-trained HGM DenseVNet models showed large standard deviations in some of the OAR's metrics. This motivated the analysis of the predicted segmentations patient by patient to determine whether certain cases were downgrading the accuracy. This evaluation showed that there were specific patients for which the models were performing especially poorly, yielding DSC smaller than 0.1, and were consequently considered as not valid segmentations. A well-segmented image can be observed in Figure 3, whereas a non-valid example is depicted in Figure 4. In this case, the patient had a massive tumor near the left parotid, so the anatomy and consequently the segmentation of that OAR differed significantly. Excluding some of these patients largely increased the prediction accuracy with both Rennes and HGM models, particularly for the left parotid. Figure 5 depicts the improvement in accuracy when the patients presenting a DSC below 0.1 are dismissed.

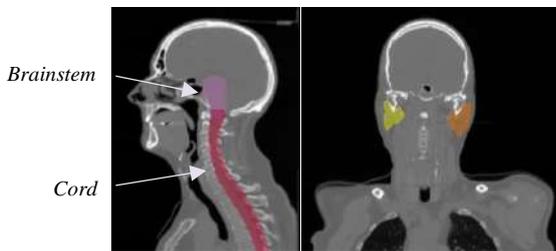


Figure 3. Segmentations acquired with Rennes' models

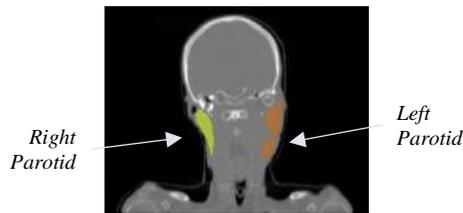


Figure 4. Predicted parotid segmentations with Rennes' models presenting a DSC of 0.1. The patient presents a large tumor close to the left parotid, which complicates the prediction

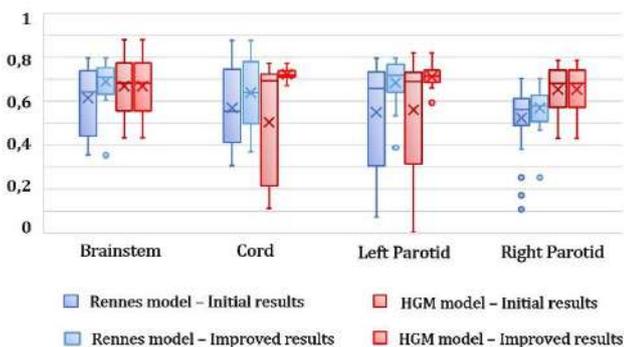


Figure 5. DSC improvement after removing the non-valid segmentations (DSC < 0.1) with both Rennes and HGM models

4. Conclusions

This study presented a deep learning-based workflow to automatically segment four different OAR from head and neck CT images. After analyzing the results, we can extract three main conclusions.

Firstly, the workflow implemented has the potential to be translated into the clinic. This would facilitate the

segmentation protocol in radiotherapy, as it would allow obtaining more objective and robust segmentations that would help both the medical team and the patient. Using the developed 3D Slicer interface, the radiation oncologist could calculate the segmentations in less than 30 seconds running on CPU, which usually would take a few hours manually. The results could be imported into the Treatment Planning System using DICOM.RT structures.

The second conclusion is that it is fundamental to build a solid database to train and test 3D convolutional neural networks. The data provided by CLCC Eugene Marquis of Rennes presents large variations that limited the networks' performance. Even if this data reflected the variability and subjectivity of the segmentation methodology and the difficulty of evaluating the performance of such artificial intelligence models in a clinical environment, it is necessary to use a consistent database to train deep learning algorithms. It would be convenient to ask an expert to resegment the CTs with a homogeneous approach for further use of the dataset.

Finally, we can confirm that DenseVNet is a good network architecture for this problem. On the one hand, the results obtained with Rennes models show that even when trained with irregular data, the network can yield acceptable approximations of the expected segmentations. On the other hand, the pre-trained HGM models produced good enough results even when trained with a different database and when tested on conflictive images. These results, together with the significant advantages that the architecture presents, motivate the further study and use of DenseVNet for OAR contouring in radiotherapy.

Acknowledgements

Research supported by projects PI18/01625 and AC20/00102 (Ministerio de Ciencia, Innovación y Universidades, Instituto de Salud Carlos III, Asociación Española Contra el Cáncer and European Regional Development Fund "Una manera de hacer Europa"), project PerPlanRT (ERA Permed) and EU Horizon 2020 research and innovation programme Conex plus UC3M (grant agreement 801538).

References

- [1] S. Nikolov et al., "Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy," ArXiv, vol. abs/1809.04430, 2018.
- [2] E. Gibson et al., "Automatic Multi-Organ Segmentation on Abdominal CT With Dense V-Networks," IEEE Transactions on Medical Imaging, vol. 37, no. 8, pp. 1822–1834, 2018.
- [3] "CARS 2019—Computer Assisted Radiology and Surgery Proceedings of the 33rd International Congress and Exhibition, Rennes, France, June 18–21, 2019," International Journal of Computer Assisted Radiology and Surgery, vol. 14, no. 1, pp. 1–194, Jun. 2019.
- [4] J. Wallner, I. Mischak, and Jan Egger, "Computed tomography data collection of the complete human mandible and valid clinical ground truth models," Scientific Data, vol. 6, no. 1, p. 190 003, Jan. 2019.
- [5] E. Gibson et al., "NiftyNet: A deeplearning platform for medical imaging," Computer Methods and Programs in Biomedicine, vol. 158, 2017.

Propuesta de modelos predictivos en salud mental para la personalización de terapias de rehabilitación en pacientes con adicción a la cocaína

J. Tapia-Galisteo^{1,2}, J.M. Iniesta¹, C. Pérez-Gandía^{2,1}, G. García-Sáez^{1,2}, D. Urgelés³, M.E. Hernando^{1,2}

¹ Grupo de Bioingeniería y Telemedicina (GBT), Centro de Tecnología Biomédica, ETSI de Telecomunicación, Universidad Politécnica de Madrid, España, {jtapia, jminiesta, cperez, ggarcia, elena}@gbt.tfo.upm.es

² Centro de Investigación Biomédica en Red de Bioingeniería, Biomateriales y Nanomedicina, CIBER, Madrid, España

³ Servicio de Psiquiatría. Responsable de I+D+I. Nuestra Señora de la Paz, Madrid, España, diego.urgeles@sjd.es

Resumen

La drogodependencia es un importante reto que presiona a los sistemas sanitarios para que respondan con tratamientos efectivos. Para mejorar las tasas de éxito en el tratamiento de pacientes drogodependientes es necesario identificar los principales factores de riesgo asociados al abandono y las investigaciones previas son escasas y usan una gran variedad de enfoques. Este trabajo presenta: 1) El diseño y selección de un modelo predictivo del éxito en el tratamiento residencial para pacientes con adicción a la cocaína a partir de un conjunto heterogéneo de características de gran dimensionalidad, con el objetivo de aprender nuevas asociaciones entre variables independientes; 2) Una herramienta de extracción automática de la información, para obtener mayores volúmenes de datos, reduciendo tiempo y costes. Se evaluaron diferentes técnicas de selección de características y algoritmos de aprendizaje automático, obteniendo mayor rendimiento el algoritmo Bosques Aleatorios: precisión 82%, área bajo la curva 0.81, sensibilidad 0.96, especificidad 0.47, F1-score 0.89 y coeficiente de correlación de Matthews 0.53. El conjunto final empleado contiene 11 variables, referentes a tratamientos previos, exploración mental, funcionamiento cognitivo, personalidad, hábitos de consumo y tratamiento farmacológico. La herramienta de extracción automática obtuvo una precisión >90%, cercano al rendimiento humano. El uso del predictor en la rutina clínica habitual podría suponer una reducción de la tasa de abandono del 36.82% y una reducción de los costes asociados al abandono del 5.48%, lo que podría representar un nuevo paso en la gestión del tratamiento, especialmente si se incluyera en sistemas de ayuda a la decisión clínica.

1. Introducción

La elevada prevalencia del consumo y adicción a la cocaína es un importante reto sanitario que ejerce presión sobre los sistemas sanitarios ante la necesidad de responder con tratamientos efectivos. En 2018, 19 millones de personas eran consumidores de cocaína a nivel mundial (0.4% de la población mundial de 15-64 años) [1]. Se estima que 18 millones de europeos han consumido cocaína en su vida, siendo España el sexto país de la Unión Europea con mayor consumo (prevalencia del 2.8% en adultos jóvenes) [2].

Las adicciones representan un gran coste para los sistemas sociosanitarios debido a los tratamientos médicos y farmacológicos y a las complicaciones asociadas: enfermedades y pruebas médicas necesarias, la repercusión en la salud mental y el bienestar de los seres queridos y la

sociedad, el impacto en la violencia y el sistema jurídico penal, y la pérdida de productividad asociada a la discapacidad y a las muertes prematuras.

El acceso a los tratamientos para la rehabilitación por drogadicción sigue siendo insuficiente y desigual, careciendo de la cobertura y especificidad necesarias. En Europa, la mayoría de los tratamientos para la adicción son ambulatorios, mientras que sólo una pequeña parte son residenciales [1]. Las rutas terapéuticas se caracterizan por el uso de diferentes servicios, por recaídas que conllevan un reingreso y por estancias de diferente duración.

La elevada tasa de abandono es una de las principales limitaciones del éxito del tratamiento, oscilando entre 23-50% [3] en tratamiento ambulatorio y entre 17-57% [4] en tratamiento residencial. La finalización de este se asocia con mayores niveles de abstinencia, menores recaídas e índices de delincuencia y mejores niveles de empleo [5].

Los factores de riesgo de abandono más consistentes en la literatura son los déficits cognitivos, la baja adherencia al tratamiento, los trastornos de personalidad y la edad [5].

La generación de modelos predictivos exitosos permitiría comprender mejor el comportamiento de los pacientes con problemas de adicción y podría ayudar a identificar los factores de riesgo más importantes que impiden el tratamiento. En la investigación médica y psicológica generalmente las interacciones entre variables heterogéneas independientes pueden ofrecer resultados potenciales que ayuden a comprender mejor los factores que explican el abandono del tratamiento. El desarrollo de sistemas de ayuda a la decisión basados en modelos predictivos se ha convertido en un tema de enorme interés científico en diversos campos de la medicina.

Los objetivos de esta investigación son: 1) Desarrollar un modelo predictivo utilizando Aprendizaje Automático (AA) con datos heterogéneos de gran dimensionalidad para estimar el éxito en el tratamiento residencial en adicción a la cocaína, como primer paso hacia el desarrollo de un sistema de ayuda a la decisión para los clínicos encargados de guiar a pacientes en su ruta terapéutica; 2) Desarrollar y evaluar una herramienta de extracción automática de la información para estructurar mayores volúmenes de información histórica, reduciendo tiempo y costes.

Las principales contribuciones de este trabajo de investigación son: 1) Hasta donde sabemos, este es el primer estudio que utiliza técnicas de AA en población europea con adicción a cocaína; 2) Utiliza el mayor tamaño muestral en este contexto (253 sujetos); 3) Hasta donde sabemos, es el primer estudio que considera las características de los tratamientos previos junto con variables mentales, cognitivas, de personalidad y farmacológicas en adicción a la cocaína; 4) El modelo predictivo desarrollado se discute en términos de posibles beneficios obtenidos de su uso en la práctica clínica; 5) Se implementa y evalúa una herramienta de extracción automática de la información.

2. Antecedentes

En los últimos tiempos se ha producido un movimiento emergente del AA en la psiquiatría computacional y en las neurociencias computacionales, situándose como una herramienta que facilite la traslación de la investigación en neurociencia hacia las prácticas clínicas en psiquiatría.

Existe un número reducido de trabajos en la literatura que hacen uso del AA sobre datos de adicciones. Sólo se han identificado 16 trabajos que combinan metodologías de AA incluyendo la selección de características, seis de ellos tratan de consumo de cocaína y 1 de policonsumo incluyendo cocaína. Los resultados son prometedores pero la comparación entre ellos es complicada, debido a la naturaleza diversa de los estudios realizados (sustancias, variables, población, diseño de estudios y métricas).

Esta heterogeneidad en la literatura, junto con el avance de las técnicas de AA y su aplicación en el campo de las adicciones, pone de relieve la necesidad de realizar más investigaciones que permitan representar el potencial de estas técnicas en el campo de la psiquiatría de adicciones.

3. Materiales y métodos

3.1. Extracción de la información

La derivación de pacientes a tratamiento residencial ocurre cuando no consigue lograrse la abstinencia de manera ambulatoria. Para ello es necesario que los médicos a cargo del paciente rellenen un informe de derivación para recomendar su ingreso en el centro residencial.

En la Comunidad de Madrid, el informe de derivación es un documento de texto formado por tablas y campos de texto libre. Para la extracción y estructuración de la información fue necesaria la colaboración clínica con el fin de definir las variables de interés y las métricas asociadas. Se identificaron un total de 121 características como potenciales predictores del éxito en el tratamiento. El conjunto de datos inicial fue creado manualmente por 3 investigadores que se encargaron de leer los informes de derivación y convertirlos en una base de datos estructurada.

La variable objetivo, seleccionada para la predicción, fue el éxito en el tratamiento residencial, entendido como la finalización del mismo. Los casos en los que el paciente abandonó el tratamiento prematuramente, de forma voluntaria en contra del consejo profesional o por expulsión debido a un incumplimiento de las normas de conducta, se consideraron como un tratamiento sin éxito.

3.2. Descripción de las variables de interés

El conjunto total de datos utilizado se compone de los informes de derivación de **253 pacientes** (72% éxito-28% abandono) ingresados en el Centro de Atención Integral al Cocainómano (CAIC) de la Clínica Nuestra Señora de la Paz. Las variables identificadas se pueden agrupar en 8 dimensiones: sociodemográficas (4), exploración mental y funcionamiento cognitivo (18), valoración de la personalidad (1), habilidades/competencias sociales, cognitivas y afectivas (9), historia de consumo (63), tratamiento farmacológico (14), tratamientos previos en adicciones (7) y otros parámetros individuales (5).

3.3. Procesamiento y criterios de filtrado

Se realizó un procesamiento para descartar variables, tomando como criterios de exclusión: >25% de valores ausentes; > 90% de registros en una misma categoría; variables nominales con >20 categorías; coeficiente de variación <0.1; coeficiente de correlación ≥ 0.66 (empleando Pearson entre variables continuas y Spearman entre variables ordinales y entre ordinales y continuas). Como resultado, se redujo el conjunto a 60 variables.

En los pasos siguientes se normalizaron los valores de las variables numéricas mediante puntuaciones z (media 0.0 y desviación estándar 1.0) y se transformaron las variables nominales a valores numéricos.

3.4. Metodología de aprendizaje automático

Se emplearon diferentes tipologías de algoritmos de AA en combinación con diferentes técnicas de selección de características, con el fin de evaluar las diferentes combinaciones en competencia y seleccionar el predictor y el conjunto de características con mejores resultados.

Se seleccionaron 4 tipologías de algoritmos en base a su utilización y resultados obtenidos en los trabajos de la literatura: Regresión Logística (RL), Bosques Aleatorios (BA), Máquinas de Vectores de Soporte (MVS) y redes neuronales de tipo Perceptrón Multicapa (PMC).

Los procesos de selección de características se componen de un algoritmo de selección de subconjuntos de características y de un evaluador. Entre los algoritmos de selección se decidió emplear: 1) Best First (BF) con búsqueda secuencial hacia adelante y criterio de parada 15 iteraciones máximas sin mejora; 2) Búsqueda genética (G) con criterio de parada 40 generaciones máximas a evaluar.

Entre los evaluadores se decidió emplear: 1) Filtro multivariante basado en correlaciones (CFS); 2) Wrapper (W), en el que el algoritmo de selección de características utiliza el propio algoritmo de aprendizaje como parte de la evaluación del subconjunto; 3) Incremental Wrapper Subset Selection (IWSS): método híbrido que combina el uso de un filtro, para establecer una clasificación de las características basada en correlación, con un wrapper.

La selección de características se hizo mediante validación cruzada (VC) de 10 ramas en todas las combinaciones mencionadas. En cada conjunto final (CS) se incluyeron las característica seleccionadas en al menos 8 ramas, siguiendo una estrategia secuencial para las demás.

3.5. Evaluación de los modelos predictivos

Cada algoritmo de AA es evaluado con los diferentes conjuntos de variables obtenidos mediante una VC estratificada de 10 ramas repetida 10 veces. Los resultados se comparan con un clasificador de tipo Zero Rule, el cual predice simplemente la clase mayoritaria.

Como medidas de evaluación se decidió emplear: exactitud, área bajo la curva (AUC), sensibilidad, especificidad, F1-score, y Coeficiente de Correlación de Matthews (Matthews Correlation Coefficient), calculando el intervalo de confianza del 95% para la media.

3.6. Extracción automática de la información

Se planteó la creación de una herramienta de extracción automática de la información, con el fin de poder obtener mayores cantidades de información estructurada. Se utilizó Python para el desarrollo del código y la interfaz gráfica.

El proceso para la extracción y almacenamiento de la información es el siguiente: 1) Lectura de informes y división en bloques; 2) Transformación de bloques en dataframes; 3) Procesamiento de dataframes para extraer variables (de tablas o del lenguaje natural); 4) Exportación de información estructurada a Excel.

Para la creación de la herramienta se emplearon 108 informes y para su validación 145, y se compararon ambas extracciones (manual y automática), obteniendo sendas tasas de precisión y discrepancia.

4. Resultados

4.1. Rendimiento de los algoritmos de AA

La Tabla 1 muestra los resultados de cada algoritmo de AA junto al mejor conjunto de características en cada caso.

	AUC	Exactitud	Sensibili.	Especifici.	F1-score	MCC
Zero Rule	0.50	71.6±0.3	1.00	0.00	0.83	---
BA-CSWBF	0.81±0.02 *	82.1±1.1 *	0.96±0.01	0.47±0.03*	0.89±0.01*	0.53±0.03
PMC-CSWG	0.78±0.02 *	76.5±1.5	0.86±0.02	0.52±0.03*	0.84±0.01*	0.41±0.04
RL-CSIWSS	0.73±0.02 *	74.2±1.2	0.91±0.01	0.32±0.03*	0.83±0.01	0.29±0.04
MVS-CSWBF	0.63±0.02 *	77.5±1.2 *	0.97±0.01	0.29±0.03*	0.86±0.01	0.38±0.04

Tabla 1. Resultados obtenidos con los algoritmos de AA. CSWBF: Conjunto obtenido con W y BF; CSIWSS: Conjunto obtenido con IWSS; CSWG: Conjunto obtenido con W y G.

El modelo con mejor rendimiento es el algoritmo BA con el CSWBF, siendo vencedor en casi todas las medidas de evaluación. Además, es el único que obtiene una mejora estadísticamente significativa en F1-score y cuyo MCC muestra un nivel de correlación moderadamente positivo. El conjunto CSWBF consta de 11 característica, y en la Figura 1 puede verse su importancia relativa dentro del BA.

Se realizó un análisis del balance de costes entre falsos negativos (FN) y falsos positivos (FP), con el fin de reducir los primeros. En la Tabla 2 puede verse una reducción del 2% en FN a costa de disminuir los demás parámetros.



Figura 1. Importancia relativa de las características

FN/FP	Exactitud	Sensibili.	Especifici.	F1-score	MCC
BA-CSWBF sensible a los costes					
1.0/1.0	82.1±1.1 *	0.96±0.01	0.47±0.03*	0.89±0.01*	0.53±0.03
1.3/1.0	80.2±1.2 *	0.97±0.01	0.38±0.04*	0.88±0.01*	0.47±0.04
1.9/1.0	76.7±1.0 *	0.98±0.01	0.24±0.03*	0.86±0.01	0.36±0.04

Tabla 2. Resultados BA-CSWBF variando la matriz de costes.

4.2. Impacto del predictor en la tasa de abandono

La Tabla 3 muestra las posibles reducciones en la tasa de abandono al usar el predictor desarrollado para decidir la derivación de pacientes (versiones con 4% y 2% de FN).

	Falsos negativos	Tasa de abandono	Reducción lograda
Sin modelo	---	28,46%	---
Con modelo	4%	17,98%	36,82%
	2%	23,59%	17,11%

Tabla 3. Impacto del predictor en la tasa de abandono.

4.3. Impacto del predictor en los costes

La Tabla 4 muestra los posibles ahorros asociados al abandono al usar el predictor desarrollado para decidir la derivación de pacientes (versiones con 4% y 2% de FN). Estos han sido calculados respecto de la dotación económica anual del CAIC y las tasas de abandono previas.

	FN	N.º Pacientes/año	Coste Anual	Ahorro
Éxito terapéutico	---	146	1.361.525€	---
Abandono	---	58	208.039€	---
Éxito terapéutico	4%	155	1.445.455€	86.082€
Abandono	4%	34	121.949€	---
Éxito terapéutico	2%	151	1.408.153€	40.041€
Abandono	2%	46	164.990€	---

Tabla 4. Impacto del predictor en los costes del abandono.

4.4. Rendimiento de la herramienta de extracción

La Tabla 5 muestra la precisión y discrepancias obtenidas en la creación y validación, por dimensiones y en total.

	Set creación (108)		Set validación (145)	
	Precisión	Discrepancias	Precisión	Discrepancias
Sociodemográficos	86.81%	13.19%	85.52%	14.48%
Mental/cognitivo	98.58%	1.42%	96.76%	3.24%
Personalidad	97.23%	2.77%	95.17%	4.83%
Competencias	98.87%	1.13%	97.09%	2.91%
Historia consumo	93.21%	6.79%	86.04%	13.96%
Fármacos	100.00%	0.00%	93.80%	6.20%
Tratamient. previos	90.59%	9.41%	79.20%	20.80%
Otros parámetros	99.39%	0.61%	93.10%	6.90%
Total	96.24%	3.76%	91.59%	8.41%

Tabla 5. Niveles de precisión de la herramienta por dimensión.

5. Discusión

Los diferentes conjuntos de características obtenidos en los procesos de selección tienen una composición heterogénea, reafirmando así la necesidad de incluir información conjunta de diferentes dimensiones para optimizar los resultados. El mejor conjunto (CS_{WBF}) contiene características de 5 de las 8 dimensiones incluidas.

La valoración de la personalidad fue incluida en todos los conjuntos, en línea con la literatura como uno de los principales factores de riesgo de abandono. De igual forma ocurre con los trastornos cognitivos, confirmando su relación con el éxito en el tratamiento, especialmente el pensamiento, la orientación y la capacidad de juicio. También el fármaco disulfiram ha mostrado su importancia en el conjunto final, en línea con la literatura y sus buenos resultados en la disminución del consumo de cocaína. Los factores de tratamientos previos también han despuntado, siendo los tres más importantes en el BA-CS_{WBF}.

La comparación con los trabajos de la literatura es complicada, dado que la mayoría se centran en datos de IRMf y en clasificar sujetos dependientes y sanos. El trabajo [7] trata de predecir abstinencia durante y tras el tratamiento, lo cual podría asemejarse en mayor medida al trabajo aquí presentado, obteniendo tasas de exactitud inferiores (64-71%). El trabajo [8] trata de predecir éxito en el tratamiento ambulatorio de policonsumidores, obteniendo también grandes resultados con el algoritmo BA, similares a los aquí presentados.

Las altas tasas de abandono en los tratamientos suponen una oportunidad de mejora, y su reducción tiene un impacto directo en los tiempos de espera para el acceso al tratamiento, en los futuros reingresos y se asocia con mayores niveles de abstinencia. Una reducción de la tasa de abandono de más de un tercio, pasando del 28.46% actual al 17.98% con el predictor, supondría un mejor aprovechamiento de los recursos, un mayor porcentaje de pacientes terminando con éxito y las posteriores sinergias.

El impacto económico del consumo de drogas es tremendamente difícil de cuantificar. El fracaso en la rehabilitación conlleva un gran número de costes futuros para la sociedad y cuanto antes se produzca la rehabilitación en el tiempo mayor es el ahorro futuro. La inclusión del modelo predictivo desarrollado en la práctica clínica habitual podría suponer un ahorro asociado al abandono del 5.48% del presupuesto total anual del CAIC.

La herramienta de extracción automática de la información desarrollada obtiene una precisión >90% en todos los conjuntos, lo cual es propio de los buenos sistemas de extracción, acercándose al rendimiento humano. Su uso permitiría aumentar el volumen de datos disponibles y extraer el histórico de información, suponiendo además un ahorro en tiempo y costes, y unificando criterios. La existencia de una interfaz gráfica permitiría incluirlo en los procesos de gestión clínica como una valiosa aportación.

6. Conclusiones

Es posible predecir el éxito en el tratamiento residencial de pacientes con adicción a la cocaína antes de la

rehabilitación. El mejor modelo predictivo es el resultante de la combinación del algoritmo BA con el conjunto de características CS_{WBF}, el cual presentó la mayor exactitud (82.12%), F1-score (0.89), MCC (0.53) y AUC (0.81), así como una elevada sensibilidad (0.96).

La implementación de una herramienta de extracción automática de la información y su posible inclusión en el flujo de trabajo actual se presenta como un método prometedor para la estructuración de mayores volúmenes de datos, con beneficios para el desarrollo de futuras investigaciones y para la gestión clínica, así como implicando importantes ahorros en costes y tiempo.

Puede concluirse, por tanto, que el uso de algoritmos de AA en conjunto con técnicas de selección de características sobre un conjunto de variables heterogéneas de diferentes dimensiones ha mostrado resultados útiles para la predicción del éxito terapéutico en pacientes adictos a la cocaína. Este trabajo de investigación ratifica que el algoritmo BA tiene una gran capacidad de adaptación a conjuntos de datos con relaciones complejas, condición inherente a la psiquiatría y la drogadicción.

Referencias

- [1] UNODC, *World Drug Report 2020 (Booklet 2). Drug use and health consequences*. United Nations publication, 2020.
- [2] European Monitoring Centre for Drugs and Drug Addiction, "Spain Country Drug Report 2019," vol. 2017, pp. 1–33, 2019.
- [3] G. A. Dakof, M. Tejada, and H. A. Liddle, "Predictors of engagement in adolescent drug abuse treatment," *J. Am. Acad. Child Adolesc. Psychiatry*, vol. 40, no. 3, pp. 274–281, Mar. 2001, doi: 10.1097/00004583-200103000-00006.
- [4] F. P. Deane, D. J. Wootton, C. I. Hsu, and P. J. Kelly, "Predicting dropout in the first 3 months of 12-step residential drug and alcohol treatment in an Australian sample," *J. Stud. Alcohol Drugs*, vol. 73, no. 2, pp. 216–225, Jan. 2012, doi: 10.15288/jsad.2012.73.216.
- [5] H. H. Brorson, E. Ajo Arnevik, K. Rand-Hendriksen, and F. Duckert, "Drop-out from addiction treatment: A systematic review of risk factors," *Clinical Psychology Review*, vol. 33, no. 8, Elsevier Inc., pp. 1010–1024, 01-Dec-2013, doi: 10.1016/j.cpr.2013.07.007.
- [6] K. K. Mak, K. Lee, and C. Park, "Applications of machine learning in addiction studies: A systematic review," *Psychiatry Res.*, vol. 275, pp. 53–60, May 2019, doi: 10.1016/j.psychres.2019.03.001.
- [7] S. W. Yip, D. Scheinost, M. N. Potenza, and K. M. Carroll, "Connectome-based prediction of cocaine abstinence," *Am. J. Psychiatry*, vol. 176, no. 2, pp. 156–164, Feb. 2019, doi: 10.1176/appi.ajp.2018.17101147.
- [8] L. Acion, D. Kelmansky, M. van der Laan, E. Sahker, D. Jones, and S. Arndt, "Use of a machine learning framework to predict substance use disorder treatment success," *PLoS One*, vol. 12, no. 4, p. e0175383, Apr. 2017, doi: 10.1371/journal.pone.0175383.

Visualización de Estudios Preoperatorios en forma de Nube de Puntos Volumétrica en dispositivos de Realidad Mixta

C. Plaza de Miguel¹, J.A. Sánchez-Margallo¹, R. Sánchez Anzules², F.M. Sánchez-Margallo¹

¹ Centro de Cirugía de Mínima Invasión Jesús Usón, Cáceres, España, {cplaza, jasanchez, msanchez}@ccmijesususon.com

² Unidad de Cirugía Torácica, Hospital Universitario de Cáceres, Cáceres, España, quevedoctor@hotmail.com

Resumen

En este trabajo presentamos la primera aplicación para dispositivos de realidad mixta que permite visualizar y navegar en tiempo real, de forma estereoscópica sobre estudios preoperatorios de imagen médica sin necesidad de hacer streaming a través de Internet, hacer uso de estaciones de trabajo o utilizar servicios de pago. De esta forma, al poder visualizar estudios preoperatorios “en crudo” sin tener que segmentar elementos anatómicos para, posteriormente, extraer modelos tridimensionales y visualizarlos en el dispositivo, se agiliza el tiempo requerido de preprocesado para poder mostrar la anatomía del paciente en las gafas HoloLens 2 de Microsoft, lo que permitirá la introducción de este dispositivo en el flujo de trabajo de las unidades de cirugía en los hospitales. Además, mediante esta aplicación se permite al cirujano filtrar la visualización obtenida para mostrar tejidos de distintas densidades, facilitando así la visualización de los elementos que considere más relevantes para la planificación de la intervención. La aplicación se ha utilizado como herramienta de ayuda a la planificación quirúrgica durante la lobectomía superior derecha videoasistida, incluyendo la linfadenectomía sistemática para el carcinoma de células escamosas en el lóbulo superior derecho en el Hospital Universitario de Cáceres. No se registraron complicaciones durante la cirugía. La aplicación permitió compartir el punto de vista del cirujano junto con los hologramas, mostrándolos en una de las pantallas del quirófano al resto del personal médico. Determinados aspectos ergonómicos, como la generación de calor en la frente del cirujano debido al dispositivo, son aspectos a mejorar en siguientes intervenciones.

1. Introducción

La imagen médica actualmente es un campo en rápido crecimiento gracias a los avances técnicos tanto en software (que permiten la renderización de entornos gráficos cada vez más exigentes) como en hardware (a través de la implementación de sucesivas generaciones de

unidades de procesamiento de gráficos -GPU- que están aumentando la capacidad de gráficos cálculos cada año).

La visualización de estudios de imagen preoperatoria, como la tomografía computarizada (TAC), permiten estudiar la anatomía (tanto interna como externa) del paciente, aportando al cirujano información crucial para el planteamiento de la intervención, principalmente en cirugía mínimamente invasiva (CMI), lo que permite una mejor comprensión de las peculiaridades anatómicas de cada paciente de forma personalizada.

La continua evolución de los dispositivos móviles (cada vez más potentes y fáciles de fabricar), junto con la constante evolución en sus capacidades de procesamiento gráfico, han convergido en la aparición de dispositivos de realidad mixta (RM) que ahora permiten fácilmente la visualización de contenido en forma de hologramas de alta resolución. Estos hologramas permiten un reconocimiento detallado del entorno y ofrecen la posibilidad de interactuar a través de los ojos, las manos o los comandos de voz, lo que representa una herramienta valiosa para su uso en entornos quirúrgicos sin afectar a su desempeño.

Específicamente, la segunda generación de gafas HoloLens (Microsoft, Redmond, Washington, Estados Unidos) han abierto un área de investigación amplia y activa basada en la evaluación de pantallas ópticas montadas en la cabeza en el contexto de MIS. Los estudios preoperatorios como los TAC, generalmente almacenados en formato DICOM, contienen un conjunto de datos tridimensional con varias capas que tradicionalmente se muestra en un monitor generando así una imagen 2D de la sección a visualizar. Mediante el teclado o el ratón se puede navegar por el estudio preoperatorio a través de una interfaz de usuario, la cual se muestra en el monitor.

Múltiples intervenciones quirúrgicas han incorporado recientemente gafas de RM como dispositivos de asistencia quirúrgica, permitiendo visualizar los estudios preoperatorios en forma de corte bidimensional no estereoscópico [1-4]. En este trabajo presentamos una variación de este sistema clásico en el que, optimizando la

carga gráfica del dispositivo, permite visualizar las nubes de puntos del estudio de imagen preoperatoria de forma tridimensional y estereoscópica, ofreciendo una visualización más precisa y completa al añadir sensación de profundidad, lo cual aumenta sustancialmente la percepción espacial de la anatomía del paciente. Esto representa una innovación significativa con respecto a los estándares actualmente disponibles en el campo.

2. Materiales y Métodos

Existen en el mercado múltiples soluciones que permiten la visualización eficiente de los estudios de imagen preparatoria, siempre que se disponga de una alta capacidad computacional para manejar la complejidad del procesamiento. En este sentido, cabe destacar los servicios Azure Remote Rendering Cloud (Microsoft Azure), que ofrecen un servicio en la nube para visualizar modelos tridimensionales complejos a través de un sistema de streaming en la nube. Otras soluciones, como la aplicación ISAR SDK (Holo-Light GmbH, Innsbruck, Austria), presentan el inconveniente de ser servicios de alto coste que requieren conexiones a Internet seguras y de alta velocidad para funcionar correctamente [5].

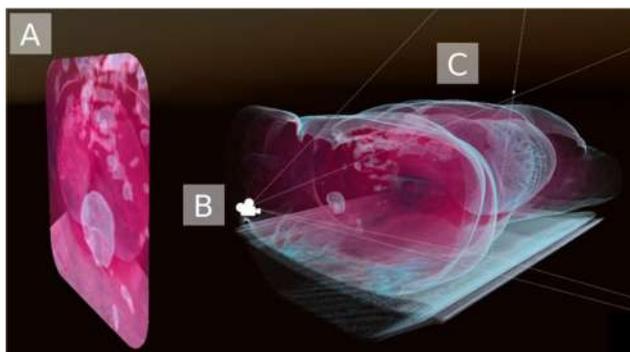


Figura 1. Esquema de la optimización para visualizar estudios de imagen preoperatoria. Plano estereoscópico que visualizará el usuario (A). Cámara interactiva que proyectará la información en el plano estereoscópico que permite navegar por el TAC (B). Nube de puntos volumétrica perteneciente al TAC, la cual no se mostrará al usuario para evitar la caída de frames (C).

En este sentido, las gafas de RM HoloLens 2 son capaces de procesar archivos preoperatorios en formato DICOM y mostrarlos en forma de imagen plana bidimensional. Sin embargo, surgen varios problemas de rendimiento al intentar visualizar nubes de puntos de modelos tridimensionales volumétricos, debido a la gran cantidad de datos que se deben procesar. En este trabajo, se ha desarrollado de forma específica e innovadora un sistema interactivo basado en dos cámaras para superar estos problemas computacionales y permitir la renderización en un plano estereoscópico de las nubes de puntos. En particular, el sistema consta de dos cámaras con perspectiva tridimensional, una de ellas interactiva y la otra fija y unida al sistema de visión del usuario, que renderiza lo que ve la primera cámara (**Figura 1**). De esta

forma, el usuario puede desplazar la cámara para visualizar la anatomía del paciente simplemente arrastrando la cámara mediante gestos intuitivos con las manos, permitiendo la navegación dentro del propio estudio preoperatorio.

Para la implementación de la aplicación se empleó la plataforma de desarrollo software Unity (Unity Technologies, San Francisco, CA, Estados Unidos), la cual permite una fácil interacción con elementos visuales (tanto en 2D como en 3D) e integra diferentes plugins y librerías que facilitan el desarrollo de aplicaciones de realidad mixta. Se adoptó Microsoft Visual Studio como entorno de programación, con C# como lenguaje de programación, para desplegar las aplicaciones en el dispositivo de RM. La aplicación se desarrolló utilizando el kit de desarrollo Mixed Reality Toolkit para Unity. La carga de los estudios de imágenes preoperatorias con el formato Digital Imaging and Communication on Medicine (DICOM) se llevó a cabo mediante una versión personalizada del conjunto de herramientas DICOM de Fellow Oak [6].

La aplicación desarrollada se utilizó como herramienta de ayuda a la planificación quirúrgica durante la lobectomía superior derecha videoasistida, incluyendo la linfadenectomía sistemática para el carcinoma de células escamosas en el lóbulo superior derecho en el Hospital Universitario de Cáceres (España). Al final de la intervención, se le pidió al cirujano indicar su experiencia con el uso del dispositivo y la aplicación desarrollada.

3. Resultados y discusión

El sistema descrito aumentó significativamente la optimización de recursos gráficos de las gafas HoloLens 2, permitiendo la visualización de la nube volumétrica del estudio preoperatorio evitando el denominado “motion sickness” [7] y permitiendo la interacción en tiempo real con el modelo preoperatorio del paciente. En este sentido, presentamos la primera aplicación en tiempo real que permite visualizar y navegar de forma estereoscópica sobre este tipo de imágenes médicas sin necesidad de hacer streaming a través de Internet o cualquier otro servicio de pago. De esta forma, es posible visualizar la nube volumétrica del estudio preoperatorio y navegar por el mismo simplemente haciendo uso del dispositivo.

Adicionalmente, nuestro sistema hace uso de la denominada función de transferencia [8], que permite visualizar la nube de puntos con diferentes preselecciones, asociando un color determinado según el nivel de densidad de cada vóxel (píxel tridimensional perteneciente a la nube de puntos). Esta característica facilita la visualización e identificación de los diferentes sistemas anatómicos del paciente (**Figura 2**).

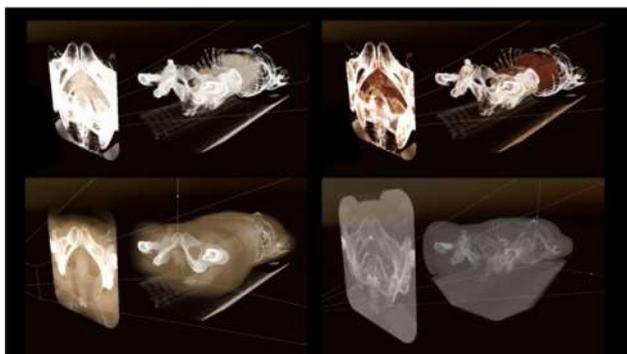


Figura 2. Diferentes visualizaciones preestablecidas para mostrar las nubes de puntos volumétricas asociando distintos colores a la densidad de cada vóxel en particular.

Como se puede ver en la **Figura 3A**, y más en detalle en la **Figura 3B**, partiendo de una nube de puntos preestablecida mostrada en un plano coronal de un paciente el sistema permite visualizar la nube de puntos con densidad de tejidos, facilitando la visualización con mayor detalle la anatomía vascular del sistema respiratorio en la primera fila de imágenes y en la segunda elementos con casi nula densidad, como puede ser el aire que se aloja en los pulmones.

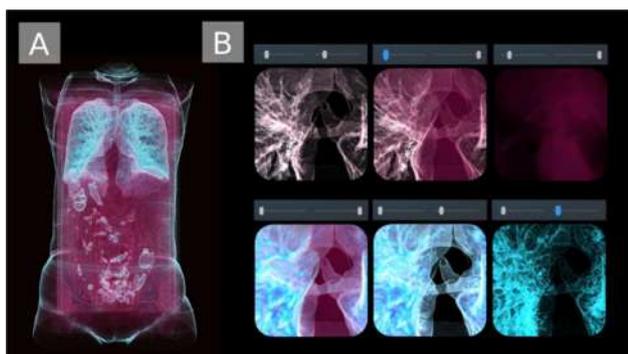


Figura 3. Ejemplo de visualización de la nube de puntos en el eje coronal (A). Modificación de los umbrales que permiten visualizar la misma nube de puntos en el plano axial, mostrando la información relativa a los umbrales de densidad seleccionados por el cirujano (B).

El ajuste predeterminado que se muestra en la **Figura 3** fue utilizado para la planificación quirúrgica de la lobectomía superior derecha. Este preajuste permitió al cirujano visualizar las ramificaciones del pulmón, así como el sistema de arterias y venas pulmonares. Esto le facilitó la sección de la zona afectada por el tumor con mucha más precisión que con los sistemas de planificación quirúrgica tradicionales. Mediante gestos con las manos, el cirujano pudo colocar el modelo sobre el paciente, lo que le permitió identificar la zona afectada en tiempo real, permitiéndole planificar la cirugía de forma personalizada y con mayor exactitud. La cirugía se llevó a cabo sin complicaciones.

4. Conclusiones

Hasta donde llega nuestro conocimiento, esta ha sido la primera vez que se ha conseguido mostrar este tipo de estudios preoperatorios en las gafas de forma estereoscópica durante una cirugía torácica. Como trabajos futuros, se integrarán funciones de transferencia bidimensional en el sistema propuesto, permitiendo un filtrado y personalización más precisos a la hora de visualizar la anatomía del paciente a partir de un estudio de imagen preoperatoria.

Por otro lado, estudiaremos otras posibilidades para mejorar la interacción del usuario con la cámara para la navegación por la nube de puntos, y utilizar la tecnología desarrollada en CMI con más sujetos, gracias a la colaboración del Centro de Cirugía de Mínima Invasión Jesús Usón y el Hospital Universitario de Cáceres.

Agradecimientos

Este estudio ha sido parcialmente financiado por la Junta de Extremadura (España), el Fondo Social Europeo, el Fondo Europeo de Desarrollo Regional (FEDER) "Una forma de hacer Europa", y el Ministerio de Ciencia, Innovación y Universidades de España (TA18023, GR18199, CPI-2019-33-1-TRE -14).

Referencias

- [1] Al Janabi HF, Aydin A., Palaneer S, Macchione N, Al-Jabir A, Khan MS, Dasgupta P, Ahmed K. Effectiveness of the HoloLens mixed-reality headset in minimally invasive surgery: a simulation-based feasibility study. *Surgical Endoscopy* (2019). <https://doi.org/10.1007/s00464-019-06862-3>.
- [2] Perkins SL, Krajancich B, Yang CJ, Hargreaves BA, Daniel BL, Berry MF. A Patient-Specific Mixed- Reality Visualization Tool for Thoracic Surgical Planning, *The Annals of Thoracic Surgery*, (2020) 10(1):290-295, ISS 0003-4975, <https://doi.org/10.1016/j.athoracsur.2020.01.060>.
- [3] Cartucho J, Shapira D, Ashrafian H, Giannarou S. Multimodal mixed reality visualisation for intraoperative surgical guidance. *Int J CARS* (2020) 15:819-826. <https://doi.org/10.1007/s11548-020-02165-4>.
- [4] Li G, Dong J, Wang J, Cao D, Zhang X, Cao Z, Lu G. The clinical application value of mixed-reality-assisted surgical navigation for laparoscopic nephrectomy. *Cancer Medicine* (2020) 9(15):5480- 5489.
- [5] Serhan G, Dimitri P, Jangwoo S, Gurdeep SB, Thomas B, Thomas S, Cornelius H. 2020. Cloud rendering-based volumetric video streaming system for mixed reality services. *Proceedings of the 11th ACM Multimedia Systems Conference*. Association for Computing Machinery, New York, NY, USA, 357-360. DOI:<https://doi.org/10.1145/3339825.3393583>
- [6] Al-Zu'bi S, Al-Ayyoub M, Jararweh Y, Shehab M. Enhanced 3D segmentation techniques for reconstructed 3D medical volumes: Robust and Accurate Intelligent

System. *Procedia Computer Science* (2017) 113. doi:
10.1016/j.procs.2017.08.318

- [7] Alla V, Fridolin W, William G, Kuula T. (2018). Simulator Sickness in Augmented Reality Training Using the Microsoft HoloLens. 1-9. 10.1145/3173574.3173783.
- [8] Kniss J, Kindlmann G, Hansen C. Multi-Dimensional Transfer Functions for Volume Rendering. *Johnson/Hansen: The Visualization Handbook Page Proof* 20. 5.2004 page 181

Evaluación de la Calidad en Registros Electrocardiográficos para Diagnósticos Automáticos Fiables Mediante Aprendizaje Profundo y Aumento de Datos

A. Huerta Herraiz¹, A. Martínez Rodrigo¹, J.J. Rieta Ibáñez², R. Alcaraz Martínez¹

¹ Grupo de Investigación en Ingeniería Electrónica, Biomédica y de Telecomunicación, Universidad de Castilla-La Mancha, Cuenca, España, {alvaro.huerta, arturo.martinez, raul.alcaraz}@uclm.es

² BioMIT.org, Dep. de Ingeniería Electrónica, Universitat Politecnica de Valencia, Valencia, España, jjrieta@upv.es

Resumen

La evaluación de la calidad de registros electrocardiográficos (ECG) es crucial para realizar diagnósticos adecuados de diversas enfermedades cardiovasculares. Actualmente, para realizar dicha tarea es común utilizar algoritmos de aprendizaje profundo, sin embargo, se precisan grandes cantidades de datos. La falta de bases de datos públicas y anotadas requiere buscar nuevos métodos para ampliar el número de muestras disponible. El presente trabajo tiene como objetivo comprobar la utilidad de técnicas de aumento de datos con el fin de evaluar la calidad del ECG. Para ello se han comparado los resultados de clasificación de una red neuronal convolucional (CNN) con dos bases de datos diferentes. Por un lado, se utilizaron 2.000 segmentos de ECG de 5 segundos de una base de datos pública y gratuita, de los cuales la mitad pertenecían a registros de ECG de alta calidad y la otra mitad a ruido. Por otro lado, a partir de técnicas de aumento de datos se generaron 1.000 señales de ECG sintéticas de baja calidad. Estos tramos sintéticos junto a los tramos de alta calidad se utilizaron para crear el segundo conjunto de datos. Los resultados de ambos experimentos se evaluaron mediante un test de McNemar, no encontrándose diferencias significativas entre los mismos, lo que sugiere que las señales ruidosas generadas artificialmente podrían utilizarse con éxito a la hora de entrenar redes CNN con el fin de evaluar la calidad del ECG como paso previo a diagnósticos automáticos.

1. Introducción

Actualmente, los avances en dispositivos portátiles y vestibles así como el desarrollo de la telemedicina han permitido la monitorización electrocardiográfica (ECG) de pacientes a largo plazo [1,2]. Los trastornos cardiovasculares más comunes son muy difíciles de detectar mediante inspecciones rutinarias, debido a que muchos de ellos tan solo duran unos pocos minutos, por lo tanto, se hace imprescindible disponer de sistemas capaces de monitorizar largos periodos de tiempo. Un claro ejemplo es la fibrilación auricular (FA), en la que los episodios iniciales son asintomáticos y duran tan solo unos pocos segundos [3], cuyas probabilidades de éxito de detección dependen intrínseca-

mente del tiempo de monitorización del paciente.

Sin embargo, el hecho de registrar el ECG de un paciente durante el desarrollo de sus tareas cotidianas puede romper la señal en determinados momentos [4]. Existen ruidos inherentes a los sistemas de adquisición incluso en entornos controlados (hospitales o ambulatorios), tales como interferencias de la red eléctrica o línea de base. En sistemas vestibles o portátiles, estas interferencias se acentúan todavía más al ser entornos no controlados, y debido a la actividad cotidiana pueden aparecer artefactos de movimiento, pérdidas de contacto del electrodo con la piel o ruido impulsivo [4]. Otros factores que pueden complicar el análisis de la señal de ECG son el cambio de la intensidad del ruido a lo largo del tiempo, así como las características no estacionarias de dicha señal [4].

El principal problema a la hora de automatizar la toma de decisiones es la presencia de grandes tramos de ruido existentes en la señal de ECG, ya que es inviable procesar de manera manual la gran cantidad de datos adquiridos con dispositivos portátiles y vestibles [4]. Así pues, la evaluación de la calidad del ECG es fundamental para poder aprovechar al máximo el potencial de la monitorización a muy largo plazo y evitando, en la medida de lo posible, diagnósticos automáticos erróneos o una mala interpretación de los segmentos de ECG con ruido [4]. En este sentido, tradicionalmente se han propuesto métodos para evaluar la calidad del ECG, los cuales estaban basados en técnicas comunes de filtrado extrayendo características morfológicas de la señal o puntos fiduciales [5]. Una vez obtenidas las características correspondientes, éstas se utilizaron como entrada de métodos basados en el aprendizaje máquina, tales como máquinas vectores de soporte o ensambles de clasificadores [6]. A pesar de haber obtenido unos resultados muy prometedores, los métodos basados en técnicas de aprendizaje profundo han mejorado los resultados respecto a las técnicas de aprendizaje máquina a la hora de tratar con señales de ECG [7]. Además, estos nuevos métodos también presentan otras características interesantes, como la capacidad de utilizar la señal de ECG en bruto (sin preprocesar), no siendo necesaria interven-

ción manual o externa y evitando la extracción de puntos fiduciales o la detección picos R [7].

Sin embargo, el entrenamiento de un algoritmo de aprendizaje profundo requiere una gran cantidad de datos, y no existen bases de datos de ECG anotadas que sean públicas y gratuitas suficientemente grandes, y las existentes presentan un número muy desigual entre las clases disponibles [8]. Para solventar este problema, actualmente se están proponiendo nuevas técnicas de aumento de datos, tales como el sobremuestreo o la utilización de redes generativas antagónicas (GAN por sus siglas en inglés *Generative Adversarial network*) [9]. En cuanto a su utilización, la técnica de aumento de datos más común es el sobremuestreo, la cual consiste en aplicar transformaciones sobre una señal original, como por ejemplo rotación, reflejo o recortes [9], sin embargo también se pueden aplicar directamente sobre las señales de ECG en bruto, tal y como se realiza en el estudio propuesto. Por lo tanto, el presente trabajo tiene como objetivo validar técnicas de aumento de datos en señales unidimensionales dentro del contexto de la evaluación de la calidad en registros de ECG.

2. Métodos

2.1. Algoritmos para la evaluación de la calidad del ECG

El método empleado para discernir automáticamente entre los intervalos de alta y baja calidad del ECG está basado en un algoritmo previamente publicado, el cual utiliza una red neural convolucional (CNN) preentrenada muy famosa dentro del aprendizaje profundo: AlexNet [7]. Básicamente, la arquitectura de esta red se compone de ocho capas con capacidad de aprendizaje, de las cuales cinco son convolucionales y tres son totalmente conectadas [10]. Además de estas capas, se utilizan funciones lineales de activación, así como capas de agrupación que se utilizan para reducir el tamaño espacial del mapa de características. Con el fin de evitar el sobreentrenamiento también se incluyen dos funciones de regularización después de las dos primeras capas totalmente conectadas. Todos los detalles sobre esta red se encuentran publicados en [10].

Debido a que la red AlexNet está preparada para recibir señales de entrada de dos dimensiones, los segmentos de ECG se convirtieron a imágenes mediante la transformada wavelet continua (TWC) [11]. Estos coeficientes obtenidos mediante la wavelet se representan de forma gráfica junto a un mapa de color que resalta las características de la señal de ECG, generando un escalograma. La representación de la información temporal y de frecuencia de manera simultánea en imágenes 2D ha demostrado generar resultados más robustos en tareas de clasificación. Se pueden encontrar más detalles de esta transformación en [7].

2.2. Desarrollo del experimento

A pesar de que la red AlexNet se encuentra preentrenada, se hace necesario aplicar una operación que se conoce como *ajuste fino* para adaptarla a la nueva tarea de evaluar de calidad del ECG [7]. Para ello, las capas con capacidad

de aprendizaje (es decir, las de convolución y las totalmente conectadas) reajustan el valor de sus pesos y adquieren la habilidad de distinguir entre segmentos de ECG pertenecientes a las clases de alta y baja calidad. Con el fin de evaluar el rendimiento del método de sobremuestreo propuesto en este trabajo, AlexNet se reentrenó de dos maneras diferentes. En primer lugar, se utilizaron 2.000 tramos de ECG de 5 segundos pertenecientes a la base de datos del PhysioNet/CinC Challenge 2017 [12]. Esta base de datos es pública y gratuita y contiene más de 8.000 registros de ECG de una sola derivación, los cuales han sido anotados por expertos en cuatro categorías diferentes: FA, ritmos sinusal normal (RSN), otros ritmos (OR) y ruido. Debido al gran desbalanceo entre las tres primeras clases (registros de alta calidad) y ruido (baja calidad) que presenta la base de datos, tan solo se tomaron 1.000 segmentos de 5 segundos de alta calidad y 1.000 de baja calidad, tal y como se ha descrito en trabajos previos [7]. Es necesario destacar que el grupo de alta calidad se creó manteniendo registros de los diferentes ritmos en proporciones similares, incluyendo 300 segmentos de FA, 400 de RSN y 300 de OR.

En segundo lugar, AlexNet se volvió a entrenar y evaluar con un conjunto de datos generado a partir del anterior. Dado que las anotaciones de ruido en las bases de datos de ECG públicas y gratuitas son menos frecuentes que las de ritmos de alta calidad, se mantuvieron los 1.000 intervalos de ECG de esta última categoría. Para la generación de las nuevas señales se partió de 1.000 segmentos de ECG originales, los cuales se han sometido a cuatro transformaciones diferentes, con un índice de ocurrencia de 0,5. La primera transformación aplicada es la *extensión de tiempo*, mediante esta técnica se modifica ligeramente la duración de la señal de ECG según un factor que varía entre 0,75 y 1,5. Este tipo de transformación se aplica sin afectar al tono. La segunda transformación es el *cambio de tono*, de esta forma, el tono de la señal de ECG se eleva o se atenúa atendiendo a un factor que oscila entre -2 y 2. Debido a que la señal de ECG se está tomando como si fuese un sonido, estos valores se miden en semitonos. La tercera transformación empleada es la *suma de ruido*, en este caso se suma ruido blanco Gaussiano, ajustando la relación señal a ruido para que esté comprendida entre 0 y 3 dB. Por último, la cuarta transformación es la *modificación de amplitud*, en el que se ajusta la ganancia de manera aleatoria comprendida entre un rango entre -5 a 5 dB.

En la Figura 1 se observa un ejemplo de cada transformación aplicada a una señal de ECG perteneciente a la categoría de ruido, clasificada así por la presencia de un artefacto en torno al primer segundo del segmento (a). La primera transformación aplicada es la extensión de tiempo (b), en la que se acorta el número de muestras de la señal. La siguiente transformación (c), cambia en gran medida la morfología y la amplitud después de aplicar el cambio de tono. La suma de ruido blanco Gaussiano distorsiona de manera notable la señal de ECG quedando casi inapreciable su morfología original. Por último, la modificación de amplitud se aplica sin distorsionar la forma de onda (e). Las imágenes generadas mediante la TWC a partir de los segmentos de ECG de ejemplo de la Figura 1 se mues-

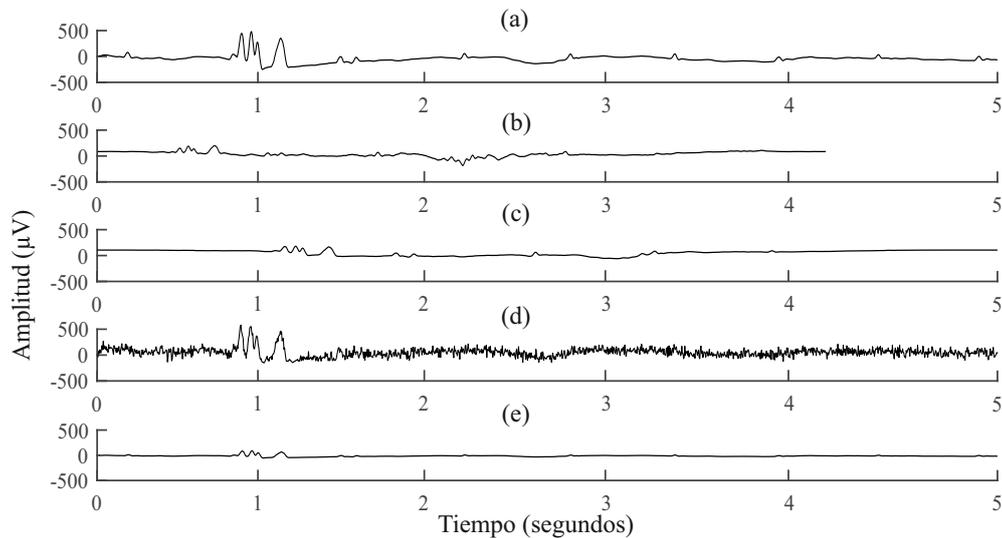


Figura 1. Ejemplo de transformaciones aplicadas para el sobremuestreo tomando a una señal ruidosa como punto de partida (a), extensión de tiempo (b), cambio de tono (c), suma de ruido Gaussiano (d) y modificación de amplitud (e).

tran en la Figura 2. Se observa claramente la correspondencia entre cada segmento ECG y su escalograma correspondiente. En este sentido, la Figura 2(a) muestra la forma en la que el artefacto destaca y los complejos QRS quedan menos definidos, tal y como sucede en la Figura 1(a), en la que se pueden distinguir los picos R y el artefacto. Las figuras 2(b) y (c) muestran cómo la extensión del tiempo y la variación del tono han modificado la morfología de la señal, cambiando de manera muy notable el escalograma. La suma de ruido Gaussiano ha sido la que más altera la imagen original difuminando los picos R, tal y como se muestra en la Figura 2(d). Por último, la Figura 2(e) muestra cómo la modificación de amplitud es la transformación que menos distorsión aplica a la señal original.

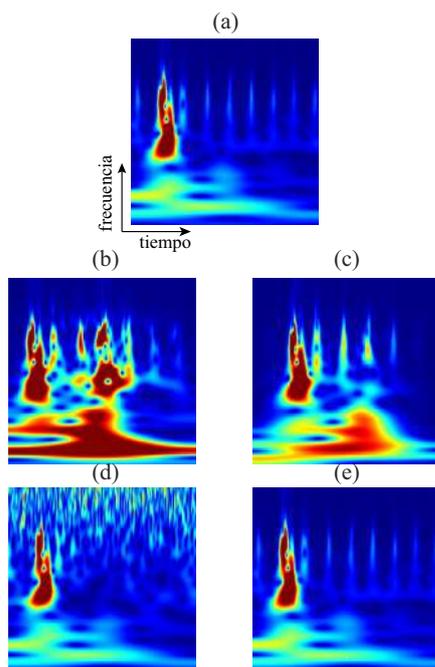


Figura 2. Escalogramas generados a partir de los ECG mostrados en la Figura 1 aplicando la TWC, y que se utilizarán como entrada de la red AlexNet para evaluar su calidad.

2.3. Análisis del rendimiento

El algoritmo se sometió a dos pruebas, en primer lugar utilizando las señales originales y posteriormente las generadas por sobremuestreo. El rendimiento se evaluó mediante un test de McNemar y una validación cruzada de 5 iteraciones. Los resultados estadísticos relativos al clasificador binario se midieron en términos de sensibilidad (Se), especificidad (Es) y exactitud (Ex). Entendiéndose la Se como el porcentaje de segmentos de ECG de alta calidad identificados correctamente, la Es como la tasa de segmentos de baja calidad correctamente clasificados y la Ex como la proporción total de todos los intervalos de ECG bien clasificados. Además, también se midió la tasa de acierto dentro de cada uno de los ritmos disponibles pertenecientes a la clase de alta calidad, tales como RSN (\mathcal{R}_{RSN}), FA (\mathcal{R}_{FA}) y OR (\mathcal{R}_{OR}) para cada uno de los experimentos.

3. Resultados

Como se puede observar en la Tabla 1, en la que se detallan los resultados obtenidos por los dos experimentos desarrollados, no hay diferencias significativas (valor $p > 0.05$) en las métricas de Se , Es y Ex . Del mismo modo, no se obtuvieron grandes diferencias en las tasas de acierto dentro de los diferentes ritmos disponibles, siendo ligeramente superior la tasa de \mathcal{R}_{RSN} que la de \mathcal{R}_{OR} . También se observaron unos resultados cercanos al 90% en todos los experimentos llevados a cabo, con una desviación estándar inferior al 6%.

4. Discusión

En el presente trabajo se ha presentado un novedoso enfoque basado en el aumento de datos utilizando la técnica de sobremuestreo, disponiendo así de más muestras y mejorando el proceso de evaluación de la calidad de los registros de ECG. Los resultados de clasificación obtenidos para las dos bases de datos analizadas sugieren que las transformaciones comunes, tales como la extensión del tiempo, el cambio del tono, la suma de ruido y la modificación de am-

Conjunto	Se	Es	Ex	\mathcal{R}_{RSN}	\mathcal{R}_{FA}	\mathcal{R}_{OR}
Original	0.89±0.02	0.87±0.045	0.88±0.02	0.92±0.03	0.88±0.03	0.87±0.04
Generado	0.90±0.03	0.90±0.04	0.90±0.01	0.91±0.02	0.89±0.06	0.89±0.05

Tabla 1. Resultados de clasificación obtenidos por el clasificador AlexNet en ambos set de datos original y generado.

plitud, que han sido utilizadas para el aumento de datos en otros campos de investigación como por ejemplo el aprendizaje profundo de audio, también podrían ser útiles en este caso. Sin embargo, hay que destacar que el presente estudio tan solo ha evaluado la posibilidad de generar señales de ECG de la clase ruidosa utilizando dichas transformaciones, quedando por estudiar en trabajos futuros el efecto del sobremuestreo para el grupo de alta calidad a la hora de realizar el ajuste fino con algoritmos basados en CNN. De hecho, en este estudio piloto se ha tenido en cuenta que los conjuntos de datos originales y generado compartiesen los mismos 1.000 intervalos de ECG de alta calidad para obtener una comparación equitativa.

Existen otros estudios que han empleado transformaciones diferentes para el aumento de datos, en dichos trabajos se ha visto la utilidad de estas técnicas a la hora de entrenar algoritmos basados en CNN y su utilización en aplicaciones con señales de ECG. En este sentido, Alghamdi et al. [13] desarrollaron un trabajo en el que mejoraban la detección de infartos de miocardio utilizando diferentes arquitecturas CNN, para ello el aumento de datos se realizó por sobremuestreo aplicado a la transformación de imágenes. En otro trabajo de gran interés técnico, Shaker et al. [8] utilizaron una red GAN con la que sintetizaron latidos aumentando así el set de datos disponible para entrenar un clasificador basado en CNN. A pesar de que obtuvieron muy buenos resultados y que las redes GAN se utilizan cada vez más para el aumento de datos en diversas aplicaciones [8], podría no ser tan efectivo a la hora de generar registros de ECG de clase ruidosa. Esto es debido a la gran aleatoriedad que presentan este tipo de señales, así como sus artefactos u otro tipo de distorsiones presentes, dificultando que la red aprenda algún tipo de patrón que luego pueda ser replicado. Esta alternativa de aumento de datos será motivo de trabajos futuros.

5. Conclusiones

Las transformaciones propuestas en el presente trabajo para generar datos con la técnica de sobremuestreo pueden ser muy útiles dentro del contexto del aumento de datos. De esta manera, se pueden generar más señales de ECG ruidosas, mejorando el rendimiento a la hora de evaluar la calidad del ECG con técnicas de aprendizaje profundo, dando lugar a algoritmos más robustos y diagnósticos automáticos más fiables. Sin embargo, para validar los resultados obtenidos es necesario el desarrollo de más estudios con el fin de ampliar las bases de datos disponibles, utilizando para ello técnicas adicionales de aumento de datos.

Agradecimientos

Este trabajo ha sido cofinanciado por los proyectos de investigación competitiva DPI2017-83952-C3 de MINECO-AEI-FEDER-UE, SBPLY/17/180501/000411 de la JCCLM, AICO/2019/036 de la GVA y FEDER 2018/11744.

Referencias

- [1] Sana F, Isselbacher EM, Singh JP, Heist EK, Pathik B, Armondas AA. Wearable devices for ambulatory cardiac monitoring: JACC state-of-the-art review. *Journal of the American College of Cardiology* 2020;75(13):1582–1592.
- [2] Ramasamy S, Balan A. Wearable sensors for ECG measurement: a review. *Sensor Review* 2018;.
- [3] Stachon P, Ahrens I, Faber T, Bode C, Zirlik A. Asymptomatic atrial fibrillation and risk of stroke. *Panminerva Med Dec* 2015;57(4):211–5.
- [4] Smital L, Haider CR, Vitek M, et al. Real-time quality assessment of long-term ECG signals recorded by wearables in free-living conditions. *IEEE Trans Biomed Eng* 2020; 67(10):2721–2734.
- [5] Satija U, Ramkumar B, Manikandan MS. An automated ECG signal quality assessment method for unsupervised diagnostic systems. *Bioc Biomed Eng* 2018;38(1):54–70.
- [6] Behar J, Oster J, Li Q, Clifford GD. ECG signal quality during arrhythmia and its application to false alarm reduction. *IEEE Trans Biomed Eng* 2013;60(6):1660–1666.
- [7] Huerta Herraiz Á, Martínez-Rodrigo A, Bertomeu-González V, Quesada A, Rieta JJ, Alcaraz R. A deep learning approach for featureless robust quality assessment of intermittent atrial fibrillation recordings from portable and wearable devices. *Entropy* 2020;22(7):733.
- [8] Shaker AM, Tantawi M, Shedeed HA, Tolba MF. Generalization of convolutional neural networks for ECG classification using generative adversarial networks. *IEEE Access* 2020;8:35592–35605.
- [9] Hatamian FN, Ravikumar N, Vesal S, Kemeth FP, Struck M, Maier A. The effect of data augmentation on classification of atrial fibrillation in short single-lead ECG signals using deep neural networks. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, may 2020; .
- [10] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In *Adv. in Neural Info. Processing Systems*. 2012; 1097–1105.
- [11] Hong S, Zhou Y, Shang J, Xiao C, Sun J. Opportunities and challenges of deep learning methods for ECG data: A systematic review. *Comput Biol Med* 2020;122:103801.
- [12] Clifford GD, Liu C, Moody B, Li-wei HL, Silva I, Li Q, Johnson A, Mark RG. AF classification from a short single lead ECG recording: The Physionet/computing in cardiology challenge 2017. In *2017 Computing in Cardiology (CinC)*. IEEE, 2017; 1–4.
- [13] Alghamdi A, Hammad M, Ugail H, Abdel-Raheem A, Muhammad K, Khalifa HS, El-Latif AAA. Detection of myocardial infarction based on novel deep transfer learning methods for urban healthcare in smart cities. *arXiv preprint arXiv190609358* 2019;.

Design, manufacture and validation of customized surgical guides for total knee replacement

L. San Martín¹, H. Losada¹, A. Tejo-Otero¹, C.M. Atienza Vicente^{2,3}

¹ Mizar Health SL, Calle Arrirudina 11, 01015, Vitoria-Gasteiz, Spain, { laurasmbdg@gmail.com, h.losada@mizaradditive.com, aitor.tejo@mizarhealth.com }

² Dpto. de Ingeniería Mecánica y de Materiales, Escuela Técnica Superior de Ingeniería Industrial, Universitat Politècnica de València (UPV), Camino de Vera, 46022 Valencia, España, { catienza@ibv.upv.es }

³ Asociación Instituto de Biomecánica de Valencia. Universitat Politècnica de València (UPV), Edificio 9D. Camino de Vera, 46022 Valencia, España, { catienza@ibv.org }

Abstract

This project focuses on the development of customized cutting guides for arthroplasty. Surgical guides are designed and manufactured using AM (Additive Manufacturing) technologies to establish the characteristics of the tibial cut and femoral cut for replacement by a surgical prosthesis. The design process of the guides is detailed, which starts with the reconstruction of the patient's anatomy by a CT (Computed Tomography) scan and is subsequently used for the decision of measures and angulations of the guides, using specific 3D design software. Subsequently, the realization of the personalized surgical planning and the choice of the specific material and method for the surgical guides are explained. Finally, the guides and the knee reconstruction model are 3D printed.

1. Introduction

Orthopedic surgery is a branch of medicine that specializes in the locomotor system; bones, muscles, ligaments, etc. Within this branch is prosthetic surgery; the key surgery that is used for the replacement of a joint when the repair of this by other means is impossible. Arthroplasty is the name of the surgery for the replacement of a joint. The most frequently replaced joint is the knee joint; currently, the number of arthroplasties performed annually is around 65,000 interventions in Spain, according to data collected by the Organization for Economic Cooperation and Development [1]. The conventional surgical procedure for knee arthroplasty consists of replacing the joint with a knee prosthesis by making cuts in the bones of the joint [2]. With the aid of standard instruments, three cuts are made in the femur and one in the tibia where the prosthesis will later be placed. Some of these surgeries present certain difficulties such as the rare anatomy of the patient or vision problems of the doctor, among others. The difficulty of these surgeries can result in more time in the operating room or greater risk for the patient, so the best option would be to avoid these cases.

Manufacturing surgical guides can be an option in order to deal with the typical problems of an operation. These guides ease the surgeon so as to work more freely. In this way, the likelihood of suffering a setback is reduced.

Then, regarding the manufacture of prostheses, the most common method used so far is using certain specific sizes

that surgeons have available for the operation. However, even though these prostheses have impacted in a great manner in patients' lives by improving their quality life, individually-customised 3D-printed prostheses are the future. This is a fact since the personalization in medicine is a reality. Using AM technologies can lead to the development of surgical guides, among other 3D printed parts, since they used the information acquired from medical images, known as DICOM (Digital Imaging Communication on Medicine). According to ISO/ASTM 52900 Standard [3], there are seven categories: binder jetting, direct energy deposition, material extrusion, material jetting, powder bed fusion, sheet lamination and vat photopolymerisation. Amongst these options, each one has its own advantages and disadvantages as well as a range of materials able to be used.

Manufacturing surgical guides has already been studied in different areas of the medical field. For example, Beom et al. [4] 3D-printed surgical guides using the material jetting technology for breast cancer patients who received neoadjuvant chemotherapy. Also, Wang et al. [5] 3D printed surgical guides in resection and reconstruction of malignant bone tumour around the knee joint. The efficacy of the surgical guides was compared with surgeries carried out without any surgical guide. All in all, results showed that there was no significant difference in the operation time between the two groups ($p > 0.05$). Nevertheless, the blood loss, resection length and complication rate were found significantly lower in the group with surgical guides than in the group without surgical guides. This shows the efficacy of the use of surgical guides.

Given the above, the present work studies the customization of arthroplasty surgery for a 77-year-old male patient in need of this intervention. The design and manufacture of personalized surgical guides using 3D printing technologies that fit perfectly to the patient's bone and that, together with standard instruments, indicate to the medical team where to make the cuts that have been previously studied digitally, is carried out.

2. Materials and methods

2.1. Software

The following software has been used for the complete realization of the work. Materialise Mimics is a software created for medical image data processing in the area of engineering and research. It has been used to segment the patient's anatomy from the images obtained by a CT scan and to create a 3D biomodel of the knee joint. Unigraphics NX software, an advanced high-end CAD/CAM program, was used for the creation of references and the subsequent design of customized surgical guides.

CES Edupak, software that allows the study of existing materials and manufacturing processes has been used to determine the material and manufacturing method used according to the specified requirements. Finally, the Magics program prepared the biomodel and surgical guides for immediate fabrication.

2.2. Design and manufacture process

The process of creating customized arthroplasty guides begins with the physician's communication about the need to operate on a patient and ends with the use of the customized surgical guides in surgery (Figure 1).

After the need to operate on the patient, medical images of the anatomy of the knee, ankle, and hip are collected; this process is performed employing a CT scan. These images are processed and segmented for the creation of a 3D biomodel of the knee, ankle and hip, Materialize Mimics software reads DICOM files obtained from the TAC and identifies all the layers of the in three planes different, the axial plane, the sagittal plane and the coronal plane. In these three planes. You can see layer by layer the anatomy of the patient's joint, where it is can distinguish bone from soft tissue by the level of absorption of the radiation. Once at this point, you have to select the area in which you are interested, to create a 3D of this. For the 3D to be correct, only the areas of interest of each layer and plane. In this case, it is intended to recreate the joint of the knee, so the areas that correspond to the bones have to be selected participants of this joint, which tend towards the target since the bone absorbs a large amount of radiation. (Figure 1a)

The next step is the generation of the axes and cutting planes (where the cut will be made in the intervention) that serve as a reference to create the customized guides with the correct size and shape. As it has been possible to assume to obtain these axes, it is necessary both the tibiofibular joint and the femorocoxal joint. For this, it could be done a CT scan of the whole leg, but in order to avoid excessive and unnecessary radiation, a CT scan is obtained only of the knee joint (as seen in the obtaining the biomodel), the ankle joint and the hip joint. As already mentioned in the section corresponding to obtaining the TAC, with all three joints is enough. Next, the process explained above is repeated to obtain of the biomodel of the knee joint with the other two CTs. The areas of interest correspond to the proximal femur and the distal tibia. With

these three biomodels (Figure 1b), the creation of the axes is carried out in which the mechanical axes of the bones are taken as a reference. The mechanical axis of the femur bone corresponds to the straight line drawn from the centre of the femoral head to the intercondylar fossa. For the tibia axes, a line is drawn from the midpoint of the tibial spine to the centre of the tibioperoneoneotragal joint (Figure 1c). The guides are then created in 3D and manufacture (Figure 1d and e).

Finally, the material and manufacturing method are studied and determined, and finally prepared for printing.

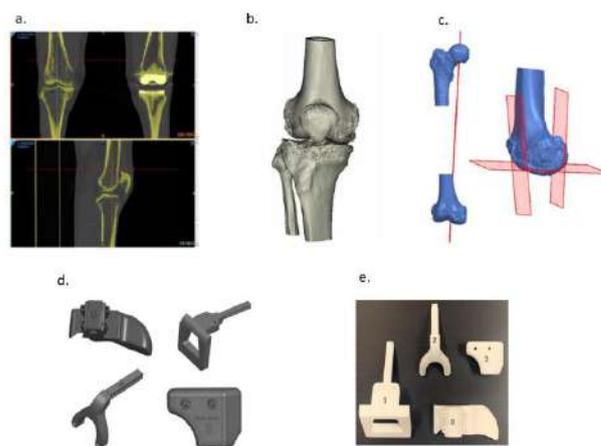


Figure 1. a. CT images processing. b. Creation of biomodels. c. Generation of axes and cutting planes. d. Design of customized surgical guides. e. Manufactured customized surgical guides.

2.3. Design of surgical guides

The customization procedure developed includes the design of four guides called guide 0, guide 1, guide 2 and guide 3.

Guide 0 (Figure 2a) is placed in the anterior cortex of the bone and is perfectly coupled to the bone. It has a rail recess and 4 fixation holes. Guide 1 (Figure 2b) has a rail with the same shape as the hole of guide 0 through which guide 1 is introduced during the operation until it collides with the medial condyle of the femur. It also has some holes that will be used by conventional instruments to perform the femoral cut. Guide 2 (Figure 1c) has a similar shape to guide 1 since it also has a rail that is introduced through guide 0, until it rests on the femoral cut already made and to be able to make the two remaining femoral cuts. Guide 3 (Figure 2d), is perfectly coupled to the tibia and has two holes that will be used to perform the tibial cut using the corresponding instruments.

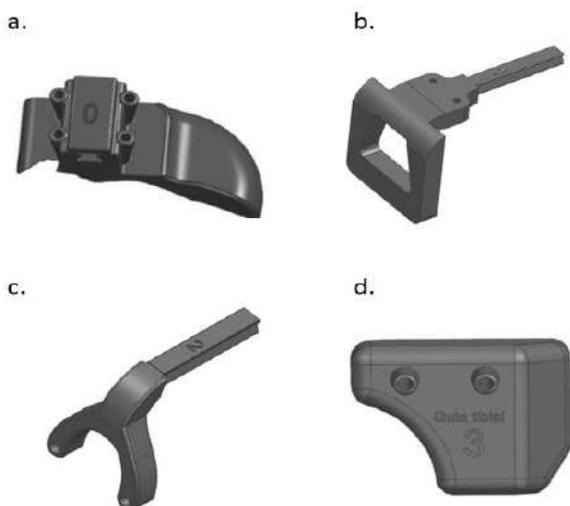


Figure 2. Surgical guides. a. Guide 0. b. Guide 1. c. Guide 2. d. Guide 3.

2.4. Material and manufacture method

Given the complexity and geometric difficulty of the product, the material and manufacturing method must meet certain requirements. The material must be (1) biocompatible, (2) non-toxic, (3) fracture-resistant and (4) sterilizable. The fabrication method must be (1) valid for creating complex 3D geometries and (2) economically cost-effective for the production of customized products.

The study of these requirements and the materials and fabrication methods has been carried out with the CES Edupak software and database, filtering the data bases of materials and manufacturing methods by the given characteristics and requirements.

The study resulted in the use of Polyamide (PA) 12 as the material and selective laser sintering technology (SLS) as the fabrication method.

3. Results

3.1. Biomodel

Figure 3 shows the result of the fabrication of the knee biomodel. This biomodel serves as the starting point for preparing the surgery, obtaining the exact anatomy of the patient before the incision on the leg, to know on beforehand what the doctors are going to find. It is important to highlight that the use of 3D models started many years, but most of the 3D printed biomodels (those with soft tissues) have been manufactured used material jetting technologies.



Figure 3. 3D knee biomodel.

3.2. Customized surgical guides

Figure 4 shows the result of the manufacturing of the customized surgical guides. The surgical guides help the surgeon to accomplish the surgery in a great manner and improve the outcome of the surgery.



Figure 4. Customized surgical guides.

3.3. Customized procedure

The customization of the guidewires results in avoiding some steps of conventional surgery and avoiding risks.

During the operation, after the approach and exposure of the joint bones, guide 0 is placed on the anterior femoral area and fixed with the pins corresponding to this guide (Figure 5a). Guide 1 is then introduced through the hole in guide 0 until guide 1 is resting against the medial condyle (Figure 5b). Holes are drilled in the holes provided in guide 1. After this step, the rail of guide 1 is broken with the aid of a chisel and the pieces are removed. The next step is to make the distal cut. For this, the distal cutting guide, a commercial guide of the corresponding brand that has been studied, is inserted into the holes made with the help of guide 1. Once fixed, the cut is made with a saw, as in conventional surgery, except that the placement of the guide has already been studied. The femur now has the distal cut. The guide rail 2 is then inserted through the corresponding hole in guide 0 (Figure 5c). Guide 2 is placed on the distal section, and holes are drilled again in

the areas indicated for this purpose. Next, the 4:1 cutting guide, also a commercial guide, is placed and the anterior cut and the posterior cut are made based on the location of the holes provided by guide 2. Finally, the tibial cut is performed. To do this, guide 3 is first placed on the tibia (Figure 5d), and following the same steps as for guides 1 and 2, holes are drilled in the prepared locations. Finally, the commercial tibial cutting guide is placed in these holes and the bone is cut with a saw.

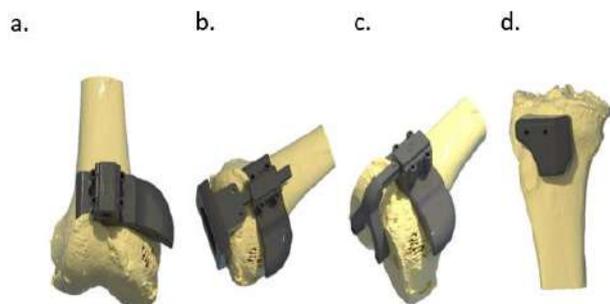


Figure 5. Placement of the surgical guides. a. Placement of guide 0. b. Placement of guide 1. c. Placement of guide 2. d. Placement of guide 3.

The final result after the digital cuts is shown in Figure 6.



Figure 6. The result after cutting the bones.

3.4. Validation

Within a company specialized in 3D printing of implants, a post-validation of the 3D printed parts is done in which several aspects are taken into consideration, mainly by the doctor. Most of the important points are covered with the design carried out within the scope of the present work.

It is true that a clinical investigation in which many and varied opinions about the design and manufacture of the guides would be the correct way to confirm the validity of the surgical guidelines. In this way, the procedure and product would be approved by the medical sector apart from the regulatory agencies for medical devices. However, this research does not validate the entire product, but offers factors whereby it can be improved, as it is a product still under development, so that after several tests and improvements it could become a marketed product

4. Conclusions

The advantages of using this personalized procedure are many, especially when it comes to people's health. With this procedure, alignments and measurements during the operation are avoided, since they are done digitally, which reduces bleeding, minimizes the possibility of error and the operative time; which translates into less risk for the person being operated on. This method also makes it possible to perform these surgeries in patients with anatomical problems or diseases; and allows anticipating the outcome of possible complications to find a solution before they arise, as well as the final result of the surgery.

In terms of design, several details remain to be specified that gain importance in the field of health, where achieving perfect accuracy is paramount. It should also be kept in mind that many of the design aspects are quite subjective and bone-dependent.

In summary, given the advantages, it is clear that it is a procedure that would offer solutions to the most complicated cases and would improve the field of health. It therefore why a very precise procedure must be created and adapted to each patient for the development of this product.

For this reason, a full clinical investigation has to be the next step to follow, in order to complete the validation of the product, making it totally useful and marketable following the European Regulation for this kind of product.

5. References

- [1] «Estadísticas del uso del servicio sanitario en intervenciones: reemplazo total de rodilla.» OCDE (Organisation for Economic Co-operation and Development).
- [2] «Reemplazo total de rodilla (Total Knee Replacement) - OrthoInfo - AAOS».
- [3] Astm, I. (2015). ASTM52900-15 standard terminology for additive manufacturing—general principles—terminology. ASTM International, West Conshohocken, PA, 3(4), 5.
- [4] Ko, B. S., Kim, N., Lee, J. W., Kim, H. J., Chung, I. Y., Kim, J., ... & Ahn, S. H. (2019). MRI-based 3D-printed surgical guides for breast cancer patients who received neoadjuvant chemotherapy. *Scientific reports*, 9(1), 1-6.
- [5] Wang, F., Zhu, J., Peng, X., & Su, J. (2017). The application of 3D printed surgical guides in resection and reconstruction of malignant bone tumor. *Oncology letters*, 14(4), 4581-4584.

Análisis Paso a Paso de Registros de Ablación por Catéter en Fibrilación Auricular Paroxística

A. Vranka¹, F. Hornero Sos², V. Bertomeu-González³,
A. Quesada Dorador⁴, R. Alcaraz Martínez⁵, J.J. Rieta Ibáñez¹

¹ BioMIT.org, Departamento de Ingeniería Electrónica, Universitat Politècnica de Valencia, España, {aivra, jjrieta}@upv.es

² Servicio de Cirugía Cardiovascular, Hospital Clínico Universitario de Valencia, España, hornero_fer@gva.es

³ Departamento de Medicina Clínica, Universidad Miguel Hernández, España, vbortomeu@umh.es

⁴ Servicio de Cardiología, Hospital General Universitario de Valencia, España, quesada_aur@gva.es

⁵ Grupo de Inv. en Electrónica, Telecom. y Bioingeniería, Univ. de Castilla-La Mancha, España, raul.alcaraz@uclm.es

Resumen

Dada la gran importancia de la ablación por catéter (AC) para afrontar la fibrilación auricular (FA), debe priorizarse conocer de forma detallada sus mecanismos durante la AC. Sin embargo, sigue sin conocerse el efecto de pasos fundamentales como la cateterización del seno coronario (SC) y la AC de las venas pulmonares izquierda (ACVPI) y derecha (ACVPD). El presente trabajo intenta indagar en estos asuntos. Se obtuvieron registros bipolares del SC y de la derivación II del electrocardiograma (ECG) durante 5 minutos antes de la AC (A), después de la ACVPI (B) y tras la finalización de la ACVPD (C) de pacientes con FA paroxística. Se calcularon las características de duración, amplitud, área y tasa de pendiente (TP) de las ondas P y las ondas de activación local (OAL). Se calcularon además las características de la variabilidad morfológica, dispersión y variabilidad del ritmo cardíaco (VRC). Se realizó un análisis paralelo de las derivaciones II y SC. Se utilizaron pruebas no paramétricas con corrección de Bonferroni para comparar los resultados entre los pasos de AC (A, B o C). ECG: la duración se redujo después de B (-13%, $p = 0,0001$). SC: no se observaron modificaciones estadísticas. La VRC tendió a aumentar después de B (por el SC: $> 73,12\%$, $p < 0,0480$) y a disminuir después de C (por el SC: $> -33,94\%$, $p < 0,0642$). El análisis de la ACVPI es clave y da como resultado la reducción de la onda P después de la AC. Las alteraciones se observan con mayor fiabilidad a partir del análisis de la onda P, pero los cambios en la VRC se aprecian mejor a partir del estudio del SC. Se recomienda un análisis combinado.

1. Introducción

La ablación por catéter (AC) sigue siendo el tratamiento estrella para afrontar la fibrilación auricular (FA), la arritmia cardíaca más común en el mundo occidental [1,2]. Una técnica ampliamente conocida para la evaluación del sustrato auricular tras la aplicación de AC es el análisis de ondas P, que contiene información sobre la despolarización de la aurícula [3]. Las ondas P de larga o corta duración están relacionadas con remodelación eléctrica, favoreciendo la recurrencia de FA [3-7]. Al mismo tiempo, la disminución de la amplitud o el área de la onda P puede revelar una sesión de AC con resultados exitosos [6,7].

Para el mismo objetivo se utiliza el análisis de la variabilidad del ritmo cardíaco (VRC) [8,9]. Mientras que personas

sanas presentan cierto grado de VRC, un valor bajo o la ausencia total de VRC está relacionada con la aparición de FA [9]. Su importancia en la predicción de recurrencia de FA pone el análisis de VRC en un lugar favorable para su adopción en estrategias postablativas. Sin embargo, los estudios anteriores se enfocan en la obtención de registros antes y después de la AC, obteniendo una perspectiva universal del efecto de la AC, pero ignorando el impacto de cada paso fundamental en el sustrato auricular.

La AC consiste en el aislamiento de las venas pulmonares (VPs), que son la fuente principal de FA. Durante el procedimiento, otras zonas que pueden iniciar o favorecer la perpetuación de FA están detectadas mediante catéteres de cartografía, que están constatemente moviéndose para detectar dichas zonas y catéteres de referencia, que se utilizan para provocar pulsos que pueden iniciar la FA [10]. La cateterización del seno coronario (SC) como estructura de referencia para la cartografía del tejido fibrótico es una opción bastante popular, debido a su posición estratégica entre aurícula derecha e izquierda, que le permite captar disparadores de FA de ambos lados [10,11]. A pesar de su importancia para el mapeo de FA, no hay evidencia que el SC es una referencia adecuada en cuanto a la famosa modificación del sustrato auricular, que se observa después de la aplicación de AC.

El presente estudio tiene como objeto indagar en los asuntos que se han mencionado anteriormente para ofrecer una mayor percepción de los mecanismos de FA paroxística y como se alteran durante cada paso de la AC.

2. Materiales

La base de datos se compuso con registros de derivación II de electrocardiogramas (ECGs) y registros bipolares del SC de 29 pacientes con FA paroxística sometidos por primera vez a AC mediante radiofrecuencia. En primer lugar, se realiza una corona de ablación alrededor de las VPs izquierdas, logrando el aislamiento de las mismas. A continuación, se repite el mismo procedimiento para las VPs

derechas, comprobando también su aislamiento.

Los registros se obtuvieron por cinco minutos antes (fase *A*), durante (fase *B*) y después de la AC (fase *C*) con frecuencia de muestreo de 1 kHz. La fase *B* coincide con el fin de la ablación de las VPs izquierdas, mientras que la fase *C* con el fin de la ablación de las VPs derechas. El catéter de SC consta de 10 electrodos, correspondientes a 5 canales de registros bipolares. Los criterios para elegir el canal analizado para cada paciente fueron la amplitud alta y la extinción de fluctuaciones en la base. Para el mismo paciente en cada fase se utilizó siempre el mismo canal.

3. Métodos

Los ECGs y los registros del SC fueron preprocesados mediante eliminación de ruido y de la media [12, 13]. Los latidos cardíacos ectópicos fueron corregidos mediante interpolación lineal [14]. Los registros del SC fueron adicionalmente sometidos a cancelación de la actividad ventricular [15].

Una vez preprocesadas las señales, las ondas P y las ondas de activación local (OALs) fueron detectadas y delimitadas en los ECGs y en los registros del SC, respectivamente [16, 17]. Para cada onda P y OAL, las características de perfil, es decir las características de duración, amplitud, área y tasa de pendiente a 20 % (TP_{20}) de su duración se computaron. La TP_{20} fue calculada con la siguiente fórmula

$$TP_{20}^n = \frac{A(t_{20}) - A(t_{inicio})}{t_{20} - t_{inicio}}, \quad (1)$$

donde TP_{20}^n es la TP_{20} de la activación n , $A(t_{20})$ y $A(t_{inicio})$ son el amplitud en 20% de la duración de la activación y en el inicio de la misma, respectivamente y t_{20} , t_{inicio} los correspondientes puntos de muestra.

Además de las características de perfil, las siguientes características del registro fueron calculadas [17]: variabilidad morfológica (VM), dispersión a 25% – 75% del rango de duración de cada registro y los índices temporales de la VRC (SDNN, VARNN, RMSSD). Estos índices fueron calculados utilizando las ondas P u OALs.

Dado que muchas de las características de perfil computadas están sujetas a alteraciones provocadas por inestabilidad del ritmo cardíaco, se ha establecido un factor corrector (FC) aplicado a cada activación i de la característica c

$$FC(c_i) = \frac{1000}{IR}, \quad (2)$$

donde IR es el intervalo de dos picos R consecutivos dentro de los cuales se encuentra la activación i .

Cada resultado de duración y área se multiplicó por el FC mientras que cada resultado de TP_{20} se ha dividido por su FC correspondiente. Las características que han pasado por el filtro del FC se van a representar con el símbolo C .

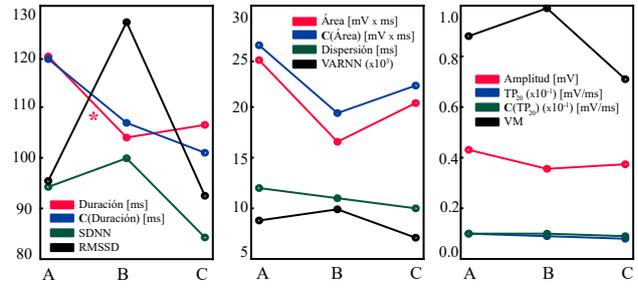


Figura 1. Valores de las características de la onda P en cada fase de análisis. Los valores significativos están marcados con un *. Por la corrección de Bonferroni, es $\alpha = 0,0167$.

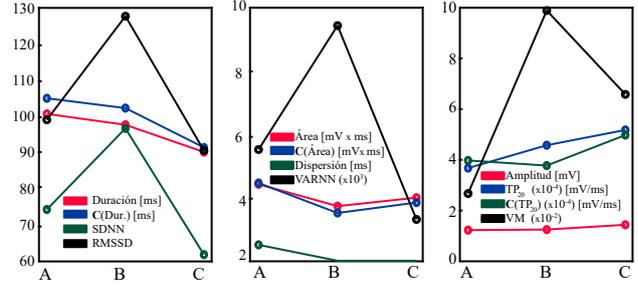


Figura 2. Valores de las características de las OALs en cada fase de análisis. Por la corrección de Bonferroni, es $\alpha = 0,0167$.

3.1. Evaluación del Método

Los valores medianos de todas las características de ECGs y registros de SC a cada fase fueron calculados. Como la prueba de normalidad de Kolmogorov-Smirnov ha indicado distribuciones no normales, la comparación de cada característica entre todas las fases fue realizada mediante el test de Kruskal-Wallis. A continuación se aplicó el test U de Mann-Whitney con corrección de Bonferroni, para comparar el valor de la misma característica cada dos fases. Finalmente, se calculó la variación de las características entre fases de análisis como alteración en porcentaje [%] de su valor inicial. Los antedichos pasos se aplicaron tanto para el análisis de ECGs como para el análisis de registros del SC.

4. Resultados

Las medianas de las características están ilustradas en las figuras 1 y 2 para el análisis de ECGs y registros del SC, respectivamente. Se puede observar en la figura 1 que solo el valor de la duración cambia estadísticamente entre fases *A* y *B* de 120 ms a 104 ms. El mismo valor aumenta en la fase *C* en el caso sin corrección, mientras que cuando se aplica el FC , hay una tendencia decreciente. Aún así, la reducción más prominente sigue siendo observada en la fase *B*. En cuanto a los demás de las características de perfil, siguen también una tendencia decreciente, aunque sin presentar valores estadísticamente significativos.

Los porcentajes de alteración se pueden ver en la tabla 1 y sus valores p en la tabla 2. Cabe destacar que, aunque se reporta un cambio significativo de duración en las ondas P entre las fases *A* y *C* ($-11,01\%$, $p = 0,0094$), eso se tiene que atribuir al cambio que ha provocado la fase *B*. En caso contrario, debería de reportarse un cambio significati-

Tabla 1. Variación provocada por la AC en cada transición de fases consecutivas (A-B, B-C) o en la transición desde el inicio hasta el final de la AC (transición A-C). Los valores significativos están marcados con un *.

Características	A-B [%]		B-C [%]		A-C [%]	
	Ondas P	OALs	Ondas P	OALs	Ondas P	OALs
<i>Duración</i>	-13,30*	-5,49	+0,00	-1,93	-11,01*	-7,46
<i>Amplitud</i>	-16,04	-7,79	+3,80	-2,15	-11,52	-5,09
<i>Área</i>	-25,68	-5,23	+6,50	-1,74	-20,72	-4,06
<i>TP₂₀</i>	+12,84	+25,89	-18,35	+12,20	-7,87	+41,24
C(Duración)	-13,73	-3,89	+1,60	-4,46	-14,49	-9,84
C(Área)	-24,54	-2,90	+6,60	-3,86	-19,65	-19,94
C(TP₂₀)	+3,35	-5,64	-6,44	+32,58	-3,30	+25,10
<i>VM</i>	+15,00	+144,9	-6,11	-5,93	-0,42	+172,1
<i>Dispersión</i>	-22,42	+0,00	+22,22	+80,00	-9,55	+0,00
<i>SDNN</i>	+28,39	+79,80	-24,27	-33,94	+0,28	+0,92
<i>VARNN</i>	+64,86	+225,9	-42,64	-55,91	+0,57	+1,93
<i>RMSSD</i>	+45,15	+73,12	-28,30	-36,30	+0,26	+5,43

Tabla 2. Valores p de la comparación entre fases consecutivos (A-B, B-C), el inicio y el final de la AC (transición A-C) o entre todas las fases (A-B-C). Por la corrección de Bonferroni, el valor α es $\alpha = 0,0167$. Los valores significativos están marcados con un *.

Características	A-B		B-C		A-C		A-B-C	
	Ondas P	OALs						
<i>Duración</i>	0,0012*	0,2411	0,5580	0,2165	0,0094*	0,0553	0,0026*	0,1078
<i>Amplitud</i>	0,0556	0,7397	0,3189	0,8371	0,0967	0,9370	0,0838	0,9416
<i>Área</i>	0,1032	0,6693	0,4383	0,7880	0,1032	0,8371	0,1436	0,8964
<i>TP₂₀</i>	0,3843	0,8868	0,1329	0,6693	0,6925	0,9621	0,1622	0,9322
C(Duración)	0,1412	0,6693	0,7397	0,2001	0,0791	0,0598	0,1588	0,1590
C(Área)	0,0847	0,6693	0,7160	0,9370	0,1101	0,6925	0,1443	0,8821
C(TP₂₀)	0,6925	0,7160	0,5166	0,5798	0,8371	0,8124	0,8011	0,8364
<i>VM</i>	0,1892	0,0176	0,6464	0,6924	0,6239	0,1134	0,4760	0,0595
<i>Dispersión</i>	0,2084	0,4613	0,9493	0,4511	0,1761	0,9234	0,3098	0,6759
<i>SDNN</i>	0,1329	0,0480	0,0598	0,0413	0,8619	0,9118	0,1359	0,0556
<i>VARNN</i>	0,1329	0,0480	0,0598	0,0413	0,8619	0,9118	0,1359	0,0556
<i>RMSSD</i>	0,0517	0,0257	0,0689	0,0642	0,9621	0,7637	0,1359	0,0489*

vo entre fases B y C. Las características de perfil de OALs se reducen ligeramente en su mayoría después de cada fase, sin que estas alteraciones sean significativas ni de valor tampoco de poder estadístico.

Mientras la evolución de VM en los ECGs es mayormente sutil, la VM en registros del SC parece avanzar rápidamente después de la fase B para luego disminuir ligeramente, manteniendo aún valores altos en comparación con los de inicio de la AC. Además, se observan unas tendencias de aumento de valores de VRC entre fases A y B para después observar la disminución de los mismos. Estos resultados son más destacados en el análisis de OALs, con modificaciones más prominentes y poder estadístico más grande.

5. Discusión

La mayoría de los estudios enfocados en el análisis de ondas P se limitan a la observación de las características antes y después de la AC, ignorando la importancia que puede tener la ablación de cada una de las VPs. El presente trabajo se ha comprometido a esclarecer este asunto, ofreciendo un estudio detallado sobre los pasos fundamentales de la

AC. Los resultados demuestran que la ablación de las VPs izquierdas es clave para la disminución de duración, que forma la característica más analizada en cuanto a la evaluación de sustrato auricular después de aplicaciones de AC [3-7]. Esta información puede ser de gran utilidad para futuros estudios, ofreciendo una perspectiva más amplia sobre los mecanismos de la FA y sugiriendo la adopción de estrategias de ablación más enfocadas en el aislamiento de dichas VPs y su análisis postablativo. Las razones por las cuales se observa esta diferencia entre VP izquierdas y derechas quedan desconocidas.

La disminución de VRC es una consecuencia esperada después de la AC, indicando un procedimiento con resultados favorables [8]. Sin embargo, en el presente análisis se reveló un aumento destacable después de la AC de VPs izquierdas. Este fenómeno no es de extrañar, según estudios anteriores que reportan un incremento en la VRC durante la exposición a la energía de radiofrecuencia [18].

En paralelo con el análisis de cada paso de AC, se ha aplicado también el análisis de registros de SC. Debido a su gran popularidad como referencia en los procedimientos

de AC [10], el conocimiento sobre las alteraciones de las características eléctricas del SC es imprescindible para un estudio exhaustivo sobre el impacto de la AC en el sustrato auricular de pacientes de FA paroxística. Aunque las alteraciones observadas desde el análisis de registros del SC están en línea con los resultados del análisis de ondas P reportadas tanto en el presente estudio como en anteriores, dichas modificaciones son de bajo nivel y sin ninguna importancia estadística. De todos modos, cabe destacar que el impacto de la radiofrecuencia durante la AC ha sido manifestado de forma más dominante y clara por el estudio de SC que por el estudio de ECGs. Una posible explicación de esta observación podría ser la proximidad del SC con el tejido directamente expuesto a la energía de radiofrecuencia, que le permite captar cambios de la morfología y la frecuencia de las señales de forma más precisa. En este caso, se sugiere la incorporación del análisis del SC para lograr un estudio más eficaz del sustrato auricular.

6. Conclusiones

El análisis de los resultados de AC en cada paso ofrece un conocimiento más amplio sobre el sustrato auricular y revela fenómenos poco conocidos por los estudios antes-después de la AC. Aunque el análisis del SC en la dirección del análisis de las ondas P no ofrece mayor conocimiento sobre los cambios provocados por la AC, es capaz de captar con mayor precisión alteraciones relacionadas con el efecto de la radiofrecuencia y la modificación de VRC.

Agradecimientos

Trabajo cofinanciado por los proyectos de investigación competitiva DPI2017-83952-C3 de MINECO-AEI-FEDER-UE, SBPLY/17/180501/000411 de la JCCLM y AICO/2021/286 de la GVA.

Referencias

- [1] Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, Boriani G, Castella M, Dan GA, Dilaveris PE, Fauchier L, et al. 2020 ESC guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardiothoracic Surgery (EACTS): The task force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. *European heart journal* February 2021;42:373–498. ISSN 1522-9645.
- [2] Shah D, Haïssaguerre M, Jaïs P. Catheter ablation of pulmonary vein foci for atrial fibrillation. *The Thoracic and Cardiovascular Surgeon* 1999;47:352–356. ISSN 0171-6425.
- [3] Simpson RJ, Foster JR, Gettes LS. Atrial excitability and conduction in patients with interatrial conduction defects. *The American journal of cardiology* December 1982; 50:1331–1337. ISSN 0002-9149.
- [4] Pranata R, Yonas E, Vania R. Prolonged p-wave duration in sinus rhythm pre-ablation is associated with atrial fibrillation recurrence after pulmonary vein isolation—a systematic review and meta-analysis. *Annals of noninvasive electrocardiology the official journal of the International Society for Holter and Noninvasive Electrocardiology Inc* September 2019;24:e12653. ISSN 1542-474X.
- [5] Auricchio A, Özkartal T, Salghetti F, Neumann L, Pezzuto S, Gharaviri A, et al. Short p-wave duration is a marker of higher rate of atrial fibrillation recurrences after pulmonary vein isolation: New insights into the pathophysiological mechanisms through computer simulations. *Journal of the American Heart Association* January 2021;10:e018572. ISSN 2047-9980.
- [6] Van Beeumen K, Houben R, Tavernier R, Ketels S, Duytschaever M. Changes in p-wave area and p-wave duration after circumferential pulmonary vein isolation. *Europace* June 2010;12:798–804. ISSN 1532-2092.
- [7] Maan A, Mansour M, Ruskin JN, Heist EK. Impact of catheter ablation on p-wave parameters on 12-lead electrocardiogram in patients with atrial fibrillation. *Journal of electrocardiology* 2014;47:725–733. ISSN 1532-8430.
- [8] Chen PS, Chen LS, Fishbein MC, Lin SF, Nattel S. Role of the autonomic nervous system in atrial fibrillation: pathophysiology and therapy. *Circulation research* April 2014;114:1500–1515. ISSN 1524-4571.
- [9] Habibi M, Chahal H, Greenland P, Guallar E, Lima JAC, Soliman EZ, Alonso A, Heckbert SR, Nazarian S. Resting heart rate, short-term heart rate variability and incident atrial fibrillation (from the multi-ethnic study of atherosclerosis (MESA)). *The American journal of cardiology* December 2019;124:1684–1689. ISSN 1879-1913.
- [10] Boles U, Gul EE, Enriquez A, Starr N, Haseeb S, Abdollah H, Simpson C, Baranchuk A, Redfearn D, Michael K, et al. Coronary sinus electrograms may predict new-onset atrial fibrillation after typical atrial flutter radiofrequency ablation (cse-af). *Journal of atrial fibrillation* 2018;11(1).
- [11] Tahara M, Kato R, Ikeda Y, Goto K, Asano S, Mori H, Iwanaga S, Muramatsu T, Matsumoto K. Differential atrial pacing to detect reconnection gaps after pulmonary vein isolation in atrial fibrillation. *International heart journal* May 2020;61(3):503–509. ISSN 1349-3299.
- [12] García M, Martínez-Iniesta M, Ródenas J, Rieta JJ, Alcaraz R. A novel wavelet-based filtering strategy to remove powerline interference from electrocardiograms with atrial fibrillation. *Physiological measurement* 2018; 39(11):115006.
- [13] Martínez-Iniesta M, Ródenas J, Rieta JJ, Alcaraz R. The stationary wavelet transform as an efficient reductor of powerline interference for atrial bipolar electrograms in cardiac electrophysiology. *Physiological measurement* 2019; 40(7):075003.
- [14] Martínez A, Alcaraz R, Rieta JJ. Detection and removal of ventricular ectopic beats in atrial fibrillation recordings via principal component analysis. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual International Conference* 2011;2011:4693–4696. ISSN 2694-0604.
- [15] Alcaraz R, Rieta JJ. Adaptive singular value cancellation of ventricular activity in single-lead atrial fibrillation electrocardiograms. *Physiological measurement* 2008; 29(12):1351–69.
- [16] Osorio DI, Alcaraz R, Rieta JJ. A fractionation-based local activation wave detector for atrial electrograms of atrial fibrillation. In *Computing in Cardiology, CinC 2007*, Rennes, France, September 24–27, 2017. 2017; 1–4.
- [17] Vraka A, Bertomeu-González V, Osca J, Ravelli F, Alcaraz R, Rieta JJ. Study on how catheter ablation affects atrial structures in patients with paroxysmal atrial fibrillation: The case of the coronary sinus. In *2020 International Conference on e-Health and Bioengineering (EHB)*. IEEE, 2020; 1–4.
- [18] Misek J, Belyaev I, Jakusova V, Tonhajzerova I, Barabas J, Jakus J. Heart rate variability affected by radiofrequency electromagnetic field in adolescent students. *Bioelectromagnetics* 2018;39(4):277–288.

Evaluación del Riesgo de Hipertensión a partir de Registros Fotopletismográficos utilizando Clasificadores de Deep Learning

Jesús Cano¹, Vicente Bertomeu-González², Lorenzo Fácil³, Roberto Zangróniz⁴, Raúl Alcaraz⁴, José J Rieta^{1*}

¹ BioMIT.org, Electronic Engineering Department, Universitat Politecnica de Valencia, Spain

² Clinical Medicine Department, Miguel Hernández University, Elche Spain

³ Cardiology Department, Hospital General Universitario de Valencia, Spain

⁴ Research Group in Electronic, Biomed. and Telecomm. Eng., Univ. of Castilla-La Mancha, Spain

Resumen

La monitorización periódica de la presión arterial (PA) es esencial para la detección temprana de las enfermedades cardiovasculares causadas por la hipertensión, una condición potencialmente mortal que no presenta síntomas en etapas tempranas. Este trabajo pretende investigar si las técnicas de Deep Learning pueden evaluar los niveles de riesgo de PA utilizando únicamente registros fotopletismográficos (PPG) sin emplear registros electrocardiográficos (ECG), como en muchos estudios anteriores. Se analizaron 15.240 registros con señales PPG y de presión arterial simultáneas de 125 Hz y submuestreadas a 25 Hz. Para la clasificación se emplearon las redes neuronales convolucionales (CNN) preentrenadas GoogLeNet y ResNet con el escalograma de las señales PPG obtenidos mediante la transformada wavelet continua (CWT) como imágenes de entrada para los modelos. La F1 score más alta se alcanzó al discriminar a los pacientes normotensos de los prehipertensos e hipertensos, obteniendo un 92,10% para GoogLeNet y 93,91% para ResNet, respectivamente. Además, la clasificación intra-paciente utilizando diferentes segmentos de un mismo paciente para el entrenamiento y la validación proporcionó una F1 score de 90,28% con GoogLeNet y 89,04% con ResNet. La transformación tiempo frecuencia de las señales PPG como imágenes de entrada de clasificadores de Deep Learning ha sido capaz de proporcionar resultados sobresalientes en la evaluación del riesgo de hipertensión sin requerir señales ECG o la extracción manual de características. Por ello, estas técnicas tienen un gran potencial en dispositivos llevables para la evaluación y monitorización continua y no invasiva del riesgo de hipertensión.

1. Introducción

La presión arterial (PA) es el biomarcador más importante de las enfermedades cardiovasculares, que son la principal causa de muerte en todo el mundo y un importante factor de reducción de la calidad de vida [1]. Por lo tanto, el control periódico de la PA y el diagnóstico precoz de la hipertensión son esenciales para su prevención. Además, la mayoría de los pacientes con presión arterial elevada no presentan signos ni síntomas de daños en los órganos. Por consiguiente, no realizar una detección precoz puede desencadenar una urgencia hipertensiva con la presencia de factores de riesgo de enfermedades como la insuficiencia cardíaca y la insuficiencia renal preexistente o la hipertensión grave no controlada [2].

Las técnicas no invasivas más comunes para la medición de

la PA se basan en la oclusión arterial mediante manguitos inflables, los cuales ofrecen una precisión adecuada pero sólo proporcionan una medición intermitente, es incómoda para la persona a la que se le realiza la medición y deben ser aplicadas por profesionales [3]. Los recientes avances en la tecnología de sensores han desarrollado dispositivos sin manguitos que permiten medir parámetros fisiológicos en cualquier momento. Así, el uso de registros fotopletismográficos (PPG) es muy prometedor por ser no invasivo, proporcionar una medición continua a lo largo del día, de bajo coste, sencillo y con una alta correlación con la PA en el dominio de la frecuencia y el tiempo [4].

En este sentido, muchos estudios han aplicado la tecnología de inteligencia artificial con el fin de estimar o discriminar los niveles de presión arterial. Las técnicas de Machine Learning que combinan las señales electrocardiográficas (ECG) y de PPG utilizan la teoría de la propagación de la onda de presión producida en el corazón con parámetros como el tiempo de tránsito del pulso (PTT), el tiempo de llegada del pulso (PAT) y la velocidad de la onda del pulso (PWV) para determinar el estado cardiovascular [5]. Estudios recientes combinan estos parámetros de propagación de la onda con la extracción de características morfológicas de la señal PPG como entradas para los modelos de clasificación [6]. En los últimos años, la técnica de Deep Learning con potentes métodos computacionales que eliminan la extracción de características de forma manual han mostrado una mejora en la estimación de la PA a partir de señales PPG [7].

El presente trabajo propone un método para la evaluación del riesgo de hipertensión utilizando las redes neuronales convolucionales (CNN) preentrenadas GoogLeNet y ResNet y el escalograma de las señales PPG a través de la transformada wavelet continua (CWT) como imágenes de entrada de estos modelos sin la necesidad de utilizar registros de ECG o la extracción manual de características de las señales para su clasificación.

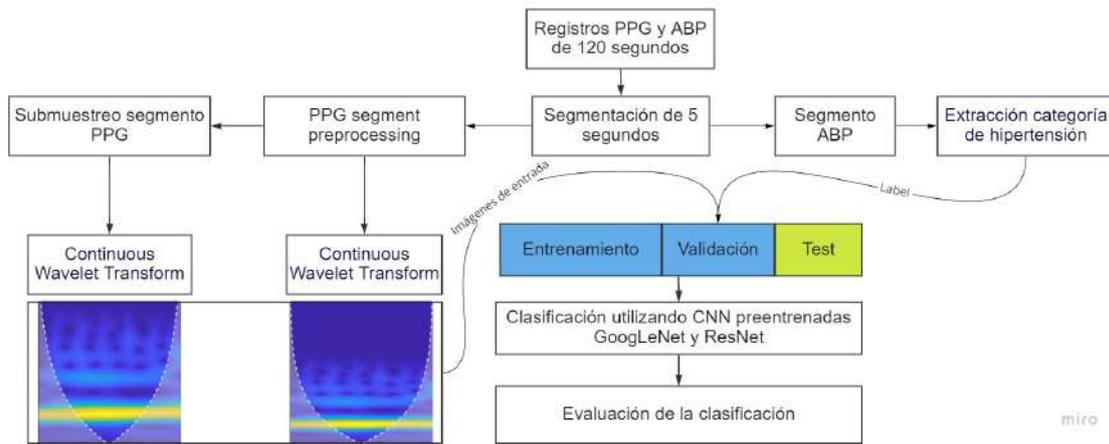


Figura 1. Proceso de clasificación Deep Learning

2. Materiales y Métodos

2.1. Adquisición de Señales y Preprocesado

Se obtuvieron un total de 635 registros de 50 pacientes diferentes de la base de datos gratuita MIMIC-III ('Medical Information Mart for Intensive Care'), que contiene información y señales vitales de pacientes ingresados en la unidad de cuidados intensivos [8]. Como las señales son obtenidas por dispositivos comerciales, a menudo contienen artefactos causados por los movimientos del sensor o la pérdida de contacto con la zona de medida. Los principales artefactos detectados y la razón por la que muchos registros no se incorporaron al conjunto de datos finales fueron la presencia de señales PPG y de presión arterial (ABP) ruidosas, falta de picos, valores de PA improbables o señales continuas que no representan la morfología característica de las señales.

En este estudio se emplearon señales ABP y PPG simultáneas y estables con una duración de 120 segundos y una frecuencia de muestreo de 125 Hz. La presión arterial sistólica (PAS) se extrajo de las señales ABP, a las que no se realizó ningún preprocesado, y se utilizó para etiquetar las señales PPG en normotensas (NT), prehipertensas (PH) e hipertensas (HT), con valores de PAS de 120 mmHg o inferiores, entre 120 y 140 mmHg y 140 mmHg o superiores, respectivamente, según la definición de los Institutos Nacionales de Salud de EE.UU. [1]. Además, se aplicó un filtro paso banda Chebyshev II de cuarto orden de 0,5-10 Hz a las señales PPG normalizadas para eliminar el ruido y mejorar la calidad de las señales [9]. Posteriormente, ambas señales de 120 segundos se dividieron en segmentos de 5 segundos de longitud, siendo 15.240 el número total de segmentos analizados.

Además, se estudió el efecto de disminuir la frecuencia de muestreo de la señal PPG a 25 Hz para poder así facilitar la introducción de este método en dispositivos llevables como pueden ser los smartwatches. De esta forma, el uso de frecuencias de muestreo más bajas reducirá el consumo de energía del dispositivo y la cantidad de datos guardados y transmitidos [10].

2.2. CNN Preentrenadas GoogLeNet y ResNet

Se evaluaron dos CNN diferentes para la clasificación del riesgo de hipertensión, GoogLeNet [11] y ResNet50 [12]. El aprendizaje de transferencia funciona con redes preentrenadas que ya han utilizado el conjunto de datos ImageNet [13] con más de un millón de imágenes y 1000 clases diferentes para aprender a extraer características profundas de las imágenes. Son un punto de partida para los problemas de identificación y clasificación de una forma mucho más rápida que el entrenamiento desde cero, utilizando de esta forma una menor cantidad de imágenes para el entrenamiento. Para ello, la última capa de aprendizaje y la capa final de clasificación se sustituyen por nuevas capas ajustadas a las nuevas imágenes de entrenamiento.

Las opciones de entrenamiento se establecieron con un tamaño de minimum Batch de 128, frecuencia de validación modificada en función del número de imágenes de entrenamiento y la época máxima fue de 25. Para minimizar el efecto del sobreentrenamiento, la técnica de early stopping detiene el entrenamiento automáticamente cuando la pérdida en el conjunto de validación comienza a aumentar.

Debido a que ambas redes neuronales utilizan imágenes RGB de entrada de tamaño $224 \times 224 \times 3$, los segmentos de PPG fueron procesados por la transformada wavelet continua (CWT) y transformados a un escalograma, una representación de la frecuencia a lo largo del tiempo de la señal, para posteriormente ser redimensionados a $224 \times 224 \times 3$ y alimentar los modelos de entrenamiento. Se ha utilizado la analytic Morse (3,60) wavelet en el `cwtfilerbank` de Matlab, ajustando el `VoicesperOctave` a 12 para crear la CWT. En el escalograma se incluye el cono de influencia, que representa las zonas dónde se producen los efectos de borde en la CWT.

2.3. Evaluación del riesgo de hipertensión

Con el fin de clasificar el riesgo de hipertensión utilizando señales PPG, se compararon los tres grupos de señales etiquetadas discriminando NT frente a PH + HT y NT + PH frente a HT. A continuación, se desarrollaron dos métodos diferentes, empezando por la división del conjunto de datos de 635 registros en un 80 % para el entrenamiento

y la validación y un 20 % para test con registros de nuevos pacientes. Además, el primer grupo se vuelve a dividir aleatoriamente en un 80 % para el entrenamiento y un 20 % para la validación. Hay que tener en cuenta que en este enfoque la división se hace a las imágenes, no a los registros, por lo que las imágenes obtenidas de segmentos de 5 segundos de los mismos registros están presentes en los conjuntos de datos de entrenamiento y validación. Además, se desarrolló una clasificación intra paciente utilizando diferentes segmentos de señal PPG completos de 120 segundos de un mismo paciente en diferentes momentos del día para el entrenamiento y la validación.

Por último, la evaluación de los resultados se realizó mediante los estadísticos de sensibilidad (Se), especificidad (Sp) y F1 score, calculada a partir del recall y la precisión.

3. Resultados

Las Tablas 1 y 2 muestran la eficacia de la clasificación cuando el conjunto de datos se divide aleatoriamente en entrenamiento y validación con la frecuencia de muestreo original de 125 Hz y reducida a 25 Hz respectivamente.

		Se	Sp	F1score
GoogLeNet	NTvsPH+HT	95.44 %	83.08 %	92.10 %
	NT+PHvsHT	84.45 %	90.99 %	84.16 %
ResNet	NTvsPH+HT	94.37 %	90.52 %	93.91 %
	NT+PHvsHT	87.67 %	93.87 %	88.23 %

Tabla 1. Rendimiento de entrenamiento-validación a partir de señales PPG con una frecuencia de muestreo de 125 Hz

		Se	Sp	F1score
GoogLeNet	NTvsPH+HT	94 %	83.83 %	91.54 %
	NT+PHvsHT	82.49 %	91.87 %	83.64 %
ResNet	NTvsPH+HT	94.56 %	89.29 %	93.58 %
	NT+PHvsHT	88.94 %	92.83 %	88.08 %

Tabla 2. Rendimiento de entrenamiento-validación a partir de señales PPG con una frecuencia de muestreo de 25 Hz

Todos los modelos siguieron el mismo patrón, en el que el porcentaje de F1 score es mayor comparando los segmentos NT frente a los PH + HT, siendo superior al 91,5 %, que comparando los segmentos NT + PH frente a los HT, siendo inferior al 88,5 %.

Las Tablas 3 y 4 muestran el rendimiento de test con nuevos pacientes cuando el conjunto de datos se divide en un 80 % para entrenamiento y validación y 20 % para test, con la frecuencia de muestreo original de 125 Hz y reducida a 25 Hz, respectivamente.

		Se	Sp	F1score
GoogLeNet	NT vs PH+HT	68.13 %	48.03 %	65.95 %
	NT+PH vs HT	9.05 %	87.67 %	10.07 %
ResNet	NT vs PH+HT	69.67 %	49.58 %	67.31 %
	NT+PH vs HT	16.11 %	83.70 %	15.38 %

Tabla 3. Rendimiento de Test a partir de señales de PPG con una frecuencia de muestreo de 125 Hz

Todos los modelos que comparan NT frente a PH+HT obtienen una F1 score similar, en torno al 65 %. Sin embargo,

		Se	Sp	F1score
GoogLeNet	NT vs PH+HT	59.70 %	51.16 %	60.97 %
	NT+PH vs HT	4.61 %	89.47 %	5.60 %
ResNet	NT vs PH+HT	67.75 %	45.68 %	65.17 %
	NT+PH vs HT	20.18 %	82.33 %	18.29 %

Tabla 4. Rendimiento de Test a partir de señales de PPG con una frecuencia de muestreo de 25 Hz

los modelos que comparan NT+PH vs HT obtienen resultados muy bajos al clasificar a casi todos los nuevos registros como negativos, con una F1 score de en torno al 10 %.

Tras obtener estos resultados, se desarrolló el estudio intrapaciente sólo con la primera distribución de riesgos de hipertensión, tal y como se representa en la Tabla 5. El mayor porcentaje de F1 score, aproximadamente el 90 %, se obtiene en los modelos que utilizan representaciones de señales submuestreadas como entradas a los modelos.

		Se	Sp	F1score
GoogLeNet	125 Hz	88.89 %	78.13 %	87.37 %
	25 Hz	90.28 %	85.42 %	90.28 %
ResNet	125 Hz	83.83 %	83.83 %	85.71 %
	25 Hz	93.06 %	76.04 %	89.04 %

Tabla 5. Rendimiento de clasificación intrapaciente a partir de las señales de PPG comparando segmentos NT con PH + HT

4. Discusión

Poder monitorizar y detectar la hipertensión con una medición continua es de gran importancia, ya que es el principal factor de riesgo de muchas enfermedades cardiovasculares. Por este motivo, se han propuesto nuevos dispositivos alternativos que no utilizan las tradicionales técnicas de medición con manguito, basándose en modelos de inteligencia artificial como Machine Learning y Deep Learning. La mayoría de estudios utilizan la señal PPG, ya que la variación en su morfología está causada principalmente por la actividad del corazón y el estado de las paredes vasculares. Además, las señales PPG son sencillas de obtener con dispositivos no invasivos, tienen un bajo coste y pueden medirse en tiempo real.

Este trabajo trata de investigar si las nuevas técnicas de Deep Learning son capaces de imponerse a la técnica más utilizada en la actualidad para obtener el riesgo de hipertensión, los clasificadores de Machine Learning. Estos clasificadores tienen algunas dificultades, ya que necesitan extraer de forma manual las características morfológicas de las señales, que tendrán diferentes propiedades y su calidad puede no ser la óptima. Además, en este tipo particular de clasificación, es necesario extraer otra señal biomédica, el ECG, para obtener uno de los parámetros discriminativos más utilizados, el PAT. El enfoque de los modelos de Deep Learning son una buena solución para evitar estas dificultades, ya que extraen las características de la imagen obtenida a partir de señales PPG (sin necesidad de la señal ECG) de forma automática y robusta.

A partir de los resultados obtenidos se ha comprobado que la F1 score para los modelos preentrenados GoogLeNet y

ResNet y utilizando tanto la señal PPG original con una frecuencia de muestreo de 125 Hz como la submuestreada a 25 Hz ha sido de alrededor del 93 %. Estos buenos resultados, similares en ambas frecuencias de muestreo, han hecho que se pueda proponer utilizar señales submuestreadas para la potencial aplicación de esta técnica en un dispositivo llevable, ya que reduciría la complejidad computacional y el uso de memoria.

Cabe destacar que en todos los modelos, los mejores resultados se obtienen comparando a los pacientes de NT con los de PH y HT. Esto es muy relevante, ya que el mejor rendimiento de la clasificación se obtiene cuando los pacientes PH son identificados como enfermos, por lo que las características profundas extraídas en los pacientes PH son más similares a las de los pacientes HT que a las de los NT. Además, hay que tener en cuenta que los pacientes PH no muestran síntomas graves hasta que se encuentran en fases muy avanzadas de la enfermedad, causando graves problemas cardiovasculares, por lo que alertar a este grupo como pacientes enfermos tiene un gran interés.

Además, los resultados de test mostraron que la clasificación de los nuevos pacientes no se realiza correctamente, con una F1 score de en torno al 65 %. Esto podría deberse a que cada sujeto tiene una dinámica cardiovascular única, por lo que no todos tienen la misma relación entre las señales de PPG y la PA. Además, las mismas formas de pulso de la señal PPG no garantizan el mismo nivel de PA, por lo que la información de cada sujeto debe estar en los conjuntos de entrenamiento, validación y test [14].

Por último, se ha estudiado la clasificación intra paciente, donde se han introducido segmentos de señal PPG del mismo sujeto de diferentes horas en los conjuntos de entrenamiento y de validación, buscando así evitar el posible sobreentrenamiento del anterior método en los que se hizo una división aleatoria de todas las imágenes. Con el modelo GoogLeNet y las señales PPG submuestreadas se ha obtenido el mayor valor de F1 score, 90,28 %, lo que demuestra la importancia de la calibración a la hora de introducir nuevos sujetos para reducir así el error en la clasificación.

5. Conclusiones

En este trabajo se han evaluado modelos de clasificación del riesgo de hipertensión a partir de registros de PPG utilizando clasificadores de Deep Learning. La combinación de la CWT y los modelos CNN preentrenados ha demostrado un rendimiento y un potencial excepcionales, ya que extrae las principales características de la representación de imágenes PPG de forma automática y no necesita señales de ECG para la extracción de características como los métodos clásicos basados en Machine Learning, con una F1 score del 93,91 %. Además, se ha demostrado la viabilidad del uso de señales PPG submuestreadas y el planteamiento intrapaciente para la implementación de estos modelos en dispositivos llevables.

Agradecimientos

El presente trabajo ha sido cofinanciado por los proyectos de investigación competitiva DPI2017-83952-C3 de

MINECO-AEI-FEDER-UE, SBPLY/17/180501/000411 de la JCCLM y AICO/2019/036 de la GVA.

Referencias

- [1] Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL, Jones DW, Materson BJ, Oparil S, Wright JT, Roccella EJ. Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Hypertension* dec 2003;42(6):1206–1252. ISSN 0194-911X.
- [2] Kessler CS, Joudeh Y. Evaluation and Treatment of Severe Asymptomatic Hypertension. *American Family Physician* feb 2010;81(4):470–476. ISSN 0002-838X.
- [3] Pielmus AG, Mühlstef J, Bresch E, Glos M, Jungen C, Mieke S, Orglmeister R, Schulze A, Stender B, Voigt V, Zauneder S. Surrogate based continuous noninvasive blood pressure measurement. *Biomedizinische Technik* feb 2021; ISSN 00135585.
- [4] Martínez G, Howard N, Abbott D, Lim K, Ward R, Elgendi M. Can Photoplethysmography Replace Arterial Blood Pressure in the Assessment of Blood Pressure? *Journal of Clinical Medicine* sep 2018;7(10):316. ISSN 2077-0383.
- [5] Yoon YZ, Kang JM, Kwon Y, Park S, Noh S, Kim Y, Park J, Hwang SW. Cuff-Less Blood Pressure Estimation Using Pulse Waveform Analysis and Pulse Arrival Time. *IEEE Journal of Biomedical and Health Informatics* jul 2018; 22(4):1068–1074. ISSN 2168-2194.
- [6] Kachuee M, Kiani MM, Mohammadzade H, Shabany M. Cuffless Blood Pressure Estimation Algorithms for Continuous Health-Care Monitoring. *IEEE Transactions on Biomedical Engineering* apr 2017;64(4):859–869. ISSN 0018-9294.
- [7] Liang Y, Chen Z, Ward R, Elgendi M. Photoplethysmography and deep learning: Enhancing hypertension risk stratification. *Biosensors* oct 2018;8(4). ISSN 20796374.
- [8] Johnson AE, Pollard TJ, Shen L, Lehman LWH, Feng M, Ghassemi M, Moody B, Szolovits P, Anthony Celi L, Mark RG. MIMIC-III, a freely accessible critical care database. *Scientific Data* may 2016;3(1):1–9. ISSN 20524463.
- [9] Liang Y, Elgendi M, Chen Z, Ward R. Analysis: An optimal filter for short photoplethysmogram signals. *Scientific Data* may 2018;5(1):1–12. ISSN 20524463.
- [10] Fujita D, Suzuki A. Evaluation of the Possible Use of PPG Waveform Features Measured at Low Sampling Rate. *IEEE Access* 2019;7:58361–58367. ISSN 21693536.
- [11] Szegedy C, Wei Liu, Yangqing Jia, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), volume 07-12-June. IEEE. ISBN 978-1-4673-6964-0. ISSN 10636919, jun 2015; 1–9.
- [12] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), volume 2016-Decem. IEEE. ISBN 978-1-4673-8851-1. ISSN 10636919, jun 2016; 770–778.
- [13] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision* dec 2015; 115(3):211–252. ISSN 0920-5691.
- [14] Esmaelpoor J, Moradi MH, Kadkhodamohammadi A. A multistage deep neural network model for blood pressure estimation using photoplethysmogram signals. *Computers in Biology and Medicine* may 2020;120:103719. ISSN 00104825.

Diferencias Entre Patrones de Ondas P y Ondas de Activación Local del Seno Coronario Durante Ablación por Catéter de Fibrilación Auricular Paroxística

A. Vraha¹, V. Bertomeu-González², F. Hornero Sos³,
A. Quesada Dorador⁴, R. Alcaraz Martínez⁵, J.J. Rieta Ibáñez¹

¹ BioMIT.org, Departamento de Ingeniería Electrónica, Universitat Politècnica de Valencia, España, {aivra, jjrieta}@upv.es

² Departamento de Medicina Clínica, Universidad Miguel Hernández, España, vbortomeu@umh.es

³ Servicio de Cirugía Cardiovascular, Hospital Clínico Universitario de Valencia, España, hornero_fer@gva.es

⁴ Servicio de Cardiología, Hospital General Universitario de Valencia, España, quesada_aur@gva.es

⁵ Grupo de Inv. en Electrónica, Telecom. y Bioingeniería, Univ. de Castilla-La Mancha, España, raul.alcaraz@uclm.es

Resumen

Aunque el seno coronario (SC) es una referencia ampliamente utilizada en la ablación por catéter (AC) de fibrilación auricular (FA), los estudios evalúan el sustrato auricular analizando solamente ondas P. Saber si el SC sigue patrones similares a los de las ondas P ofrece una perspectiva más detallada sobre el procedimiento de la AC. Se utilizaron registros bipolares del SC y de derivación II de 29 pacientes con FA paroxística que se obtuvieron durante 5 minutos antes de la AC (A), después de la AC de las venas pulmonares (VPs) izquierdas (B) y después de la AC de las VPs derechas (C). Se calculó la duración, la amplitud y la variabilidad (VRC) del ritmo cardíaco (RC), entre otras características. Se realizó una corrección de las fluctuaciones del RC. Las correlaciones entre ondas P y activaciones del SC se calcularon mediante correlación de Pearson (CP), regresión lineal (RL) con validación cruzada de 10 pliegues y entropía de muestra cuadrática cruzada (EMCC). Las correlaciones del efecto de la AC se calcularon mediante CP. Tras la corrección, la correlación en la duración aumentó tanto estadísticamente como en magnitud para CP (activaciones: > 43,82%, $p < 0,0001$, efecto de AC: > 53,90%, $p < 0,0210$) y RL (> 18,97%, $p < 0,0001$). La EMCC mostró valores entre 0,6 y 1,3, indicando relaciones insignificantes. El efecto de la AC mostró alta sintonía en el análisis de la VRC (< +94,20%, $p < 0,0297$) y moderada/alta concordancia tras la corrección (< +85,72%, $p < 0,0210$). Las similitudes entre las activaciones del SC y las ondas P son leves. El efecto de la AC también es diferente. Para los estudios de correlación, la corrección de las fluctuaciones del RC puede potenciar los resultados.

1. Introducción

La alta prevalencia de fibrilación auricular (FA) y su impacto significativo en la salud y la economía destacan la necesidad de un tratamiento con alto rendimiento [1]. Actualmente, la ablación por catéter (AC) es la estrategia más conocida de afrontar esta enfermedad, presentando alto rendimiento para pacientes de FA paroxística [2]. Sin embargo, la fibrosis y la remodelación eléctrica en los pacientes de FA persistente hacen que no presenten tasas igualmente satisfactorias [2]. La acción clave para diseñar la estrategia más adecuada es la observación constante del sustrato au-

ricular, que se realiza principalmente a través del análisis de ondas P [3–5].

La duración de ondas P es la característica más utilizada para lograr este objetivo, ya que puede revelar información sobre la remodelación eléctrica [3, 5]. Otras características de ondas P que son extensamente estudiadas son su amplitud y área y una tendencia descendiente está considerada favorable para el resultado de la AC [3]. El análisis de ondas P contribuye de manera incuestionable en el conocimiento de los mecanismos de la FA y su interacción con la AC. Sin embargo, su estudio corresponde a la alteración masiva de las aurículas. Cómo se alteran las estructuras auriculares en función del procedimiento de la AC sigue siendo un misterio.

El problema principal de analizar ciertas estructuras auriculares por separado es la falta de registros estables que permitan observar dichas estructuras por cierto tiempo. Una estructura mayormente utilizada como referencia en sesiones de AC es el seno coronario (SC), que se ubica entre aurícula derecha e izquierda [6]. Teniendo en cuenta su importancia en el procedimiento de la AC y la disponibilidad de registros estables a lo largo del procedimiento, utilizar estos registros para estudiar la modificación de la función del SC permitirá un conocimiento más amplio sobre las estructuras críticas de la AC. Al mismo tiempo, la eficacia del estudio del sustrato auricular mediante ondas P indica la adopción de una estrategia similar para analizar las ondas de activación local (OALs) observadas en los registros del SC y la comparación entre OALs y ondas P para definir el grado de alteración del SC en comparación con la alteración observada por las ondas P.

2. Materiales

Se utilizaron registros de derivación II de electrocardiogramas (ECGs) y registros bipolares del SC con frecuencia de muestreo de 1 kHz de 29 pacientes con FA paroxística

Tabla 1. Correlaciones CP y RL para la comparación de cada activación. Los resultados estadísticamente significativos se muestran en color gris. A: antes; B: ACVPI; C: ACVPD.

Características	A			B			C		
	CP [%]	RL [%]	valor <i>p</i>	CP [%]	RL [%]	valor <i>p</i>	CP [%]	RL [%]	valor <i>p</i>
Duración	-2,19	0,56	0,1291	-3,08	0,69	0,1040	4,53	0,22	0,2221
Amplitud	-1,98	2,12	0,0090	15,19	2,34	0,0033	7,26	3,49	0,0016
Área	-3,30	2,94	0,0032	16,18	3,24	0,0015	0,99	1,40	0,0142
TP ₂₀	0,61	2,38	0,0839	0,97	0,09	0,3927	-0,64	0,38	0,1426
C(Duración)	45,25	20,32	< 0,0001	69,91	47,25	< 0,0001	43,82	18,97	< 0,0001
C(Área)	10,00	1,90	0,0142	35,40	12,22	< 0,0001	10,66	1,03	0,0425
C(TP ₂₀)	-2,78	0,47	0,1431	0,60	0,78	0,0843	-0,44	0,47	0,1187

sometidos por primera vez a AC mediante radiofrecuencia (RF). Los registros se obtuvieron durante 5 minutos en tres fases de la AC: antes del comienzo de la AC (fase A), después de la AC de venas pulmonares (VPs) izquierdas (ACVPI) que se va a referir como fase B y la fase C que consta de los registros después de la AC de VPs derechas (ACVPD). La fase C coincide con el fin de la sesión de AC. Para cada registro del SC, se ha elegido un solo canal para su posterior análisis, teniendo en cuenta que para un paciente se tiene que utilizar el mismo canal en todas fases, asegurando así la fiabilidad de los resultados. El canal con la mayor amplitud y la menor presencia de ruido fue elegido para cada paciente por dos expertos de electrofisiología.

3. Métodos

El preprocesado se inició con la eliminación de ruido y la eliminación media de los registros de derivación II y SC. Estos últimos se sometieron a cancelación ventricular [7–9]. Se observó la presencia de latidos ectópicos en los registros de la derivación II con una prevalencia del < 4% del total de latidos. Se aplicó la interpolación lineal para la corrección de los latidos ectópicos [10]. La detección y delineación de las ondas P para la derivación II y las ondas de activación local (OALs) para los registros del SC fue el último paso del preprocesamiento, cuyos detalles se describen en otros estudios [11–14].

El análisis principal consistió en características basadas en la activación y en el registro. Las características basadas en la activación se calcularon en cada onda P y OAL del SC y luego se promediaron para cada registro. La duración, la amplitud, el área y la tasa de pendiente (TP) a 20% de la duración de la activación se calcularon así [14]. Se utilizó la siguiente ecuación para calcular la TP de la activación *n* a 20% de su duración

$$TP_{20}^n = \frac{A(t_{20}) - A(t_{inicio})}{t_{20} - t_{inicio}}, \quad (1)$$

donde TP_{20}^n es la TP_{20} de la activación *n*, $A(t_{20})$ y $A(t_{inicio})$ son el amplitud en 20% (TP_{20}) de la duración de la activación y en el inicio de la misma, respectivamente y t_{20} , t_{inicio} los correspondientes puntos de muestra.

El efecto de las fluctuaciones del ritmo cardíaco fue mitigado para la duración, el área y la TP, que fueron corregidos

por un factor (FC) de la siguiente forma:

$$FC(caracteristica_i) = \frac{1000}{R_i - R_{i-1}}, \quad (2)$$

donde $R_i - R_{i-1}$ es el intervalo entre dos picos R consecutivos dentro de los cuales se encuentra la activación *i*.

Las características basadas en el registro se calcularon para cada señal, proporcionando un valor único que describe los índices reclutados. Para ello, se utilizaron los índices de VRC en el dominio del tiempo (SDNN, VARNN, RMSSD) [15] considerando los intervalos entre activaciones auriculares de las grabaciones de la derivación II (ondas P) y del SC (OALs).

Para el análisis estadístico, se investigaron las correlaciones lineales entre cada onda P y OAL mediante regresión lineal (RL) con validación cruzada de 10 pliegues y correlación de Pearson (CP). Esta última también se utilizó para comparar la variación que cada transición de fase causaba a las características de las ondas P y a las de las OALs. Las correlaciones no lineales en cada onda P/OAL se investigaron con la entropía de muestra cuadrática cruzada (EMCC) con $m = 1$ y $r = 0,35$ [16].

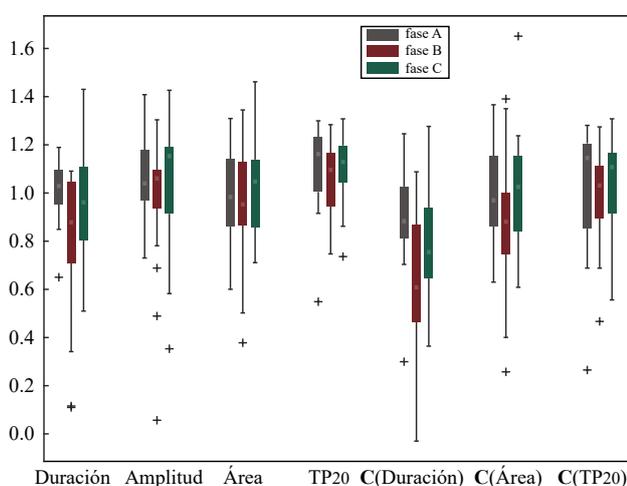
4. Resultados

La tabla 1 muestra los resultados de las correlaciones lineales entre las características de cada onda P y OAL. Como la RL se realiza utilizando sólo una variable independiente, la potencia estadística es la misma que en el análisis de Pearson. Por lo tanto, el valor *p* sólo se muestra una vez y corresponde a ambos análisis.

Los resultados de RL revelaron correlaciones más débiles que el análisis de CP debido a la validación cruzada, que añade solidez a la prueba ya que la base de datos es relativamente pequeña. La amplitud y el área mostraron correlaciones poco significativas en la fase A, mostrando una mayor concordancia en la fase B, justo después de la ACVPI. Aunque en principio no se encontraron patrones similares en la duración de las ondas P y OALs, la corrección de la duración reveló correlaciones moderadas estadísticamente significativas entre sí (hasta 69,91% para CP y 47,25% para RL, $p < 0,0001$). La corrección también aumentó la correlación del área, aunque en menor medida.

Tabla 2. Correlaciones CP para la variación provocada por la AC en cada transición de fases. Los resultados estadísticamente significativos se muestran en color gris. A: antes de la AC; B: ACVPI; C: ACVPD.

Características	Δ_{A-B}		Δ_{A-C}		Δ_{B-C}	
	CP [%]	valor p	CP [%]	valor p	CP [%]	valor p
Duración	34,82	0,1567	13,31	0,5984	-17,58	0,4854
Amplitud	8,67	0,7326	3,56	0,8885	39,20	0,1077
Área	-43,37	0,0722	-20,56	0,4131	36,60	0,1352
TP ₂₀	26,29	0,2919	-26,10	0,2955	7,77	0,7593
C(Duración)	85,72	< 0,0001	78,89	< 0,0001	53,90	0,0210
C(Área)	2,76	0,9135	9,01	0,7220	42,64	0,0776
C(TP₂₀)	28,33	0,2546	-28,46	0,2523	5,88	0,8167
SDNN	48,33	0,0422	92,75	< 0,0001	89,15	< 0,0001
VARNN	17,44	0,4883	92,23	< 0,0001	86,63	< 0,0001
RMSSD	42,12	0,0817	51,23	0,0297	94,20	< 0,0001

**Figura 1.** Resultados del análisis no lineal mediante EMCC. Correlación perfecta hay en caso de EMCC = 0.

La tabla 2 muestra los resultados para el CP, cuando se investiga el efecto de cada transición de fase (A-B, B-C, A-C) tanto sobre las ondas P y las OALs como sobre las características de la VRC. Para las características basadas en activaciones pre-corrección, no hay correlaciones estadísticamente significativas. No obstante, la corrección potenció significativamente la correlación entre la alteración de la duración en las ondas P y OALs después de cada paso de AC, que se convirtieron una vez más en estadísticamente significativas ($p < 0,0210$) mostrando además valores de moderados a altos (53,90 % – 85,72 %). En cuanto a los índices de VRC, se observa alta similitud entre su evolución en ECGs y registros del SC (hasta 94,20 % para RMSSD).

Por último, el análisis de correlación no lineal se muestra en la figura 1. Como la correlación más alta se expresa con valor de EMCC de 0, los valores observados entre 0,6 y 1,3 no implican la existencia de patrones no lineales uniformes entre las características de las ondas P y OALs. A diferencia del análisis de correlación lineal, la corrección no tuvo un efecto significativo en el estudio de correlación no lineal. A pesar de que C(Duración) mostró los valores EMCC más bajos y, por lo tanto, las correlaciones no lineales más altas, éstas siguieron siendo no significativas.

5. Discusión

El SC es una estructura crítica tanto para la perpetuación de la FA como para la terapia más importante de la FA, la AC [6]. Hasta ahora, su comportamiento con respecto al procedimiento de AC ha sido poco estudiado. El propósito del presente estudio era esclarecer esta misma cuestión. Para ello, se han utilizado los métodos más populares para la evaluación del resultado de la AC y se han analizado los registros adquiridos tras los pasos cruciales del procedimiento.

Los estudios anteriores que investigan cualquier tipo de correlación entre ECGs y registros del SC indican la existencia de correlaciones lineales o no lineales entre las características de la aurícula derecha y los registros de superficie [17–19]. Aunque los hallazgos del presente estudio no muestran los mismos resultados, existen muchas diferencias entre la base de datos de los estudios actuales y los anteriores. Los estudios anteriores utilizan señales registradas durante la FA, un conjunto de pacientes con patologías adicionales y registros de la aurícula derecha.

En el presente estudio, las correlaciones entre las ondas P y OALs del SC fueron, en su mayoría, insignificantes o moderadas, no revelando un comportamiento del SC similar al que se ha observado cuando se analiza la aurícula como una entidad. Antes de la corrección, la correlación fue mayor cuando se midió la amplitud y el área. No obstante, la amplitud depende de la derivación y del canal del SC que se analice. En este trabajo se utilizaron los registros de derivación II, donde la onda P muestra la mayor amplitud [20] y se seleccionaron los canales de SC con alta amplitud como criterio. Solamente se observó alta correlación cuando se investigaron las alteraciones de las características de la VRC en el dominio del tiempo, posiblemente debido a las aplicaciones de energía de RF, como se ha informado en estudios anteriores [21].

Cuando se corrigieron las características, la correlación se amplificó significativamente tanto en términos de poder estadístico como en términos de magnitud. Después de esta corrección, se encontró una correlación moderada entre la duración de las ondas P y las OALs del SC y una corre-

lación de moderada a alta entre la modificación de la duración en los registros de derivación II y del SC para cada transición de la AC. Dada la gran importancia de la duración de las ondas P para la predicción del resultado de la AC [3, 5], la corrección se considera un paso fundamental para la comparación entre los registros de superficie e invasivos. Sin este paso, las correlaciones lineales de la duración quedarían camufladas. No se ha observado que la corrección tenga el mismo efecto en el análisis de correlación no lineal, en el que tampoco se han reportado patrones significativos.

6. Conclusiones

El efecto de la AC es diferente en el SC que el efecto observado por el análisis de ondas P. Se cree que estos últimos describen con mayor precisión las alteraciones del sustrato auricular, ya que contienen información sobre la totalidad de las aurículas. La corrección de acuerdo con las fluctuaciones de la frecuencia cardíaca puede potenciar las correlaciones entre grabaciones superficiales e invasivas y se recomienda encarecidamente su adopción.

Agradecimientos

Trabajo cofinanciado por los proyectos de investigación competitiva DPI2017-83952-C3 de MINECO-AEI-FEDER-UE, SBPLY/17/180501/000411 de la JCCLM y AICO/2021/286 de la GVA.

Referencias

- [1] Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, et al. 2020 esc guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the european association for cardiothoracic surgery (eacts): The task force for the diagnosis and management of atrial fibrillation of the european society of cardiology (esc) developed with the special contribution of the european heart rhythm association (ehra) of the esc. *European heart journal* February 2021;42:373–498. ISSN 1522-9645.
- [2] Oral H, Knight BP, Tada H. Pulmonary vein isolation for paroxysmal and persistent atrial fibrillation. *Circulation* 2002;11:83. ISSN 1062-1458.
- [3] Maan A, Mansour M, Ruskin JN, Heist EK. Impact of catheter ablation on p-wave parameters on 12-lead electrocardiogram in patients with atrial fibrillation. *Journal of electrocardiology* 2014;47:725–733. ISSN 1532-8430.
- [4] Hu X, Jiang J, Ma Y, Tang A. Novel p wave indices to predict atrial fibrillation recurrence after radiofrequency ablation for paroxysmal atrial fibrillation. *Medical science monitor international medical journal of experimental and clinical research* July 2016;22:2616–2623. ISSN 1643-3750.
- [5] Chen Q, Mohanty S, Trivedi C, Gianni C, Della Rocca DG, Canpolat U, et al. Association between prolonged p wave duration and left atrial scarring in patients with paroxysmal atrial fibrillation. *Journal of cardiovascular electrophysiology* October 2019;30:1811–1818. ISSN 1540-8167.
- [6] Boles U, Gul EE, Enriquez A, Starr N, Haseeb S, Abdollah H, Simpson C, Baranchuk A, Redfearn D, Michael K, et al. Coronary sinus electrograms may predict new-onset atrial fibrillation after typical atrial flutter radiofrequency ablation (cse-af). *Journal of atrial fibrillation* 2018;11(1).
- [7] García M, Martínez-Iniesta M, Ródenas J, Rieta JJ, Alcaraz R. A novel wavelet-based filtering strategy to remove powerline interference from electrocardiograms with atrial fibrillation. *Physiological measurement* November 2018; 39:115006. ISSN 1361-6579.
- [8] Martínez-Iniesta M, Ródenas J, Rieta JJ, Alcaraz R. The stationary wavelet transform as an efficient reductor of powerline interference for atrial bipolar electrograms in cardiac electrophysiology. *Physiological measurement* July 2019;40:075003. ISSN 1361-6579.
- [9] Alcaraz R, Rieta JJ. Adaptive singular value cancellation of ventricular activity in single-lead atrial fibrillation electrocardiograms. *Physiological measurement* 2008; 29(12):1351–69.
- [10] Martínez A, Alcaraz R, Rieta JJ. Detection and removal of ventricular ectopic beats in atrial fibrillation recordings via principal component analysis. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual International Conference* 2011;2011:4693–4696. ISSN 2694-0604.
- [11] Martínez A, Alcaraz R, Rieta JJ. A new method for automatic delineation of ecg fiducial points based on the phasor transform. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 2010; 2010:4586–4589. ISSN 2375-7477.
- [12] González F, Alcaraz R, Rieta JJ. Electrocardiographic p-wave delineation based on adaptive slope gaussian detection. In *Computing in Cardiology, CinC 2007, Rennes, France, September 24-27, 2017*. www.cinc.org, 2017; .
- [13] Osorio DI, Alcaraz R, Rieta JJ. A fractionation-based local activation wave detector for atrial electrograms of atrial fibrillation. In *Computing in Cardiology, CinC 2007, Rennes, France, September 24-27, 2017*. 2017; 1–4.
- [14] Vranka A, Bertomeu-González V, Osca J, Ravelli F, Alcaraz R, Rieta JJ. Study on how catheter ablation affects atrial structures in patients with paroxysmal atrial fibrillation: The case of the coronary sinus. In *2020 International Conference on e-Health and Bioengineering (EHB)*. IEEE, 2020; 1–4.
- [15] Heart rate variability: standards of measurement, physiological interpretation and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology. *Circulation* March 1996; 93:1043–1065. ISSN 0009-7322.
- [16] Alcaraz R, Rieta JJ. Nonlinear synchronization assessment between atrial and ventricular activations series from the surface ecg in atrial fibrillation. *Biomedical Signal Processing and Control* 2013;8(6):1000–1007.
- [17] Liu Z, Hayano M, Hirata T, Tsukahara K, Quin Y, Nakao K, et al. Abnormalities of electrocardiographic p wave morphology and their relation to electrophysiological parameters of the atrium in patients with sick sinus syndrome. *Pacing and clinical electrophysiology PACE* January 1998; 21:79–86. ISSN 0147-8389.
- [18] Yahyazadeh S, Firoozabadi S, Haghjoo M, Parvaneh S. Quantitative relation between chaotic features of surface electrocardiogram and intracardiac electrogram. In *2010 Computing in Cardiology*. ISSN 2325-8853, Sep. 2010; 593–596.
- [19] Alcaraz R, Hornero F, Rieta JJ. Validation of surface atrial fibrillation organization indicators through invasive recordings. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2011; 5519–5522.
- [20] Meek S, Morris F. Introduction. ii—basic terminology. *BMJ Clinical research* ed February 2002;324:470–473. ISSN 1756-1833.
- [21] Misek J, Veternik M, Tonhajzerova I, Jakusova V, Janoušek L, Jakus J. Radiofrequency electromagnetic field affects heart rate variability in rabbits. *Physiological Research* 2020;69(4).

Reconocimiento Automático de Posiciones Radiográficas en Radiología

R. C. Giménez¹, C. F. Del Cerro^{1,2}, K. Sosenko³, M. Desco^{1,2,4,5}, J. García-Blas⁶, M. Abella^{1,2,4}

¹ Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, España

² Dep. de Bioingeniería e Ingeniería Aeroespacial, Universidad Carlos III de Madrid, Madrid, España

³ SEDECAL S.A., Madrid, España

⁴ Centro Nacional Investigaciones Cardiovasculares Carlos III (CNIC), Madrid, España

⁵ Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Madrid, España

⁶ Dep. de Informática, Universidad Carlos III de Madrid, Madrid, España

Resumen

La calidad de una radiografía depende de parámetros que deben ser ajustados por un técnico antes de la adquisición. Una mala configuración de estos parámetros puede dar lugar a un error de exposición, obligando a la repetición de la prueba, con el consiguiente exceso de radiación ionizante para el paciente. La mayoría de los sistemas radiológicos modernos cuentan con un control automático que interrumpe la radiación en caso de superar cierta dosis, pero no existe ningún mecanismo para evitar el problema de una infra exposición o seleccionar la configuración óptima para cada adquisición. Este trabajo presenta un nuevo método basado en aprendizaje profundo para la detección automática de la posición radiográfica, ayudando a la correcta selección de los parámetros de adquisición. La base de datos está formada por imágenes a color de setenta y cinco voluntarios, adquiridas en dos salas de rayos X. La evaluación del método muestra la viabilidad del mismo para ser utilizado en tiempo real permitiendo reducir la ocurrencia de errores de exposición, sin obstaculizar una adquisición diagnóstica de radiografía.

1. Introducción

La radiografía constituye entre un 51.7 % y un 55.1 % de los estudios de imagen médica realizados en la clínica [1], [2]. Por ello, cualquier mejora en la seguridad y calidad de imagen de los sistemas radiológicos tienen un gran impacto en el sistema de salud.

La calidad de una radiografía depende de los parámetros de exposición: voltaje de la fuente (kVp), corriente y tiempo de exposición (mAs) y distancia de la fuente al detector. Una errónea selección de estos parámetros puede resultar en imágenes inválidas para su uso diagnóstico y requiriendo la repetición de la adquisición. Algunos estudios sitúan la proporción de estas repeticiones en la práctica clínica entre el 10-15 % [3]. Estos parámetros de exposición dependen de la parte del cuerpo bajo estudio, lo que se conoce como posición radiográfica, y son ajustados por un técnico antes de la adquisición usando tablas de valores precalculados en base a estándares [4] y directrices del fabricante del sistema. La mayoría de máquinas modernas cuentan con un Control Automático de Exposición (CAE) [5], un dispositivo que interrumpe la radiación si detecta una dosis superior a la necesaria en el detector. El CAE, por tanto, es capaz de mitigar errores de

exposición en los que los parámetros son mayores de los óptimos, pero no son útiles cuando los parámetros de exposición son inferiores a los necesarios. Por ello, sería beneficioso contar con un mecanismo que ajuste de forma supervisada los parámetros de exposición a valores óptimos para cada posición radiográfica. Un posible enfoque para este problema es el uso de técnicas de reconocimiento de imágenes mediante aprendizaje profundo. En radiografía, esta tecnología se ha limitado en su mayor parte al etiquetado diagnóstico de imágenes ya adquiridas [6]–[8].

Este trabajo presenta RadPosNet, un método basado en aprendizaje profundo para obtener la posición radiográfica de un paciente a partir de una imagen en color. La salida de este método permitiría ajustar automáticamente el kVp y los mAs de una adquisición a los valores óptimos, lo que reduciría la incidencia de errores de exposición en la práctica clínica.

2. Construcción de la base de datos

Para la construcción de la base de datos se ha contado con la colaboración de la compañía SEDECAL, uno de los principales fabricantes de equipo radiológico del mundo.

Un total de 75 voluntarios fueron reclutados, de los cuales 50 fueron adquiridos en una sala de rayos experimental en el departamento de investigación de SEDECAL en Algete (Madrid) y los 25 restantes en el laboratorio de radiología del campus de Leganés de la Universidad Carlos III de Madrid (UC3M). De cada voluntario se obtuvieron 66 posiciones radiográficas, correspondientes a las más usadas en clínicas y escogidas en colaboración con técnicos del Hospital Universitario Gregorio Marañón de Madrid.

La adquisición de cada posición radiográfica se llevó a cabo usando imágenes a color (RGB) con una cámara de lente gran angular, la *YI 4K Action*, con un campo de visión de 155° y un tamaño de matriz de 3000×4000 píxeles. Dicha cámara fue atornillada a la caja del colimador en la misma posición en ambas salas de rayos, como se muestra en la Figura 1.

A partir de las imágenes adquiridas, se generaron dos bases de datos: una obtenida sin la luz del colimador (*RGB*) y otra

con el colimador encendido (*RGBcl*). Ambas bases de datos fueron divididas en dos conjuntos de datos:

- *Conjunto de entrenamiento*. Construido con 40 voluntarios de SEDECAL y 20 de UC3M, dando lugar a 3960 imágenes en total.
- *Conjunto de validación*. Construido con los restantes 10 voluntarios de SEDECAL y 5 de UC3M, con un total de 990 imágenes.

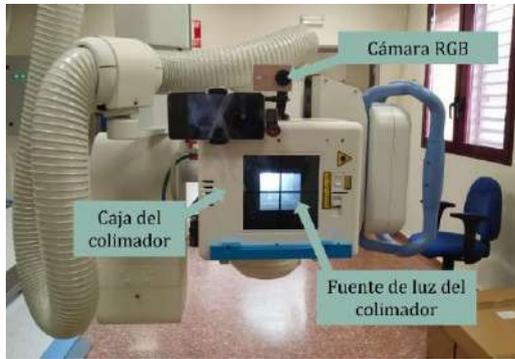


Figura 1. Configuración de la cámara durante la adquisición de la base de datos.

3. Método propuesto

Para construir RadPosNet, se generaron los modelos de aprendizaje profundo a través de un script de entrenamiento desarrollado en Python (versión 3.6). Los modelos congelados fueron posteriormente trasladados a la aplicación final que se instala en el PC de control de la sala de rayos.

3.1. Módulo de entrenamiento

Para la selección de la arquitectura de aprendizaje profundo se tuvieron en cuenta los resultados de la competición de reconocimiento de imágenes *ImageNet Large Scale Visual Recognition Challenge* [9]. La red escogida fue la ganadora de la edición de 2015, ResNet [10], por su rendimiento y la gran cantidad de documentación disponible online.

Para mitigar el efecto negativo del pequeño tamaño de nuestra base de datos, se ha incluido utilizamos aprendizaje de transferencia [11], una técnica en la que se reutilizan modelos pre-entrenados con bases de datos de gran escala y se reentrenan con otra base de datos para otro propósito. El módulo de entrenamiento hace uso de la librería PyTorch (versión 1.8.1) para cargar un modelo de ResNet-34 pre-entrenado con la base de datos ImageNet [12]. Dicho modelo fue construido para 1000 categorías, por lo que se modificó su capa de salida para ajustarse a las 66 posiciones radiográficas de este trabajo. Durante el entrenamiento del clasificador se usa la entropía cruzada como función de coste, minimizándola con AdamW [13] e introduciendo *mini-batches* de 32 imágenes. Este optimizador utiliza un enfoque de tasa de aprendizaje adaptativa similar al de Adam [14], pero consigue implementar correctamente la regularización L2, con la intención de reducir el sobreajuste del modelo final.

En un primer paso, para determinar si las diferencias entre salas de rayos tenían influencia sobre la clasificación del modelo, se entrenó la red únicamente con imágenes de la

sala de SEDECAL, y se evaluó el modelo resultante con las imágenes de la UC3M. La baja precisión obtenida (59.18 %) motivó la implementación de estrategias para limitar la imagen de entrada a la región donde el paciente sería irradiado, es decir, la región del colimador. Para ello, se evaluaron dos estrategias:

- *Table*. Recorte cuadrado alrededor de la luz del colimador del tamaño de una mesa de radiografía estándar, de 90 centímetros.
- *Collim*. Recorte cuadrado de una región un 20% mayor que la limitada por la luz del colimador.

La Figura 2 muestra un ejemplo de los dos recortes para una adquisición de cadera.

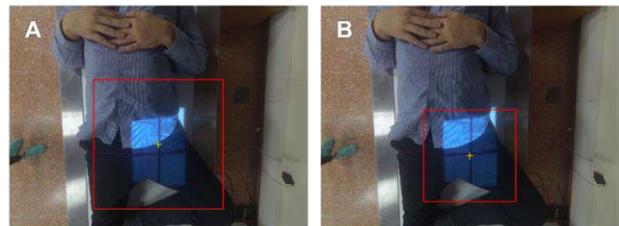


Figura 2. Ejemplo de las dos estrategias de recorte: región cuadrada de 90 cm (A) y región proporcional a la delimitada por la luz del colimador (B).

Las imágenes recortadas son remuestreadas a 224 píxeles y normalizadas teniendo en cuenta la media y desviación estándar de las imágenes de la base de datos ImageNet.

Para aumentar la capacidad de generalización del modelo y evitar la influencia de errores de calibración de la cámara, se implementó una estrategia de aumento de datos basada en pequeñas rotaciones aleatorias y transformación de histograma en las imágenes del conjunto de entrenamiento.

La tasa de aprendizaje inicial se seleccionó mediante un test desarrollado durante este trabajo, basado en pruebas previamente presentadas por Leslie, *et al.* [15]. El algoritmo de Leslie genera una estimación del valor de la función de coste en un rango de tasas de aprendizaje. Mediante esos datos, calculamos el cambio en la función de coste con la primera derivada y elegimos la tasa de aprendizaje que disminuye más rápido el coste. Un ejemplo de esta prueba se presenta en la Figura 3.

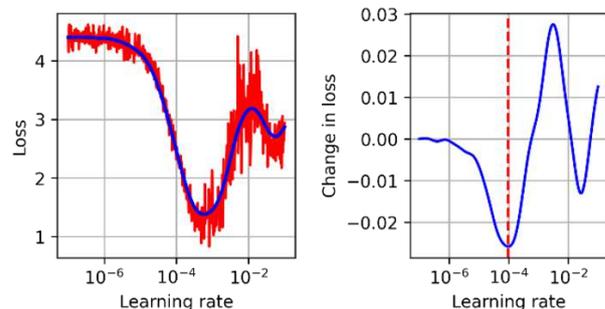


Figura 3. Gráfica ajustada de los valores generados por el algoritmo de Leslie (izquierda) y cambio de la función de coste con respecto a la tasa de aprendizaje (derecha).

Para acelerar la convergencia, se desarrolló un nuevo algoritmo para calcular la reducción de la tasa de aprendizaje a lo largo del entrenamiento, combinando la

reducción exponencial con la estrategia de *cosine annealing* presentada por Loshchilov, *et al.* [16].

El módulo de entrenamiento completo incluye el preprocesado de la base de datos (recorte, remuestreo y normalización), su división en los conjuntos de entrenamiento y validación, las mejoras de aprendizaje (aumento de datos, test de tasa de aprendizaje inicial y reducción de tasa de aprendizaje) y el bucle de entrenamiento.

3.2. Implementación en sistema real

El modelo generado mediante el módulo de entrenamiento en Python fue convertido a C++ usando la API *Libtorch*. Para el preprocesado de las imágenes de entrada de la aplicación final se utilizó la librería de C++ *OpenCV*. La compilación e inclusión de todas las librerías dinámicas con el ejecutable generado permite el uso de estos modelos en sistemas que no tengan Python/PyTorch instalados. Además de ello, la aplicación hace uso únicamente del procesador del ordenador para evaluar las imágenes, evitando así la necesidad de una tarjeta gráfica discreta. La Figura 4 muestra el esquema del proceso completo.

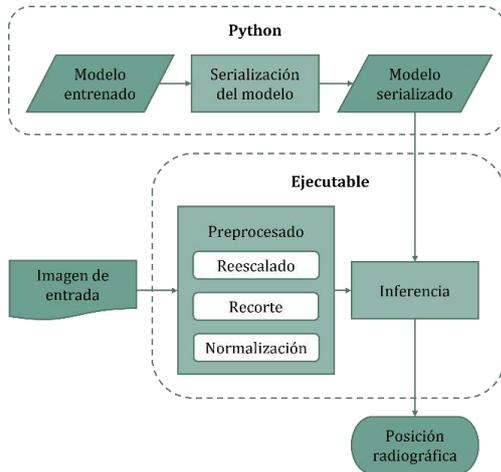


Figura 4. Diagrama de flujo de la conversión de modelos entrenados a serializados y su uso en el ejecutable final.

El prototipo final usa una interfaz de línea de comandos y fue compilado para ser utilizado en máquinas con Windows® y con un procesador de arquitectura x86-64.

4. Evaluación y resultados

Se han evaluado los cuatro modelos resultantes de la combinación de las dos estrategias de recorte (*Table* y *Collim*) y las dos bases de datos (*RGB* y *RGBcl*). La Tabla 1 presenta estos resultados, con su entropía cruzada y precisión.

Modelo	Entropía cruzada	Precisión (%)
Table RGB	0.3564	91.94
Table RGBcl	0.1874	95.56
Collim RGB	0.2391	94.49
Collim RGBcl	0.1966	95.44

Tabla 1. Rendimiento de los modelos usados por RadPosNet

Los errores de clasificación de los modelos se evaluaron mediante matrices de confusión. Como se puede ver en la que se muestra en la Figura 5, para el caso de “Table RGB”, la mayoría de las predicciones forman parte de la diagonal, lo que indica que se corresponden con su etiqueta correcta (lo mismo se observó para los otros tres modelos). Los puntos fuera de la diagonal se corresponden mayoritariamente con posiciones anatómicas muy parecidas de la misma parte del cuerpo (mano, por ejemplo). Generalmente, este tipo de errores no tiene impacto en la aplicación final, ya que comparten parámetros de exposición (kVp y mAs).

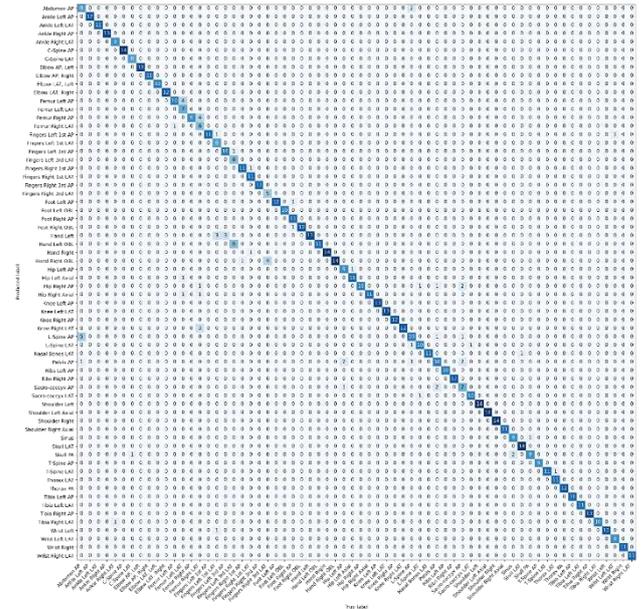


Figura 5. Matriz de confusión del modelo Table RGB. Las imágenes correctamente clasificadas se encuentran en la diagonal, y los errores fuera de ella.

Para evaluar el impacto de los errores en la aplicación final, se obtuvieron los pares de posiciones en los que fallaba el modelo en más de un 10% de los casos y que además generaban error de exposición (distinto kVp y/o mAs). La Tabla 2 muestra estos pares de posiciones con error, denominados confusiones relevantes y los errores de exposición correspondientes.

Modelo	Confusiones relevantes	Error medio kVp	Error medio mAs
Table RGB	9	5 ± 6	1 ± 3
Table RGBcl	5	11 ± 8	13 ± 14
Collim RGB	4	15 ± 19	10 ± 14
Collim RGBcl	2	18 ± 14	8.0 ± 1.3

Tabla 2. Confusiones relevantes provocadas por cada modelo y sus respectivos errores de exposición medios

El tiempo medio de ejecución del prototipo final fue de 1.01 ± 0.14 segundos medido en 6 imágenes tomadas a

tiempo real, evaluadas en un ordenador equipado con un procesador Intel® Core™ i7-6700HQ.

5. Discusión y conclusiones

RadPosNet presenta un novedoso método para obtener la posición radiográfica de imágenes de pacientes en tiempo real de forma automática. Esto permite la configuración de los parámetros de exposición, evitando la irradiación innecesaria de pacientes por las repeticiones de adquisición debidas a errores en la selección de estos parámetros.

Gracias al uso de la transferencia de aprendizaje, el método propuesto alcanza una precisión de clasificación de hasta el 95.56 % pese a la limitada base de datos utilizada. El estudio de las matrices de confusión de muestra que la mayoría de los errores de clasificación no tienen un impacto en la aplicación final, ya que se producen entre posiciones radiográficas que comparten parámetros de exposición y por tanto no resultarían en un error en la adquisición. No obstante, aunque en menor medida, también se han producido confusiones que resultarían en errores de exposición. El impacto de estas confusiones podría ser minimizado usando el método de forma supervisada o aumentando el tamaño de la base de datos para entrenamiento. ResNet se seleccionó frente a otras arquitecturas más modernas y con mayor rendimiento, debido a la cantidad de documentación disponible. El uso de arquitecturas de reconocimiento de imágenes del estado del arte como EfficientNet [17] podría también aumentar la precisión.

Se ha observado cierta influencia del fondo de la imagen (suelo y paredes de la sala de rayos) en la clasificación de los modelos, que se ha mitigado en parte con la implementación de estrategias de recorte de la región de interés. En el futuro esta influencia se podría reducir aún más mediante la inclusión de más salas en la base de datos de entrenamiento, métodos de eliminación de fondo o la generación de datos sintéticamente mediante el reemplazo del fondo de las imágenes.

El tiempo de ejecución y rendimiento del prototipo final muestra la viabilidad del método para ser utilizado en tiempo real permitiendo reducir la ocurrencia de errores de exposición, sin obstaculizar una adquisición diagnóstica de radiografía. La implementación del método en un sistema real sólo requiere una cámara RGB y un ejecutable instalado en el PC de control.

Agradecimientos

Este trabajo ha sido financiado por el Programa de apoyo a la realización de proyectos interdisciplinarios de I+D para jóvenes investigadores de la Universidad Carlos III de Madrid 2019-2020 en el marco del Convenio Plurianual Comunidad de Madrid- Universidad Carlos III de Madrid (proyecto DEEPCT-CM-UC3M) y por CRUE Universidades, CSIC y el Banco Santander (Fondo Supera, proyecto RADCOV19). El CNIC está financiado por el Ministerio de Ciencia, Innovación y Universidades y la fundación PRO-CNIC y es un centro de excelencia Severo Ochoa (SEV-2015-0505).

Referencias

- [1] «Diagnostic Imaging Dataset Annual Statistical Release 2019/20», NHS England, oct. 2020.
- [2] R. Smith-Bindman *et al.*, «Use of Diagnostic Imaging Studies and Associated Radiation Exposure for Patients Enrolled in Large Integrated Health Care Systems, 1996-2010», *JAMA*, vol. 307, n.º 22, jun. 2012, doi: 10.1001/jama.2012.5960.
- [3] D. Waaler y B. Hofmann, «Image rejects/retakes—radiographic challenges», *Radiat. Prot. Dosimetry*, vol. 139, n.º 1-3, pp. 375-379, abr. 2010, doi: 10.1093/rpd/ncq032.
- [4] European Commission, *European guidelines on quality criteria for diagnostic radiographic images: Eur 16260 EN*. European Commission, 1996.
- [5] J. T. Bushberg y J. M. Boone, *The essential physics of medical imaging*. Lippincott Williams & Wilkins, 2011.
- [6] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, y H. J. W. L. Aerts, «Artificial intelligence in radiology», *Nat. Rev. Cancer*, vol. 18, n.º 8, pp. 500-510, ago. 2018, doi: 10.1038/s41568-018-0016-5.
- [7] T. K. Kim *et al.*, «Deep Learning Method for Automated Classification of Anteroposterior and Posteroanterior Chest Radiographs», *J. Digit. Imaging*, vol. 32, n.º 6, pp. 925-930, dic. 2019, doi: 10.1007/s10278-019-00208-0.
- [8] X. Fang, L. Harris, W. Zhou, y D. Huo, «Generalized Radiographic View Identification with Deep Learning», *J. Digit. Imaging*, vol. 34, n.º 1, pp. 66-74, feb. 2021, doi: 10.1007/s10278-020-00408-z.
- [9] O. Russakovsky *et al.*, «ImageNet Large Scale Visual Recognition Challenge», *Int. J. Comput. Vis.*, vol. 115, n.º 3, pp. 211-252, dic. 2015, doi: 10.1007/s11263-015-0816-y.
- [10] K. He, X. Zhang, S. Ren, y J. Sun, «Deep Residual Learning for Image Recognition», jun. 2016.
- [11] M. Raghu, C. Zhang, J. Kleinberg, y S. Bengio, «Transfusion: Understanding Transfer Learning for Medical Imaging», en *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Red Hook, NY, USA: Curran Associates Inc., 2019.
- [12] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, y Li Fei-Fei, «ImageNet: A large-scale hierarchical image database», en *2009 IEEE Conference on Computer Vision and Pattern Recognition*, jun. 2009, pp. 248-255. doi: 10.1109/CVPR.2009.5206848.
- [13] I. Loshchilov y F. Hutter, «Decoupled Weight Decay Regularization», 2019. [En línea]. Disponible en: <https://openreview.net/forum?id=Bkg6RiCqY7>
- [14] D. P. Kingma y J. Ba, «Adam: A Method for Stochastic Optimization», *CoRR*, vol. abs/1412.6980, 2015.
- [15] L. N. Smith, «Cyclical Learning Rates for Training Neural Networks», en *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, mar. 2017, pp. 464-472. doi: 10.1109/WACV.2017.58.
- [16] I. Loshchilov y F. Hutter, «SGDR: Stochastic Gradient Descent with Warm Restarts», 2016.
- [17] M. Tan y Q. Le, «EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks», en *Proceedings of the 36th International Conference on Machine Learning*, jun. 2019, vol. 97, pp. 6105-6114. [En línea]. Disponible en: <https://proceedings.mlr.press/v97/tan19a.html>

EEG-InceptionGen: Una Red Convolutiva de Propósito General para la Clasificación de señales EEG

Eduardo Santamaría-Vázquez^{1,2}, Sergio Pérez-Velasco¹, Víctor Martínez-Cagigal^{1,2}, Roberto Hornero^{1,2}

¹ Grupo de Ingeniería Biomédica, Universidad de Valladolid, Valladolid, España

² Centro de Investigación Biomédica en Red de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), España

Resumen

La electroencefalografía (EEG) es una técnica ampliamente utilizada en campos como la investigación del sistema nervioso central, el diagnóstico clínico y los sistemas brain-computer interface (BCI), que requieren del análisis automático de esta señal para la detección de patrones de interés. Sin embargo, este análisis es de gran complejidad, y habitualmente se necesita una gran inversión en tiempo y esfuerzo para identificar las características más relevantes de estos patrones y diseñar un algoritmo de detección. Con el objetivo de resolver esta limitación, hemos diseñado desarrollado y validado una red neuronal convolutiva multipropósito, llamada EEG-InceptionGen, aplicable a una gran cantidad de tareas de clasificación EEG con un alto rendimiento. El modelo ha sido validado en 3 tareas distintas como el control de un sistema BCI basado en potenciales relacionados con eventos, la detección de la imaginación motora de las manos derecha e izquierda y la detección de episodios epilépticos. En la primera se alcanzó una precisión del 74.27% en una base de datos de 31 sujetos con grave discapacidad. En la segunda tarea, un 87.46% en una base de datos de 105 sujetos. En la tercera, se detectaron correctamente el 96.52% de los episodios epilépticos utilizando una base de datos de 24 sujetos. En todos los casos, los análisis fueron inter-sujeto. Los resultados alcanzados son comparables, o superan, al estado del arte en cada una de estas tareas, poniendo de manifiesto la utilidad del modelo presentado.

1. Introducción

La electroencefalografía (EEG) es una técnica que permite registrar la actividad eléctrica de las neuronas superficiales del córtex mediante la colocación de electrodos sobre el cuero cabelludo [1]. Debido a su excelente resolución temporal y a la utilización de un equipo portátil, no invasivo y de bajo coste, la señal de EEG es utilizada en numerosas aplicaciones. Entre las más importantes se encuentran la investigación del sistema nervioso central, el diagnóstico de patologías como la epilepsia, la monitorización del nivel de consciencia, o la caracterización de las fases sueño [1]. Además, la irrupción en los últimos años de los sistemas brain-computer interface (BCI) ha incrementado en gran medida las aplicaciones del EEG, fomentando su desarrollo. Entre las aplicaciones de los sistemas BCI, se encuentran aplicaciones asistivas para la comunicación y el control de dispositivos orientadas a personas con grave discapacidad, la rehabilitación motora y cognitiva, e incluso el entretenimiento [1]. Cabe destacar que todas estas tareas requieren de la detección (manual o automática) de patrones específicos dentro de la señal EEG asociados a la actividad cerebral concreta que se está estudiando.

Desafortunadamente, el EEG no está exento de limitaciones, entre las que destacan su baja resolución espacial y su gran sensibilidad a artefactos ruidosos [1]. Por estas razones, sumadas a la inherente complejidad de la actividad cerebral, el análisis de la señal de EEG plantea numerosos retos.

La metodología habitual para la detección de patrones en el EEG de forma automática consiste en la extracción, selección y clasificación de características específicamente diseñadas para un problema concreto, destacando aquellas basadas análisis temporal, espectral y no lineal [2]. Sin embargo, el diseño de estas características requiere un nivel alto de conocimiento previo sobre el problema a tratar, siendo altas las probabilidades de seleccionar un conjunto subóptimo que afecte al rendimiento del modelo [2]. En este sentido, los modelos basados en *deep learning* han permitido en los últimos años mejorar la detección de patrones en el EEG por su capacidad para extraer conjuntos de características optimizados a partir de la señal cruda [2].

Debido a las características especiales del EEG, este avance no ha sido tan relevante como en otros campos [2]. La mayoría de los modelos presentados hasta el momento han sido optimizados para tareas concretas, y no son validados de forma más general. Por ejemplo, se han diseñado modelos para detección de episodios epilépticos [3], ritmos sensoriomotores (*sensorimotor rhythms*, SMR) durante tareas de imaginación motora [4], potenciales relacionados con eventos (*event-related potentials*, ERP) [5] o esfuerzo mental [6]. Sin embargo, ninguno de los mencionados ha sido validado en otros escenarios. Esto es debido a las diferentes características intrínsecas de cada análisis (número de canales, frecuencias de interés, longitud de los segmentos, etc). Por otro lado, muchos de los avances presentados en estos modelos podrían ser aplicados en otras ramas del análisis del EEG. En este contexto, el diseño de arquitecturas que se puedan adaptar a diferentes tareas resulta de especial relevancia, acelerando los avances de este campo.

El objetivo de este estudio es doble: (i) diseñar y desarrollar una red convolutiva aplicable en múltiples escenarios de clasificación EEG; y (ii) validar el modelo en 3 tareas diferentes como son el control de un sistema BCI asistivo basado en ERPs, la detección de SMR asociados a la imaginación del movimiento de las manos derecha e izquierda, y la detección de episodios epilépticos en sujetos pediátricos.

2. Métodos

2.1. EEG-InceptionGen

En este estudio, se ha utilizado como base la red convolucional EEG-Inception para el análisis de la señal de EEG. Esta red fue presentada en nuestro anterior trabajo [5], mostrando unos excelentes resultados para una tarea de clasificación de ERPs. Sin embargo, la estructura fija descrita en [5] no permite la aplicación de la red a tareas que requieran parámetros de entrada diferentes a los definidos (i.e., número de canales, longitud de época y número de clases). En este trabajo, se han introducido las modificaciones necesarias para permitir la configuración de estos parámetros a la vez que se mantienen las proporciones internas de la red de forma automática, aumentando la versatilidad del modelo para que pueda aplicarse en otras tareas de clasificación EEG con éxito. El nuevo modelo, varios ejemplos de uso y una descripción detallada de todos los hiperparámetros, pueden encontrarse en este [repositorio](#).

La arquitectura de EEG-InceptionGen (Figura 1) está compuesta por 2 bloques Inception y un módulo de salida. Este modelo permite configurar la longitud de la época, el número de canales EEG y el número de clases de salida. El primer bloque Inception está formado por tres ramas que realizan convoluciones 2D en el eje temporal con diferentes campos receptivos (tamaño del núcleo) en cada una de las ramas, realizando un análisis multiescala que permite extraer características más complejas a partir de la señal cruda del EEG [5]. En EEG-InceptionGen, las escalas en cada una de las ramas son configurables, ya que este hiperparámetro puede variar en función de la tarea a realizar. Posteriormente, se realiza una convolución tipo *depthwise* en el eje espacial, seleccionando los canales EEG más relevantes para cada una de las características temporales extraídas anteriormente [5]. El segundo bloque Inception vuelve a realizar un análisis temporal, utilizando las mismas escalas que el primero, pero teniendo en cuenta toda la información espacial. Por último, el módulo de salida reduce la dimensionalidad del conjunto paulatinamente para evitar el efecto de sobreentrenamiento. El número de capas convolucionales en este bloque varía en función de la longitud de la época a la entrada, aplicando una capa *GlobalAveragePooling2D* al final. Las características resultantes a la salida de este módulo son clasificadas por una capa *softmax*. Adicionalmente, a lo largo de la red se utiliza *average pooling*, activaciones tipo *exponential linear unit* (ELU), *batch normalization*, y *dropout* [5].

2.2. Tarea 1: sistema BCI basado en ERPs

La primera tarea de clasificación utilizada para validar EEG-InceptionGen es el control de un *speller* BCI basado en ERPs. Este sistema utiliza el paradigma *oddball* para permitir al usuario seleccionar un comando entre varias posibilidades mediante la detección automática de ERPs generados en el EEG mediante estimulación visual [5]. La principal aplicación de este sistema es la comunicación y el control de dispositivos para aumentar la calidad de vida de personas con grave discapacidad [5].

Uso	Estudios	Sujetos	Trials	Eventos
Pre-entrenamiento	[7]–[11]	66 sanos	4490	699405
Validación cruzada	[9], [10]	31 discapacitada motora	2333	309410

Tabla 1. Base de datos de la tarea 1. La columna *trials* indica el total de selecciones realizadas por los sujetos. La columna *eventos* es el número total de estimulaciones.

La base de datos utilizada estaba compuesta por 5 subconjuntos diferentes donde se utilizaba un sistema BCI basado en ERPs [7]–[11] (Tabla 1). Todos los subconjuntos fueron adaptados y homogeneizados con el siguiente algoritmo de preprocesado: submuestreo a 128 Hz, filtrado paso-banda entre 0.5 y 45Hz, selección de 8 canales monopares, extracción de épocas de 1s para cada evento de estimulación (128 muestras x 8 canales) y normalización *z-score*.

Para la validación de EEG-Inception en esta tarea, primero se realizó un pre-entrenamiento utilizando registros provenientes de 66 sujetos sanos. Posteriormente, para la evaluación se seleccionó un subconjunto formado por 31 sujetos con grave discapacidad, analizando el rendimiento de la red utilizando el algoritmo de validación cruzada *leave-one-subject-out* (LOSO). Por tanto, se realizaron 31 iteraciones, en las que la red se entrenaba con 30 sujetos y se evaluaba en el restante. Es importante destacar que la clasificación fue inter-sujeto, eliminando la habitual etapa de calibración con el sujeto de test [8].

2.3. Tarea 2: clasificación de imaginación motora

La segunda tarea de validación consistió en una tarea de clasificación de imaginación motora. En esta tarea, los sujetos imaginan el movimiento de la mano derecha o mano izquierda, provocando la aparición de una actividad, denominada SMR, asociada principalmente a la banda de los 12-15 Hz en la zona motora contralateral a la mano imaginada [12]. Los SMR han sido ampliamente estudiados para el control de los sistemas BCI, aplicándose en los últimos años para rehabilitación motora y cognitiva de pacientes con ictus [1].

Los datos utilizados, provenientes de 4 bases de datos públicas diferentes, fueron adaptados y homogeneizados utilizando el siguiente pre-procesado: submuestreo a 128 Hz, filtrado paso-banda entre 0.1 y 45 Hz, selección de 32 canales monopares, extracción de épocas de 3s relativas a los eventos de imaginación motora (384 muestras x 32 canales) y normalización *z-score*.

Para la validación del modelo en esta tarea, primero se realizó un pre-entrenamiento utilizando 3 bases de datos [12]–[14] (Tabla 2). Posteriormente, se evaluó la precisión del modelo en la base de datos Physiobank Motor/Mental Imagery database [15]. De forma análoga a la tarea anterior, se aplicó LOSO como validación cruzada, realizando 105 iteraciones correspondientes al número de sujetos incluidos en esta última base de datos. De esta forma, tampoco se realizó una calibración del modelo utilizando datos del sujeto testado en cada iteración [4].

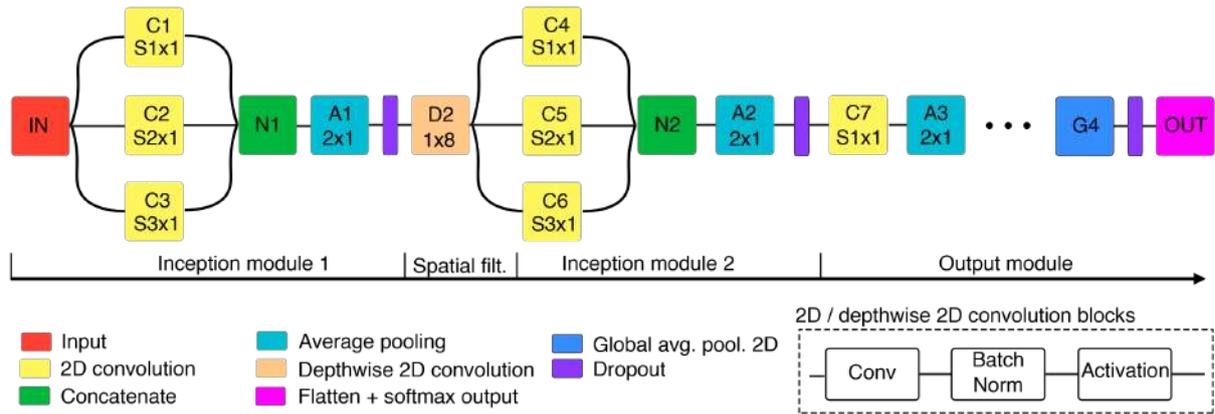


Figura 1. Representación esquemática de EEG-InceptionGen. El número de capas en el módulo de salida se adapta automáticamente en función de los parámetros de entrada para evitar el sobre-entrenamiento. $S1$, $S2$ y $S3$ representan las escalas temporales (campos receptivos) de las convoluciones y son configurables. En este estudio $S1=64$, $S2=32$ y $S3=125$.

Uso	Estudios	Sujetos	Eventos
Pre-entrenamiento	[12]–[14]	158 sanos	171523
Validación cruzada	[17]	105 sanos	4714

Tabla 2. Base de datos de la tarea 2. La columna eventos se refiere al número total de imaginaciones realizadas.

Tarea 3: detección de episodios epilépticos en registros pediátricos

La tercera tarea consistió en la detección automática de crisis epilépticas. Estas crisis consisten en periodos de actividad cerebral anormal, de unas decenas de segundos de duración, caracterizados por la aparición de patrones rítmicos detectables en el EEG y localizados en cualquier punto del córtex [16]. Entre las motivaciones más importantes para la detección de eventos epilépticos de forma automática se encuentra la optimización del proceso de diagnóstico, en el que es habitual que un especialista tenga que analizar registros de varias horas o incluso días de duración [16]. Otra aplicación es la detección en tiempo real del episodio mediante la monitorización EEG, facilitando una rápida intervención de ser necesario [16].

La base de datos utilizada para la validación de EEG-InceptionGen en este problema de clasificación fue CHB-MIT [16]. Esta base de datos contiene registros de 22 sujetos de entre 9 y 42 horas de duración con un total de 198 crisis epilépticas marcadas por especialistas al inicio y final de cada evento [16]. Estos registros continuos fueron preprocesados aplicando: selección de 23 canales bipolares comunes a la mayoría de los registros, submuestreo a 128 Hz y filtrado paso-banda entre 0.1 y 30 Hz. Posteriormente se extrajeron épocas de 5s de duración con solapamiento del 50% (640 muestras x 23 canales). Debido a la desproporción entre las épocas no ictales e ictales (i.e., asociadas a crisis epilépticas), se aplicó un submuestreo a la clase mayoritaria para balancear al 50% el número de ejemplos de cada clase en el entrenamiento del modelo.

El proceso de validación fue similar a las otras tareas, aplicando validación cruzada utilizando el algoritmo LOSO. En este caso, no se pre-entrenó la red previamente.

3. Resultados

Los resultados de los experimentos de validación, correspondientes a las 3 tareas se muestran en la Tabla 3. En los casos de las tareas 1 y 2, la comparación se realiza de forma directa con un modelo de referencia validado anteriormente en cada escenario, mientras que para la tarea 3 se realiza una comparación con la literatura existente.

Los resultados de la tarea 1 corresponden a la precisión, promediada por sujetos, de la decodificación de los comandos utilizando 15 secuencias de estimulación (15 estimaciones por fila y columna de la matriz). El modelo de referencia empleado fue EEG-Inception [5]. Los resultados de la tarea 2 corresponden a la precisión en la clasificación en la tarea de imaginación motora. El modelo de referencia empleado fue EEGNet [4], una red muy utilizada en la actualidad en el campo de los sistemas BCI basados en este paradigma. Al igual que en el caso anterior, se trata de una comparación directa repitiendo el mismo experimento. El modelo fue entrenado utilizando los hiperparámetros recomendados por estudio original [4]. Por último, los resultados de la tarea 3 muestran el porcentaje de episodios epilépticos correctamente detectados (i.e., ictal vs no ictal). En cuanto a los resultados por épocas, lo que equivaldría a una monitorización continua del paciente, EEG-InceptionGen alcanzó una precisión del 74.18%, una sensibilidad del 63.10% y una especificidad del 85.25%. En este caso, no se aplicó un modelo de referencia ya que los estudios revisados no incluían el código utilizado.

4. Discusión y conclusiones

En cuanto a la arquitectura de EEG-InceptionGen, cabe destacar que para cada una de las tareas se han utilizado épocas de señal de diferente longitud y número de canales sin que el rendimiento de la red se haya visto afectado. Esta robustez se debe al empleo de los módulos Inception y a un control efectivo del sobre-entrenamiento. En este sentido, los modelos basados en *deep learning* aplicados al EEG son muy sensibles al sobre-entrenamiento por la baja relación señal a ruido de esta señal [5]. EEG-InceptionGen controla este efecto mediante una adaptación automática

Tarea	Referencia	EEG-InceptionGen
1	73.81 ± 23.03 %	74.27 ± 19.92 %
2	83.08 ± 10.92 %	87.46 ± 09.33 %
3	* 96.00 % [17], 97.5% [3]	96.52 ± 10.37%

Tabla 3. Resultados de los experimentos de validación. En la tarea 3, el asterisco indica que la comparación es con estudios previos en la misma base de datos.

del módulo de salida para homogeneizar el número de características extraídas por la red independientemente de los parámetros de entrada y la introducción de una capa GlobalAveragePooling2D antes de la capa de clasificación. Gracias a estas características, se espera que esta arquitectura pueda alcanzar resultados satisfactorios en cualquier tarea de clasificación EEG supervisada.

Respecto a los resultados de validación, se puede apreciar como EEG-InceptionGen supera a la versión original del modelo en la tarea de clasificación de ERPs [5]. Además, es destacable la adaptación de la red a los otros escenarios de validación. En concreto, en la tarea 2 la mejora en la precisión ha sido del 4.4% respecto a EEGNet [4] en la comparación directa, lo que representa un incremento significativo. Además, Dose *et al.* [17] alcanzaron en una precisión de 80.38% en la misma base de datos, muy por debajo del resultado presentado en este estudio. Por su parte, los resultados de la tarea 3 son comparables a los de la literatura, con un 96.52 % de precisión en la detección de eventos epilépticos. Por ejemplo, Shoeb [16] alcanzó una precisión del 96.00% en la misma base de datos, mientras que en el estudio de Zhou *et al.* se mejoró ligeramente este valor, alcanzando un 97.5% de precisión utilizando transformaciones al dominio frecuencial [3].

En este estudio se ha presentado un nuevo modelo de red convolucional, denominado EEG-InceptionGen, con un diseño multipropósito adaptable a la mayoría de las tareas de clasificación EEG. Además, este modelo ha sido validado en 3 tareas diferentes: control de un sistema BCI basado en ERPs, clasificación de imaginación motora y detección de episodios de epilepsia; alcanzando unos excelentes resultados que se comparan, e incluso superan, a los alcanzados por modelos desarrollados para estos escenarios específicos.

Agradecimientos

Este trabajo ha sido financiado por los proyectos PID2020-115468RB-I00 y RTC2019-007350-1 del 'Ministerio de Ciencia, Innovación y Universidades' y 'Análisis y correlación entre la epigenética y la actividad cerebral para evaluar el riesgo de migraña crónica y episódica en mujeres' ('Cooperation Programme Interreg V-A Spain-Portugal POCTEP 2014–2020') de la Comisión Europea y FEDER, así como por CIBER-BBN a través del 'Instituto de Salud Carlos III'.

Referencias

[1] J. Wolpaw and E. W. Wolpaw, *Brain-computer interfaces: principles and practice*. OUP USA, 2012.

- [2] A. Craik *et al.*, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *J. Neural Eng.*, vol. 16, no. 3, 2019.
- [3] M. Zhou *et al.*, "Epileptic seizure detection based on EEG signals and CNN," *Front. Neuroinform.*, vol. 12, no. December, pp. 1–14, 2018.
- [4] V. J. Lawhern *et al.*, "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 056013, pp. 1–17, 2018.
- [5] E. Santamaria-Vazquez *et al.*, "EEG-Inception: A Novel Deep Convolutional Neural Network for Assistive ERP-based Brain-Computer Interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 2773–2782, 2020.
- [6] Z. Jiao, *et al.* "Deep Convolutional Neural Networks for mental load classification based on EEG data," *Pattern Recognit.*, vol. 76, pp. 582–595, 2018.
- [7] V. Martínez-Cagigal *et al.*, "An Asynchronous P300-Based Brain-Computer Interface Web Browser for Severely Disabled People," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 8, pp. 1332–1342, 2017.
- [8] V. Martínez-Cagigal *et al.*, "Towards an accessible use of smartphone-based social networks through brain-computer interfaces," *Expert Syst. Appl.*, vol. 120, pp. 155–166, 2019.
- [9] E. Santamaria-Vázquez *et al.*, "Asynchronous Control of ERP-Based BCI Spellers Using Steady-State Visual Evoked Potentials Elicited by Peripheral Stimuli," *IEEE Trans. neural Syst. Rehabil. Eng.*, vol. 27, no. 9, pp. 1883–1892, 2019.
- [10] B. Blankertz *et al.*, "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 153–159, 2006.
- [11] C. Guger *et al.*, "How many people are able to control a P300-based brain-computer interface (BCI)?," *Neurosci. Lett.*, vol. 462, no. 1, pp. 94–98, 2009.
- [12] J. Meng and B. He, "Exploring Training Effect in 42 Human Subjects Using a Non-invasive Sensorimotor Rhythm Based Online BCI," *Front. Hum. Neurosci.*, vol. 13, Apr. 2019.
- [13] M. H. Lee *et al.*, "EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy," *Gigascience*, vol. 8, no. 5, pp. 1–16, 2019.
- [14] J. R. Stieger, *et al.*, "Mindfulness Improves Brain-Computer Interface Performance by Increasing Control over Neural Activity in the Alpha Band," *Cereb. Cortex*, vol. 31, no. 1, pp. 426–438, 2021.
- [15] G. Schalk *et al.*, "BCI2000: A general-purpose brain-computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, 2004.
- [16] A. Shoeb "Application of machine learning to epileptic seizure onset detection and treatment," *Diss. Massachusetts Inst. Technol.*, pp. 157–162, 2009.
- [17] H. Dose *et al.*, "An end-to-end deep learning approach to MI-EEG signal classification for BCIs," *Expert Syst. Appl.*, vol. 114, pp. 532–542, Dec. 2018.

Procesamiento de TAC de tórax para su uso en simulaciones de tomosíntesis

A. Galán^{1,2}, P.M. Berdón^{1,2}, J.L. Manuel-Cano¹, J. G. Blas³, M. Desco^{1,2,4,5}, M. Abella^{1,2,4}

¹ Departamento de Bioingeniería e Ingeniería Aeroespacial, Universidad Carlos III de Madrid, Madrid, España

² Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, España

³ Departamento de Informática, Universidad Carlos III de Madrid, Madrid, España

⁴ Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), Madrid, España

⁵ Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Madrid, España

Resumen

La tomosíntesis de tórax permite un diagnóstico más preciso de las alteraciones de las estructuras torácicas que la radiografía, manteniendo una dosis de radiación inferior a la involucrada en Tomografía Axial Computerizada (TAC). Al igual que en otras modalidades de imagen médica, están surgiendo enfoques basados en el aprendizaje profundo para mejorar la calidad de las imágenes de tomosíntesis. Para obtener los datos necesarios para entrenar los modelos, una posibilidad es simular adquisiciones de tomosíntesis a partir de imágenes de TAC, al abundar los datos de dicha modalidad para tórax. Sin embargo, la limitación espacial de los TAC torácicos a la zona de pulmón, resulta en un artefacto de truncamiento en las simulaciones de tomosíntesis. En este trabajo se propone un método de procesado de los estudios de TAC torácico para reducir el efecto de truncamiento.

1. Introducción

La tomosíntesis de tórax ha demostrado tener utilidad clínica, especialmente en la detección de cáncer de pulmón [1, 2]. En esta modalidad se adquieren una serie de proyecciones en un rango angular limitado, obteniéndose imágenes pseudotomográficas que permiten una mejor identificación de las alteraciones de las estructuras torácicas que la radiografía. Si bien las imágenes no ofrecen tanta información como el TAC, al tener un grosor de corte mayor y no ser cuantitativas, la dosis recibida por el paciente es mucho menor que en el primero.

Recientemente se han propuesto algoritmos basados en aprendizaje profundo para mejorar la calidad de las imágenes de tomosíntesis de tórax. Lee et al. [3] entrenaron una red con arquitectura U-net para reducir el ruido y los artefactos causados por la limitación en el número de proyecciones. Para ello utilizaron simulaciones a partir de TAC torácico y probaron la validez del método con adquisiciones de un maniquí de tórax. Gomi et al. [4] entrenaron una red neuronal convolucional profunda en el espacio de la proyección para mejorar el contraste de pequeños nódulos y reducir artefactos 'ripple'. El conjunto de datos estuvo formado por proyecciones de un maniquí de tórax, introduciendo pequeñas rotaciones y translaciones para incrementar la variabilidad y el tamaño del mismo. En ninguno de los dos casos se probaron los modelos en pacientes o bien en un maniquí que incluyera

una mayor extensión anatómica. Por otra parte, sería interesante entrenar modelos de este tipo para solucionar el problema de la radiación dispersa, como Roser et al. propusieron para TAC con geometría cónica [5].

Dada la falta de datos de tomosíntesis y TAC para un mismo paciente, la opción de simular adquisiciones de tomosíntesis a partir de TAC torácico planteada en [3] resulta atractiva, más aún cuando existen datos revisados y de libre acceso de TAC torácico [6-8]. Sin embargo, no tenemos constancia de que se haya estudiado la validez de estas simulaciones, las cuales se ven afectadas por un artefacto de truncamiento como resultado de la limitación espacial de los TAC. En este trabajo se propone un método de procesamiento de imágenes de TAC torácico para reducir el efecto de truncamiento y se evalúa su validez para la simulación de datos de tomosíntesis de tórax.

2. Método

En tomosíntesis de tórax, la fuente se mueve siguiendo una trayectoria lineal, de arriba a abajo, mientras el detector se mantiene fijo (Figura 1). Cada proyección tiene un campo de visión distinto, que viene dado por la posición de la fuente y las dimensiones del detector.

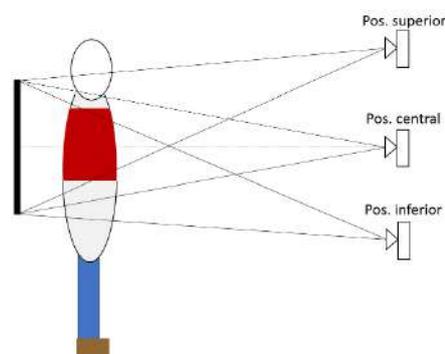


Figura 1. Esquema de la geometría de una adquisición de tomosíntesis de tórax. La parte correspondiente al tórax (en rojo) ocupa sólo parte del campo de visión (delimitado por las líneas negras) en las tres posiciones mostradas de la fuente.

Como se puede observar en la Figura 2, en la simulación de proyecciones de tomosíntesis a partir de un TAC de tórax, algunos de los rayos no atraviesan parte del tejido que sí atravesarían en una adquisición real (líneas

discontinuas en la Figura 2). Este hecho provoca que los valores de atenuación en ciertas zonas del detector, las indicadas en azul en la Figura 2, se vean infraestimados, produciendo un artefacto de truncamiento con forma de óvalo.

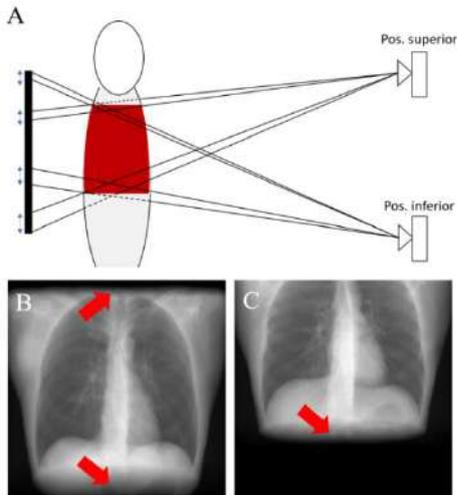


Figura 2. Esquema de la geometría de una adquisición de tomosíntesis de tórax (A). Ejemplos de artefacto de truncamiento en una proyección simulada con la fuente en la posición superior (B) y en la posición inferior (C). Las flechas rojas señalan las zonas con artefacto en B y C.

Por otro lado, la Figura 3 muestra el mapa de radiación obtenido a partir de un TAC toraco-abdominal sin artefacto de truncamiento y otro obtenido a partir de un TAC de tórax, donde se aprecia dicho artefacto, en este caso en la parte inferior. La presencia de dicho artefacto en los datos de entrenamiento de una red neuronal profunda puede introducir un sesgo en los resultados obtenidos, ya que dichos datos no son representativos del caso real, en el que no existe truncamiento. Así, la red aprende a producir resultados no realistas con componentes de alta frecuencia en la zona del artefacto.

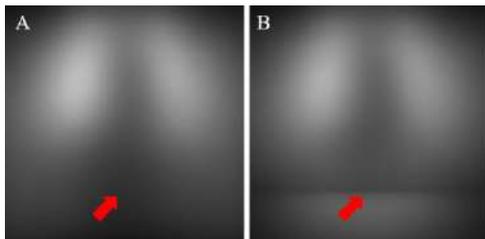


Figura 3. Mapa de radiación dispersa sin artefacto de truncamiento (A), mapa de radiación con artefacto de truncamiento (B). Las flechas señalan la posición del artefacto.

De la misma forma, en el dominio de la reconstrucción también puede existir un sesgo al entrenar con datos con artefacto de truncamiento. Por ejemplo, si se entrena la red para obtener el volumen del TAC a partir de la reconstrucción de tomosíntesis, esta aprenderá a compensar las zonas de baja atenuación causadas por dicho artefacto. Para paliar el efecto del truncamiento, se propone un método para la extensión del volumen del TAC de forma que ocupe la totalidad del campo de visión en todas las proyecciones.

2.1. Métodos de extensión de los volúmenes

En una adquisición de tomosíntesis de tórax, a diferencia de lo que ocurre en TAC, la espalda del paciente está pegada al bucky, por lo que el primer paso es recortar el TAC para eliminar el aire en la parte posterior del paciente. Con dicho recorte se consigue eliminar la cama, pero no el apoyo de cuello para el paciente. Para segmentar este soporte se utiliza el algoritmo Connected Threshold Grower y a la región segmentada se le asigna el valor del aire en unidades Hounsfield (-1000). Posteriormente, se transforman los valores originales a densidades.

El cálculo de la dimensión de la extensión necesaria para que el volumen ocupe la totalidad del campo de visión en todas las proyecciones, que se muestra en la Figura 4, se hace con la siguiente fórmula:

$$ext = \tan(\alpha) \times D \quad (1)$$

donde α y D se calculan como:

$$\alpha = \text{atan}\left(\frac{S-P}{SDD}\right) \quad (2)$$

$$D = DOD + \frac{AP}{2} \quad (3)$$

siendo SDD la distancia de la fuente al detector, DOD la distancia del detector al centro del objeto y AP la dimensión del volumen en el eje anteroposterior. Finalmente, la extensión necesaria es:

$$ext' = P - L + ext \quad (4)$$

siendo L la parte del volumen situada por debajo del rayo central que va de la fuente al centro del detector. El cálculo de la extensión necesaria en la parte superior es análogo al anterior, tomando la posición inicial de la fuente en lugar de la final.

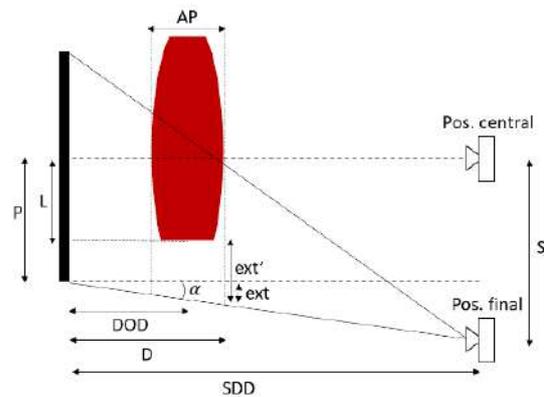


Figura 4. Esquema del cálculo de la extensión de los volúmenes.

El TAC extendido se genera repitiendo la primera y la última rodaja hasta completar la extensión calculada en la parte superior y en la parte inferior, respectivamente.

2.2. Evaluación

El método propuesto se ha evaluado para dos escenarios: reconstrucción de imagen y reducción de radiación dispersa. Para la evaluación se ha utilizado un TAC toraco-abdominal de la base de datos NBIA [8], con un tamaño de $535 \times 535 \times 682$ y un tamaño de voxel isotrópico de 0.748

mm, del cual hemos tomado 455 rodajas axiales correspondientes a la zona del tórax. La Figura 5 muestra el volumen original (TAC toraco-abdominal), el volumen cortado (TAC torácico) y el volumen extendido con el método propuesto (TAC torácico extendido). Si bien las diferencias entre el volumen extendido y el original son apreciables, la adición de las rodajas en la parte superior e inferior pretende compensar la falta de datos de las estructuras anexas al tórax en el TAC torácico, que por la geometría de esta modalidad influyen en la reconstrucción del mismo.

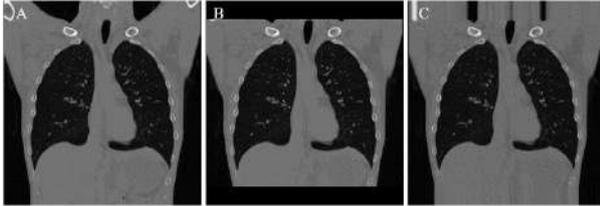


Figura 5. Ejemplo de una de las rodajas coronales del TAC toraco-abdominal (A) del TAC torácico (B) y del TAC torácico extendido (C).

Para simular las proyecciones de tomosíntesis de cada volumen se ha usado FuxSim [9], un software que permite simular diferentes geometrías de adquisición a partir de un mapa de densidad y unos parámetros del sistema. La simulación se ha basado en un protocolo estándar de tomosíntesis de tórax: 61 proyecciones con una distancia total $SOD + DOD$ de 1800 mm y un ángulo de barrido de 30 grados. Para las simulaciones se ha usado $DOD = 164.118$ mm. Las proyecciones tienen un tamaño de 1072×1072 píxeles con un tamaño de píxel de 0.4×0.4 mm.

Para ver el efecto en reconstrucción, se ha reconstruido la zona del tórax a partir de los tres conjuntos de proyecciones utilizando un algoritmo de retroproyección filtrada con compensación del artefacto de truncamiento en proyección debido a la limitación espacial del detector. La evaluación se ha basado en los perfiles sobre los cortes coronales y axiales que se muestran en la Figura 6, obteniéndose el NRMSE de los mismos.

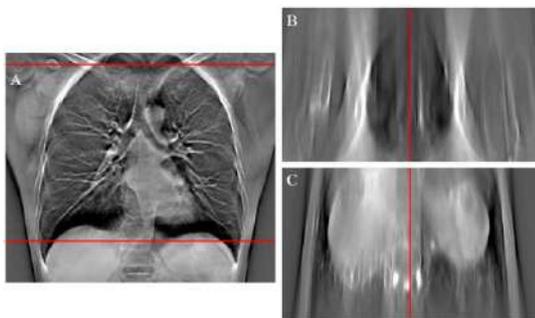


Figura 6. Corte coronal (A), axial superior (B) y axial inferior (C) del volumen reconstruido correspondiente al TAC toraco-abdominal. Las líneas rojas en A muestran la posición de los cortes axiales.

Para ver el efecto en el mapa de radiación dispersa, se ha utilizado el simulador MC-GPU [10], utilizando el mismo protocolo y parámetros geométricos que en las simulaciones con FuxSim salvo por el tamaño de las

proyecciones (4288×4288 píxeles con un tamaño de píxel de 0.1×0.1 mm). Las proyecciones han sido simuladas con 10^{11} fotones muestreados a partir del espectro del aluminio a 120kVp. El volumen ha sido segmentado en tres materiales: aire, tejido blando y hueso, con densidades de 0.001205, 1.00 y 1.92 g/cm^3 respectivamente. Se ha simulado exclusivamente la proyección central de cada volumen y se ha calculado el NRMSE de las simulaciones correspondientes al TAC torácico y al TAC torácico extendido.

3. Resultados

La Figura 7 muestra los tres volúmenes reconstruidos. Se observan zonas con alto brillo en la imagen correspondiente al TAC torácico debidas al filtro rampa en el proceso de reconstrucción. Debido al corte de la proyección por la limitación espacial del TAC torácico, los bordes del óvalo se ven realzados, provocando un aumento de los valores en las zonas involucradas. Como se puede observar en C, el método propuesto permite eliminar este efecto.



Figura 7. Cortes coronales del volumen reconstruido correspondiente al TAC toraco-abdominal (A) TAC torácico (B) y TAC torácico extendido con el método propuesto (C). Las flechas rojas señalan las zonas afectadas por el artefacto de truncamiento.

La Figura 8 muestra los perfiles de intensidad en la parte superior del volumen. Se observa que el método propuesto, más allá de eliminar el efecto de brillo mencionado anteriormente, es capaz de recuperar los valores correspondientes al volumen toraco-abdominal tanto en el corte coronal como en el axial.

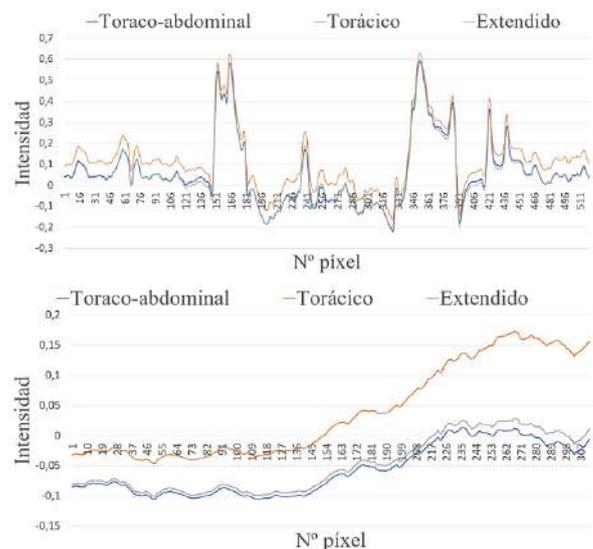


Figura 8. Perfiles de intensidad en el corte coronal para la parte superior (arriba). Perfiles de intensidad en el corte axial para la parte superior (abajo).

La Figura 9 muestra resultados similares a los anteriores en la parte inferior de los volúmenes. Con el método propuesto se ha obtenido en los perfiles un NRMSE de un 8.7 % en el peor caso examinado, aproximadamente un orden de magnitud inferior al obtenido para el TAC torácico.

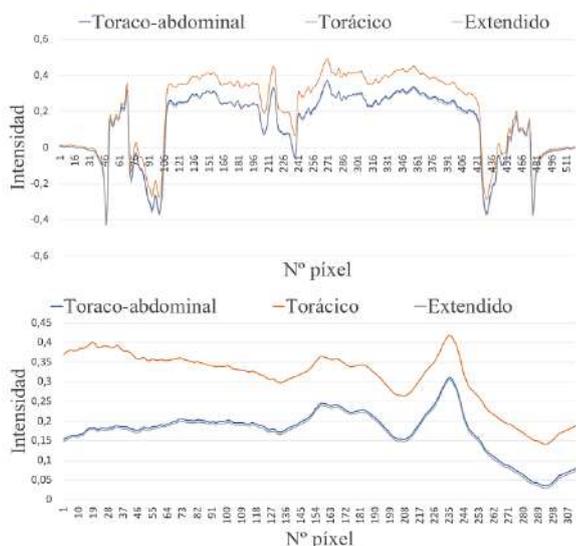


Figura 9. Perfiles de intensidad en el corte coronal para la parte inferior (arriba). Perfiles de intensidad en el corte axial para la parte inferior (abajo).

En la Figura 10 se muestran los mapas de radiación dispersa para los tres volúmenes simulados. Se observa que el artefacto de truncamiento ha sido eliminado en la simulación correspondiente al volumen torácico extendido.

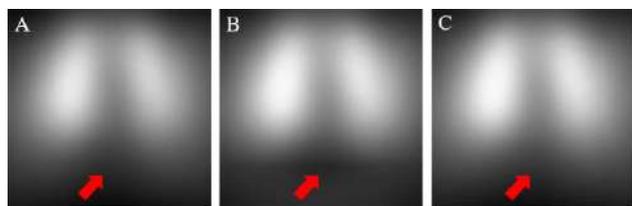


Figura 10. Mapas de radiación dispersa simulados a partir del TAC toraco-abdominal (A), TAC torácico (B) y TAC torácico extendido (C). Las flechas señalan la posición del artefacto de truncamiento.

Los valores de NRMSE calculados a partir de los mapas de radiación dispersa muestran una ligera disminución del error en la simulación del TAC torácico extendido (NRMSE = 3.92 %) con respecto al error en la simulación del TAC torácico (NRMSE = 5.37 %).

4. Discusión

En este trabajo se ha propuesto un método para reducir el artefacto de truncamiento presente en simulaciones de tomosíntesis hechas a partir de TAC torácicos.

La evaluación preliminar en un TAC toraco-abdominal muestra una recuperación de los valores del tórax en reconstrucción y una corrección del artefacto de truncamiento en el mapa de radiación dispersa.

Como trabajo futuro, se evaluará el impacto de esta mejora en el funcionamiento de algoritmos basados en aprendizaje

profundo, al mejorar los datos usados para el entrenamiento de los modelos. Se espera que el método propuesto permita evitar sesgos con origen en el artefacto de truncamiento debido a la limitación espacial de los TAC de tórax.

Agradecimientos

Este trabajo ha sido financiado por el Programa de apoyo a la realización de proyectos interdisciplinarios de I+D para jóvenes investigadores de la Universidad Carlos III de Madrid 2019-2020 en el marco del Convenio Plurianual Comunidad de Madrid- Universidad Carlos III de Madrid (proyecto DEEPCT-CM-UC3M) y por CRUE Universidades, CSIC y el Banco Santander (Fondo Supera, proyecto RADCOV19). El CNIC está financiado por el Ministerio de Ciencia, Innovación y Universidades y la fundación PRO-CNIC y es un centro de excelencia Severo Ochoa (SEV-2015-0505).

Referencias

- [1] Dobbins JT, McAdams HP, Sabol JM. Multi-institutional evaluation of Digital Tomosynthesis, Dual-energy radiography, and conventional chest radiography for the Detection and Management of Pulmonary nodules. *Radiology*, vol 282, sup 1, 2017.
- [2] Söderman C, Johnsson ÅA, Vikgren J. Detection of Pulmonary Nodule Growth with Chest Tomosynthesis: A Human Observer Study Using Simulated Nodules. *Academic Radiology*, vol 26, sup 4, 2019, pp 508-18.
- [3] Lee D, Kim H-J. Restoration of Full Data from Sparse Data in Low-Dose Chest Digital Tomosynthesis Using Deep Convolutional Neural Networks. *Journal of Digital Imaging*, vol 32, sup 2019, pp 489-98.
- [4] Gomi T, Hara H, Watanabe Y. Improved digital chest tomosynthesis image quality by use of a projection-based dual energy virtual monochromatic convolutional neural network with super resolution. *PLOS ONE*, vol 15, sup 12, 2020.
- [5] Roser P, Birkhold A, Strobel N. Feasibility of CBCT Scatter Estimation Using Trivariate Deep Splines. *16th International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, Lovaina, 2021.
- [6] Página web de The Cancer Imaging Archive. <https://public.cancerimagingarchive.net/nbia-search/> (Consultado septiembre 2021).
- [7] Página web de Medical Imaging and Data Resource Centre. <https://data.midrc.org/explorer> (Consultado septiembre 2021).
- [8] Página web de National Biomedical Imaging Archive. <https://imaging.nci.nih.gov/nbia-search/> (Consultado septiembre 2021).
- [9] Abella M, Serrano E, Blas JG-. FUX-Sim: Implementation of a fast universal simulation/reconstruction framework for Xray systems. *PLoS ONE*, vol 12, sup 7, 2017.
- [10] Badal A, Badano A. Accelerating Monte Carlo simulations of photon transport in a voxelized geometry using a massively parallel graphics processing unit. *Medical Physics*, vol 36, sup 11, 2009, pp 4878-80.

Electrophysiological modelling of left ventricle with heterogeneous scar to assess inducibility of reentry circuits

C. Sales Bellés¹, B. Jauregui Garrido², D. Soto Iglesias², A. Berruezo Sánchez², R. Sebastián Aguilar³, O. Camara Rey¹

¹ Universitat Pompeu Fabra, Barcelona, España, clara.sales01@estudiant.upf.edu, oscar.camara@upf.edu

² Heart Institute, Teknon Medical Center, Barcelona, España, {[beatriz.jauregui](mailto:beatriz.jauregui@quironsalud.es), [antonio.berruezo](mailto:antonio.berruezo@quironsalud.es)}@quironsalud.es, sotin1985@gmail.com

³ Universitat de València, Valencia, España, rafael.sebastian@uv.es

Abstract

Modelling and simulation of the heart in patients that have suffered a myocardial infarction could help in arrhythmia risk stratification and therapy planning of radiofrequency ablation. Computational simulations can cast predictions about the risk of a given myocardial infarct to develop ventricular tachycardia (VT) associated to conduction channels through the infarcted zone. In this work, we have built a patient-specific LV model that includes the scar region segmented from cardiac magnetic resonance imaging (cMRI). It corresponds to an individual who has not suffered VT, but has a high chance to develop it. Cellular and tissue heterogeneity as well as remodelling in the infarcted zone have been considered in the models. The aim of the study is to determine if this patient is VT inducible by virtually applying the same pace-mapping stimulation protocol that it is commonly performed in a clinical context. The simulation study revealed large action potential duration (APD) heterogeneity in the whole LV model and slow electrical conduction in the infarcted zone. However, no reentry circuits were obtained suggesting that at the moment this patient is not VT inducible with its actual scar configuration.

1. Introduction

Myocardial infarction (MI) is one of the most common causes of cardiovascular disease worldwide [1]. The heterogeneous scar produced by a MI is formed by a border zone (BZ), which contains myocytes and fibrosis, and a core zone (CZ) where fibrosis predominates over cellularity. Indeed, the presence of heterogeneous scar in the LV is considered a highly potential precursor to generate a reentry circuit along the slow conducting channels (CCs) made of BZ tissue that crosses the CZ. This arrhythmogenic electrical behaviour could lead to a potentially lethal VT [2].

Nowadays, clinical cardiologists use computed tomography (CT) and magnetic resonance imaging (MRI) for evaluation of both heart's structure and function, enabling the characterization of the infarcted area and CCs [3]. Radiofrequency ablation (RFA) is a procedure where a part of the heart with abnormal electrical activity is burnt aiming to prevent potential arrhythmias. In fact, RFA is a common practice in patients with heterogeneous scar that are suffering from ventricular arrhythmias. Its main goal is to isolate the CCs crossing the scar that could lead to reentry circuits associated with VT.

Different works have reproduced in-silico clinical VTs by virtually constructing the image-based patient heart and

simulating the specific conditions that produce the reentry circuit. For example, López et al. [4] used delayed enhancement (DE)-MRI for constructing a patient-specific 3D model, including the infarcted area and a detailed electrophysiological cellular modelling, that was able to reproduce the clinical VT. Cedilnik et al. [5] used CT and Eikonal models for the same purpose. These works showed the potential of cardiac computational simulations for improving patient stratification by predicting the appearance of the reentry circuits on patients with MI. In addition, being able to predict the reentry and exit point or points in the CCs that sustain the VT can considerably improve the outcome of RFA.

Therefore, the aim of this work is to construct the cMRI-based patient-specific model for an individual that has suffered MI but has not showed any VT episode yet, and virtually simulate the clinical protocol to assess whether the scar configuration is VT inducible or not.

2. Methodology

2.1. Clinical data and preprocessing

In this work, a patient-specific LV model with heterogeneous scar regions has been constructed from segmented cMRI. This data, provided by *Centro Médico Teknon* (Barcelona), was part of a dataset of patients who have suffered a MI but no VT episodes were observed, although being considered at high risk of developing it in the future. cMRI was acquired using a MRI scanner Ingenia 1.5 (Philips) and had an isotropic resolution of 1 mm. Epicardial and endocardial LV surfaces were manually segmented from cMRI using ADAS 3D (www.adas3d.com) which gave rise to smooth triangular surfaces meshes. The BZ and CZ triangular surfaces were automatically obtained from ADAS 3D by exploiting the voxel intensity information encoded in the MRI. Specifically, the BZ was defined with a probabilistic similarity index between 40-70% of the maximum voxel intensity. The patient presented a 29.70 g of infarcted zone mass, being 25.74 g BZ and 3.96 g CZ, and 14.94 g of CCs mass.

2.2. Patient-specific model construction

LV computational hexahedral-based volume mesh was constructed from epi and endocardial triangular surface mesh using MeshGems-Hexa software. The hexahedral-based volume mesh was comprised by 1.50 million nodes

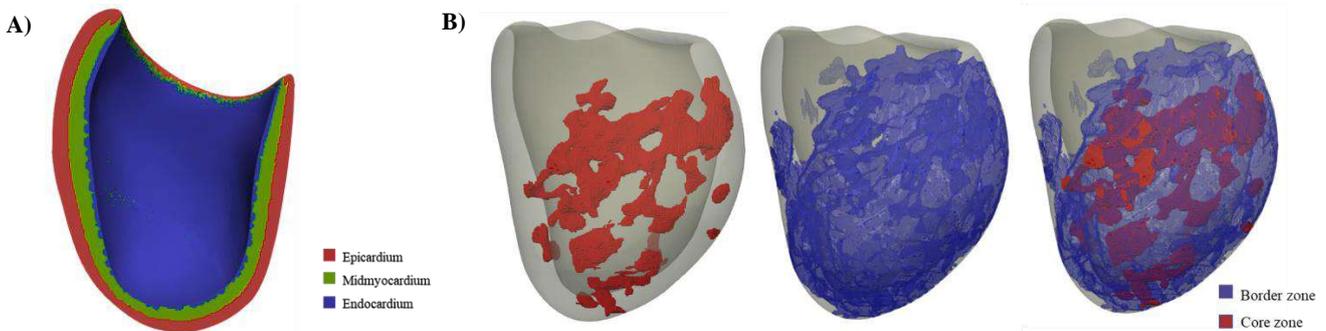


Figure 1. (A) Anterior view of the transmural heterogeneity of the LV hexahedral-based volume model of the myocardium. (B) Lateral view of the specific geometry of the LV hexahedral-based volume mesh considering the border zone and core zone.

and 1.39 million elements. The average edge length was 0.4 mm.

Based on the work of López et al. [4], cardiac transmural heterogeneity was included and the LV wall thickness was represented by 17%, 41% and 42% for the endo, mid and epicardium, respectively. A custom MatLab code, based on these percentages, was used to label and define specific ion models for each layer (see Figure 1A). *Inpolyhedron* [6] MatLab function was used to map the BZ and CZ surface meshes into the volume mesh (see Figure 1B) in order to set up patient-specific properties for healthy tissue and infarcted tissue. Only connected regions with more than 200 elements were kept from BZ and CZ surfaces meshes because smaller ones were considered noise.

Electrophysiological modelling was defined differently depending on the cardiac region and scar presence based on the work of López et al. [4]. At cellular level, ten Tusscher model of human ventricular action potential (AP) considering transmural heterogeneity was used as detailed electrophysiological model [7]. Moreover, to reproduce the BZ electrical remodelling the maximum conductivities of specific ionic channels (g_{Na} , g_{CaL} , g_{Kr} and g_{Ks}) were reduced by 28%, 31%, 30% and 20%, respectively. At tissue level, a monodomain approach was considered where the conductivity depends on the fibre orientation. Healthy tissue conductivities were set to 0.24 S/m in the longitudinal and 0.0456 S/m in the transversal direction, and BZ tissue conductivities were set to 0.05 S/m and 0.01 S/m, respectively, and considered isotropic. CZ was considered as dead tissue and its conductivity was set to 0 S/m representing no propagation in any direction. To include the anisotropy in the cardiac muscle, fibre orientation was included using the Streeter's rule-based method described by the equations defined in the work of Sebastian et al. [8].

2.3. Electrophysiology simulation

ELVIRA software was used to perform finite element method-based simulations and solve the monodomain reaction-diffusion equations for cardiac electrophysiology. The conjugate gradient method was the solver used, with a time step of 0.02 ms, and spatial discretization of 0.4 mm. Simulations were run on a 48 cores machine (2.387 GHz) with 86 GB of RAM. Simulating 1000 ms took approximately 2 hours.

2.4. Stimulation protocol

The stimulation protocol considers the pacing sites and stimulation frequency. It was defined following the standard stimulation protocol in any electrophysiology laboratory. Regarding pacing sites, their locations are at the apical-septal endocardial right ventricular wall, which was located in the external side of the septal wall in our LV model. To locate the pacing site, the biventricular model constructed in the work of López et al. [4] was used as a reference.

Regarding the stimulation pulses, the LV patient-specific model was first stabilized by a train of 7 beats with a basic cycle length (CL) of 430 ms aiming to get a steady state of the complete model, that is, when electrotonic currents adjust and balance themselves to a fast rhythm. Then, a second set of three fast stimuli were coupled aiming to induce the reentry circuit. Different simulations were performed by reducing the coupling interval (CI) in steps of 20 ms after the last stabilization pulse. This protocol was followed until either a VT was generated, propagation blocks at the pacing site or CI arrives to 200 ms. Coupled stimuli aimed to produce a unidirectional block in the infarcted region that leads to VT.

3. Results and discussion

3.1. Activation time maps

The patient-specific model showed a complete LV activation (including BZ) after 240 ms from virtual catheter stimulation. As expected, the depolarization wavefront did not propagate through the regions labelled as CZ. The other electrophysiological regions did propagate the electrical impulse but with some differences, as each zone had specific properties. Moreover, slower conduction was found in presence of BZ showing a complex activation pattern (see Figure 2).

3.2. APD heterogeneity

Six electrophysiological ionic models were taken into account corresponding to epicardium, midmyocardium and endocardium, considering healthy tissue and BZ remodelling. This fact introduced differences in the AP depending on the studied region that is translated to heterogeneous APD maps. Moreover, these differences defined at cellular level led to complex APD heterogeneity

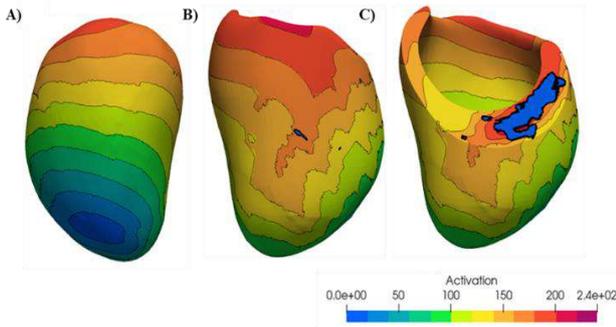


Figure 2. Activation time representation for patient-specific LV model of the stabilization phase seventh beat. (A) shows the LV septal wall, (B) shows LV lateral wall and (C) shows a heart's cut in order to display the presence of the CZ of the infarction. Units are in ms.

maps when considering the whole LV which is a condition required to trigger a reentry circuit [9]. This patient presents a complex APD heterogeneity map due to the large and complex infarcted region (see Figure 3) showing potential to induce reentrant activity.

Midmyocardium cells are defined to present the longest APDs when using the ten Tusscher model, and considering the cardiac transmural heterogeneity. Nevertheless, in the presented LV model, healthy endocardium cells presented also a long APD, which was due to the contact with midmyocardial cells and a very thin endocardial layer (see Figure 3).

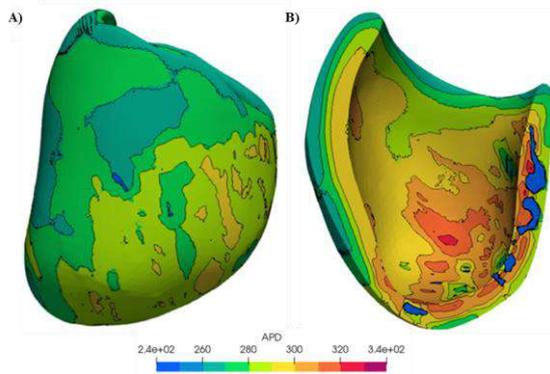


Figure 3. Action potential duration heterogeneity found in (A) lateral view and (B) anterior cut view of the LV model for the seventh stabilization pulse. Blue regions represent the scar which does not activate. Units are in ms.

BZ electrophysiological remodelling presented longer APD than healthy areas due to changes introduced in the potassium currents, which are the ones responsible for repolarization. Therefore, APD heterogeneity was due to the presence of different electrophysiological cell types and remodelling effects, which is a key substrate to trigger and sustain a reentry circuit.

3.3. APD adaptation and block propagation

The defined stimulation protocol increased the heartbeat frequency and reduced the overall APD in the different regions. Several nodes of the 3D model (corresponding to different cell types) were recorded during the stabilization phase showing the highest APD reduction after the second

pulse (see Table 1). The tendency observed was a stabilization around that value considering the seventh pulse result a steady state. APD reduction also occurred in the coupled phase. It decreased for every cell when the following pulse is reduced. This decrease was bigger as the CI was reduced (see Table 2). The coupled stimulus did not trigger the activation when it was set to 320 ms.

This APD adaptation was explained by the well-known restitution curve. The ten Tusscher model included in ELVIRA software is stabilized to a CL of 1 second, but it is able to adapt its AP to other heart rates. This adaptation is reflected in the APD reduction when increasing the heartbeat frequency.

Pulse	1	2	3	4	5	6	7
APD (ms)	305	257	267	266	266	266	267

Table 1. Action potential duration of healthy epicardium cell for the seven beats of the stabilization phase.

CI	400	380	360	340	320
APD (ms)	258	252	246	241	--

Table 2. Action potential duration of healthy epicardium cell reducing at steps of 20 ms the CI.

Nevertheless, this adaptation has a limit related to the minimum CI and the refractory period. If the stimulus is given with a short enough CI and the cells are still in refractory period, it will not trigger a new activation and the propagation block occurs [10].

3.4. Non reentry circuit appearance

This model was stimulated with a coupled pulse that was reduced in steps of 20 ms up to the propagation block was produced at the pacing site. It did not show a reentry circuit for any of the stimulation protocol and virtual catheter locations. A slower propagation was observed in this patient in zones located in the BZ suggesting the possibility of a reentry appearance in this region. Nevertheless, this was not produced for any case as the wavefront was able to activate the whole cardiac tissue (no unidirectional blocks were observed) and any region appeared viable (see Figure 4).

López et al. [4] generate DE-MRI based patient-specific biventricular model defining both healthy, BZ and CZ regions and cardiac transmural heterogeneity. They performed detailed electrophysiological simulations and were able to in-silico obtain the clinical VT reproducing the pacing protocol used in the electrophysiology laboratory. This work shows the power of image-based cardiac models together with detailed electrophysiological simulations to in-silico induce a reentry circuit.

In this work, we followed a similar methodology trying to induce reentrant activity in a patient with MI, considered at high risk of developing VT. The presence of the heterogeneous scar was the principal substrate for a reentry circuit to appear in this patient as this anatomical obstacle has been considered a highly potential precursor of arrhythmias [2]. However, not every MI generates reentry

circuits as specific CCs spatial configuration is needed. Therefore, one of the potential reasons for not having induced a reentry could be the infarcted zone configuration as this patient presented with CCs and not large scar (see Figure 1B).

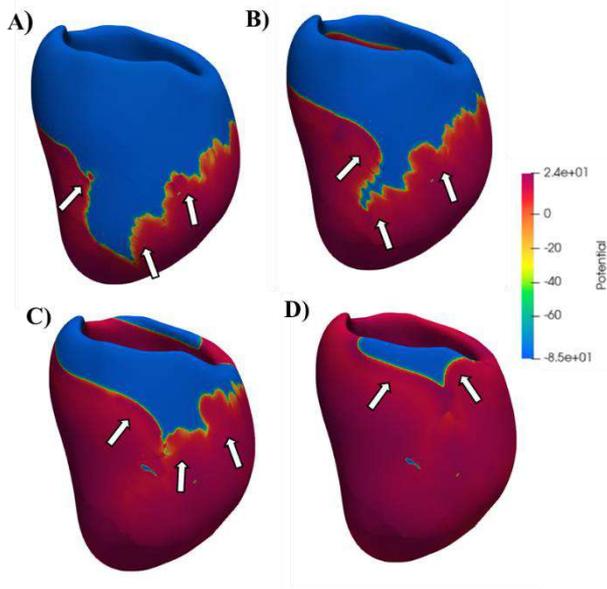


Figure 4. Lateral LV view of wavefront propagation from third coupled pulse given at 340 ms. (A) 180 ms, (B) 200 ms, (C) 220 ms and (D) 240 ms after the third coupled pulse is given at 340 ms. Units are in mV.

4. Conclusions and future work

The presented work aimed to construct the cMRI-based patient-specific LV model including specific anatomical characteristics and electrophysiological properties for each cardiac region defined. Moreover, the model was able to reproduce APD heterogeneity due to the inclusion of different electrophysiological regions and APD reduction effect related to high-frequency stimulation. Nevertheless, this patient did not show reentrant activity suggesting it is not VT inducible.

Despite the infarcted zone configuration is the most probable reason for reentries non-appearance in our work, some approaches done in the methodology could explain this fact. Firstly, just one pacing site has been performed to induce the VT. The same stimulation protocol should be considered from different pacing sites to force the reentry to appear, just in case the patient is inducible. Moreover, despite stabilization phase was done using seven pulses, following the clinical protocol, it could be that the APD obtained was not stabilized, that means, not short enough, to be able the reentry circuit to appear. Following this idea, also simulated AP could present a considerable different shape from real patient ones as the presence of MI could have remodelled the passive electrical properties of the myocardium [11].

References

- [1] E. M. Aliot *et al.*, "EHRA/HRS expert consensus on catheter ablation of ventricular arrhythmias," *Europace*, vol. 11, no. 6, pp. 771–817, 2009, doi: 10.1093/europace/eup098.
- [2] J. Bhar-Amato, W. Davies, and S. Agarwal, "Ventricular arrhythmia after acute myocardial infarction: 'The perfect storm,'" *Arrhythmia Electrophysiol. Rev.*, vol. 6, no. 3, pp. 134–139, Sep. 2017, doi: 10.15420/aer.2017.24.1.
- [3] P. J. Sheth *et al.*, "Cardiac physiology for radiologists: Review of relevant physiology for interpretation of cardiac MR imaging and CT," *Radiographics*, vol. 35, no. 5, pp. 1335–1351, Sep. 2015, doi: 10.1148/rg.2015140234.
- [4] A. Lopez-Perez, R. Sebastian, M. Izquierdo, R. Ruiz, M. Bishop, and J. M. Ferrero, "Personalized cardiac computational models: From clinical data to simulation of infarct-related ventricular tachycardia," *Front. Physiol.*, vol. 10, no. MAY, 2019, doi: 10.3389/fphys.2019.00580.
- [5] N. Cedilnik, J. Duchateau, R. Dubois, P. Jaïs, H. Cochet, and M. Sermesant, "VT scan: Towards an efficient pipeline from computed tomography images to ventricular tachycardia ablation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10263 LNCS, pp. 271–279, doi: 10.1007/978-3-319-59448-4_26.
- [6] "inpolyhedron - are points inside a triangulated volume? - File Exchange - MATLAB Central." <https://es.mathworks.com/matlabcentral/fileexchange/37856-inpolyhedron-are-points-inside-a-triangulated-volume> (accessed Jun. 11, 2021).
- [7] K. H. W. J. ten Tusscher and A. V. Panfilov, "Alternans and spiral breakup in a human ventricular tissue model," <https://doi.org/10.1152/ajpheart.00109.2006>, vol. 291, no. 3, pp. 1088–1100, 2006, doi: 10.1152/AJPHEART.00109.2006.
- [8] R. Sebastian, V. Zimmerman, F. Sukno, B. B. Bijnens, and A. F. Frangi, "Cardiac modelling for pathophysiology research and clinical applications. The need for an automated pipeline," in *IFMBE Proceedings*, 2009, vol. 25, no. 4, pp. 2207–2210, doi: 10.1007/978-3-642-03882-2_586.
- [9] M. J. Bishop, E. J. Vigmond, and G. Plank, "The functional role of electrophysiological heterogeneity in the rabbit ventricle during rapid pacing and arrhythmias," *Am. J. Physiol. - Hear. Circ. Physiol.*, vol. 304, no. 9, p. H1240, 2013, doi: 10.1152/ajpheart.00894.2012.
- [10] D. Attwell, I. Cohen, and D. A. Eisner, "The effects of heart rate on the action potential of guinea-pig and human ventricular muscle.," *J. Physiol.*, vol. 313, no. 1, pp. 439–461, Apr. 1981, doi: 10.1113/jphysiol.1981.sp013675.
- [11] N. A. Trayanova, P. M. Boyle, H. J. Arevalo, and S. Zahid, "Exploring susceptibility to atrial and ventricular arrhythmias resulting from remodeling of the passive electrical properties in the heart: A simulation approach," *Frontiers in Physiology*, vol. 5, no. Nov. Frontiers Media S.A., p. 435, Nov. 12, 2014, doi: 10.3389/fphys.2014.00435.

Inclusión de compensación de movimiento en un esquema *deep learning* basado en modelo para la reconstrucción de imágenes de cine cardíaco

J. R. Rodríguez Galván¹, E. Martín González¹,
M. T. Sevilla-Ruiz², A. Revilla-Orodea²,
C. Alberola-López¹

¹ Laboratorio de Procesado de Imagen (LPI), Universidad de Valladolid

² Instituto de Ciencias del Corazón (ICICOR). Hospital Clínico Universitario de Valladolid.

Abstract

En este artículo se propone la inclusión de compensación de movimiento en una arquitectura de reconstrucción de imágenes de resonancia magnética basada en aprendizaje supervisado. La arquitectura –conocida como MoDL– no hace uso de redundancia temporal en modalidades de resonancia que involucran movimiento, como el cine cardíaco. Nuestra propuesta introduce esta información y para ello se llevan a cabo las derivaciones analíticas oportunas. Los resultados obtenidos, según las métricas de calidad, son similares para ambas soluciones; sin embargo, las imágenes reconstruidas utilizando la solución de nuestra propuesta reproducen mejor los detalles finos tras un examen subjetivo.

1. Introducción

La Imagen por Resonancia Magnética Nuclear (MRI de *Magnetic Resonance Imaging*) es una prueba de diagnóstico que obtiene imágenes de los órganos internos sin necesidad de utilizar radiación ionizante, a diferencia de los rayos X o la tomografía axial computerizada. Su carácter inocuo y su alta calidad para tejidos *blandos* la convierte en una prueba ampliamente utilizada. El principal inconveniente de la misma es el tiempo que lleva un estudio completo; no son infrecuentes los protocolos de 30 minutos de duración y, teniendo en cuenta la incomodidad del paciente, que en ocasiones debe pasar por prolongados períodos de apnea, y la alta demanda que tienen este tipo de pruebas, la reducción de los tiempos de adquisición es un objetivo central de la comunidad científico-técnica. Las posibilidades para la reducción de este tiempo son, esencialmente, dos: mejorar el hardware o hacer modificaciones en el proceso de adquisición. La adquisición de una MRI se lleva a cabo en el dominio de Fourier (llamado espacio K), el cual se rellena secuencialmente siguiendo diferentes trayectorias (cartesianas, radiales y espirales, fundamentalmente). Cabe destacar que el mayor coste en tiempo de este proceso no se da en el muestreo de cada línea de forma individual, sino a la hora de realizar un cambio de línea [1] por lo que la meta será reducir el número de líneas a muestrear y,

posteriormente, reconstruir la información empleando diferentes técnicas que se explicarán a continuación.

En primer lugar, son conocidas las técnicas englobadas bajo el término *Parallel Imaging* (PI) de entre las que destaca SENSE [2]. Este procedimiento consiste en utilizar la información redundante de las bobinas para, al reducir las líneas del espacio K que se muestrean, evitar los efectos de solapamiento en el dominio de la imagen que se producirían si no hubiese un procesamiento específico. Al poderse modelar el proceso de adquisición mediante expresiones matriciales, la solución se obtiene mediante la resolución de un problema de mínimos cuadrados (optimización ℓ_2).

Avances posteriores han hecho uso de las llamadas técnicas de *Compressed Sensing* (CS) [3] que consisten en un muestreo con un patrón pseudoaleatorio del espacio K y una reconstrucción de la imagen mediante técnicas de optimización, en este caso, $\ell_1 - \ell_2$. Este procedimiento se basa en el hecho de que señales expresables mediante un reducido número de coeficientes en un espacio transformado son recuperables con exactitud siempre que las mediciones de las señales tengan suficiente grado de incoherencia.

Las técnicas de aprendizaje profundo (DL) han entrado también en el terreno de la reconstrucción y han convertido el problema de la reconstrucción en un problema de entrenamiento de una red neuronal convolucional (CNN) a partir de reconstrucciones de baja calidad. En el caso de entrenamientos supervisados se debe disponer también de las reconstrucciones a partir de datos completamente muestreados, los cuales se toman como patrón oro. Algunos ejemplos son [4, 5].

Sin embargo, como la reconstrucción de MRI está muy asociada a un modelo matricial de reconstrucción, se han planteado esquemas que hacen uso combinado de la potencia de ambos métodos, a saber, los basados en modelo y los basados en aprendizaje. La solución se denomina MoDL [6] (*Model-based reconstruction using Deep Learned priors*). Esta solución, si bien matemáticamente elegante y con resultados satisfactorios, en su versión original obvia la información existente entre las diferentes fases

cardíacas para la reconstrucción de MRI de cine cardíaco. Posteriormente se ha descrito una extensión, MoDL con STORM [7], pero esta solución requiere del uso de datos adicionales del escáner.

En este artículo planteamos una solución que hace uso de esta redundancia mediante la incorporación de técnicas de compensación de movimiento a partir, exclusivamente, de los datos que usa MoDL, es decir, sin necesidad de adquisiciones adicionales.

2. Materiales y métodos

2.1. Materiales

La base de datos utilizada consiste en adquisiciones de resonancia magnética de cine cardíaco en 2D con orientación de eje corto *multi-slice* de 89 pacientes, tanto sanos como afectados por miocardiopatía hipertrófica (en la proporción 1/3 y 2/3, respectivamente). Esta base de datos es la utilizada en [8]. Los diferentes cortes han sido extraídos de cada sujeto como secuencias dinámicas 2D independientes. El conjunto de datos se ha separado en tres subconjuntos, entrenamiento, test, y validación, que contienen 575, 199 y 190 secuencias bidimensionales, respectivamente. Se ha tenido especial cuidado en que los datos de un mismo paciente se encuentren en un único conjunto, como forma de evitar sesgos en los resultados. Cada una de las secuencias dinámicas contiene $N = 30$ fases temporales. Cabe destacar que el ciclo cardíaco se divide en N fases cardíacas espaciadas por igual, de modo que la separación entre los *frames* depende del paciente. Las imágenes fueron preprocesadas para conseguir una resolución de 1 mm^2 y luego se recortaron a 160×160 píxeles. Finalmente, cada una de las secuencias dinámicas fue normalizada en intensidad entre 0 y 1.

2.2. Métodos

En esta sección se va a explicar en qué consiste la propuesta del presente artículo. Dado que ésta es una modificación del *framework* MoDL, comenzaremos explicando el funcionamiento de éste para después completarlo con nuestra propuesta.

2.2.1. Model-based deep learning (MoDL)

MoDL [6] parte de la siguiente asunción central

$$\mathbf{b} = \mathcal{A}(\mathbf{x}), \quad (1)$$

donde, en nuestro caso, $\mathcal{A} = \mathbf{S}\mathbf{F}$, donde \mathbf{b} son muestras tomadas del espacio \mathbf{K} , \mathbf{x} la imagen a reconstruir, \mathbf{F} la transformada de Fourier 2-D y \mathbf{S} es una matriz de muestreo que toma líneas de la matriz transformada.

Para forzar una solución consistente, el problema es formulado como uno de optimización [6]

$$\mathbf{x} = \arg \min_{\mathbf{x}} \underbrace{\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2}_{\text{Consistencia de datos}} + \lambda \underbrace{\mathcal{R}(\mathbf{x})}_{\text{Regularización}}, \quad (2)$$

escogiendo el prior de regularización $\mathcal{R}(\mathbf{x})$ cuidadosamente para forzar que las soluciones ocupen el espacio de

imágenes deseadas.

El método propuesto en [6] utiliza como función de regularización de (2)

$$\mathcal{R}(\mathbf{x}) = \|\mathcal{N}_{\mathbf{w}}(\mathbf{x})\|_2^2, \quad (3)$$

siendo $\mathcal{N}_{\mathbf{w}}(\mathbf{x})$ un estimador del ruido de la imagen obtenido mediante los parámetros \mathbf{w} , los cuales se determinan mediante aprendizaje supervisado. La imagen de ruido se expresa

$$\mathcal{N}_{\mathbf{w}}(\mathbf{x}) = (\mathcal{I} - \mathcal{D}_{\mathbf{w}})(\mathbf{x}) = \mathbf{x} - \mathcal{D}_{\mathbf{w}}(\mathbf{x}), \quad (4)$$

donde $\mathcal{D}_{\mathbf{w}}$ representa la estimación de la versión sin ruido de la imagen. La versión final de la ecuación (2) utilizada por MoDL es la siguiente [6]

$$\mathbf{x}_{n+1} = \arg \min_{\mathbf{x}} \|\mathcal{A}(\mathbf{x}) - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x} - \mathbf{z}_n\|_2^2, \quad (5)$$

$$\mathbf{z}_n = \mathcal{D}_{\mathbf{w}}(\mathbf{x}_n), \quad (6)$$

donde el subíndice n , $1 \leq n \leq K$ denota la naturaleza iterativa de este *framework* y K es el número de iteraciones previsto en el modelo.

El problema de la ecuación (5) puede ser resuelto mediante las ecuaciones normales

$$\mathbf{x}_{n+1} = \underbrace{(\mathcal{A}^H \mathcal{A} + \lambda \mathcal{I})^{-1}}_{\mathcal{Q}} (\mathcal{A}^H(\mathbf{b}) + \lambda \mathbf{z}_n) \quad (7)$$

2.2.2. Método propuesto

Nuestra propuesta modifica el problema de reconstrucción (5) añadiendo un nuevo *prior*, esta vez tomando la norma ℓ_1

$$\mathbf{x}_{n+1} = \arg \min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda_1 \|\mathbf{x} - \mathbf{z}_n\|_2^2 + \lambda_2 \|\Phi \mathbf{T}_{\Theta} \mathbf{x}\|_1, \quad (8)$$

donde Φ es el operador de Temporal Total Variation [9] y \mathbf{T}_{Θ} es el registrado grupal de la imagen de cine [8].

Para la resolución de este problema, en primer lugar se aproximará la norma ℓ_1 por la siguiente expresión

$$f_{\mu}(\mathbf{u}) = \begin{cases} \frac{1}{2\mu} \|\mathbf{u}\|_2^2, & \text{si } |\mathbf{u}| \leq \mu, \\ \|\mathbf{u}\|_1 - \frac{\mu}{2}, & \text{en otro caso,} \end{cases} \quad (9)$$

siendo μ un parámetro de suavizado de la norma ℓ_1 en las proximidades del cero. El gradiente de la función, cuando el argumento es el vector $\mathbf{U}\mathbf{u}$ (con $\mathbf{U} = \Phi \mathbf{T}_{\Theta}$)

$$\nabla f_{\mu}(\mathbf{U}\mathbf{u}) = \begin{cases} \frac{1}{\mu} \mathbf{U}^H \mathbf{U} \mathbf{u}, & \text{si } |\mathbf{U}\mathbf{u}_{\mathbf{k}}| \leq \mu, \\ \mathbf{U}^H - \text{sgn}(\mathbf{U}\mathbf{u}), & \text{en otro caso.} \end{cases} \quad (10)$$

En este desarrollo la función $\text{sgn}(\cdot)$ aplicada al vector \mathbf{u} significa el apilamiento de funciones escalares $\frac{\mathbf{u}_i}{|\mathbf{u}_i|}$, con \mathbf{u}_i la componente i -ésima del vector. La solución en este caso pasaría a ser

$$(\mathbf{A}^H \mathbf{A} + \lambda_1 \mathbf{I})\mathbf{x} + \lambda_2 \nabla f_{\mu}(\Phi \mathbf{T}_{\Theta} \mathbf{x}) = \mathbf{A}\mathbf{b} + \lambda_1 \mathbf{z}_n, \quad (11)$$

de forma que, para valores en los que $|\Phi \mathbf{T}_\Theta \mathbf{x}| \leq \mu$ podemos proceder

$$\mathbf{J}_z(\mathbf{x}) = \left(\mathbf{A}^H \mathbf{A} + \lambda_1 \mathbf{I} + \frac{\lambda_2}{\mu} \mathbf{T}_\Theta^H \Phi^H \Phi \mathbf{T}_\Theta \right)^{-1} \quad (12)$$

Sin embargo, para el caso en que $|\Phi \mathbf{T}_\Theta \mathbf{x}| > \mu$, dado que el gradiente depende del propio punto en el que evaluamos, procederíamos de la forma

$$\begin{aligned} (\mathbf{A}^H \mathbf{A} + \lambda_1 \mathbf{I}) \mathbf{x}_n + \lambda_2 \nabla f_\mu(\Phi \mathbf{T}_\Theta \mathbf{x}_{n-1}) &= \\ = \mathbf{A} \mathbf{b} + \lambda_1 \mathbf{z}_n & \end{aligned} \quad (13)$$

obteniendo finalmente

$$\begin{aligned} \mathbf{x}_n = (\mathbf{A}^H \mathbf{A} + \lambda_1 \mathbf{I})^{-1} (\mathbf{A} \mathbf{b} + \lambda_1 \mathbf{z}_n - \\ - \lambda_2 \mathbf{T}_\Theta^H \Phi^H \text{sgn}(\Phi \mathbf{T}_\Theta \mathbf{x}_{n-1})) \end{aligned} \quad (14)$$

3. Resultados

Presentamos resultados de reconstrucción de imagen cardíaca de cine con un factor de submuestreo igual a diez. Se define este parámetro como el número de líneas que conforman el espacio K completamente muestreado dividido por el número de líneas realmente muestreado. Para tal fin, las imágenes se han submuestreado retrospectivamente mediante un patrón pseudo-aleatorio.

Para la comparación de los resultados, aparte de la calidad subjetiva de las imágenes, se calcularán tres métricas con respecto a la imagen completamente muestreada: PSNR (Relación Señal a Ruido Pico), SSIM (Índice de similitud estructural) [10] y la raíz cuadrada del error cuadrático medio normalizado (NRMSE).

Los modelos comparados utilizarán valores del parámetro K , iguales a 1 y 3. El número de épocas (que es el número de veces que pasan todos los bloques de datos de entrenamiento por la red) será de 100 y 50 para los K anteriormente mencionados. Las figuras 1, 2 y 3 muestran los resultados de la reconstrucción para $K = 3$, habiendo resaltado las diferencias encontradas con flechas. La tabla 1 muestra el valor medio de las métricas calculadas para ambas arquitecturas para ambos valores de K .

Modelo	PSNR	SSIM	NRMSE
Submuestreado	19,63	0,6033	0,5037
MoDL 1K	22,38	0,7349	0,3618
MoDL MC 1K	22,69	0,7335	0,3483
MoDL 3K	23,69	0,7783	0,3097
MoDL MC 3K	23,68	0,7430	0,3081

Tabla 1. Resultados medios obtenidos para las diferentes métricas reconstruyendo modelos entrenados.

4. Discusión

La Tabla 1 indica que las diferencias entre ambos esquemas son sutiles y los resultados numéricos son comparables. Por este motivo, el juicio subjetivo en cuanto a la calidad de la reconstrucción es importante y, en particular, la

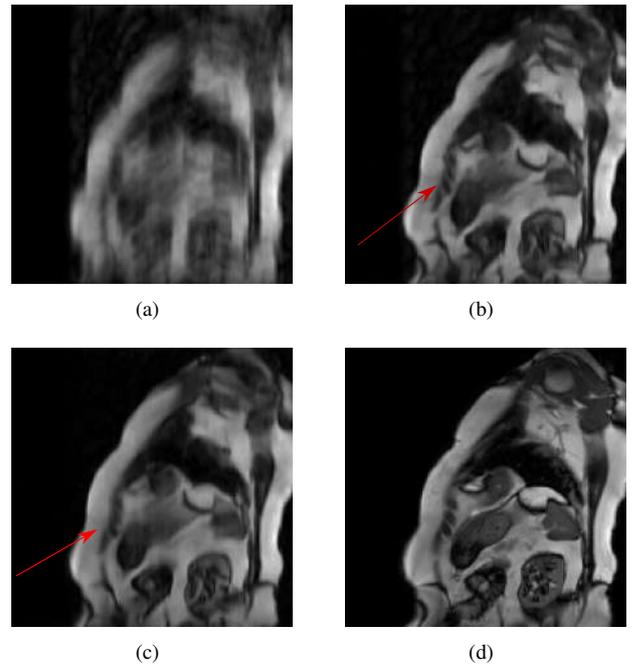


Figura 1. Reconstrucción para distintos modelos con $K = 3$. (a) Datos submuestreados, (b) MoDLMC, (c) MoDL, (d) datos originales.

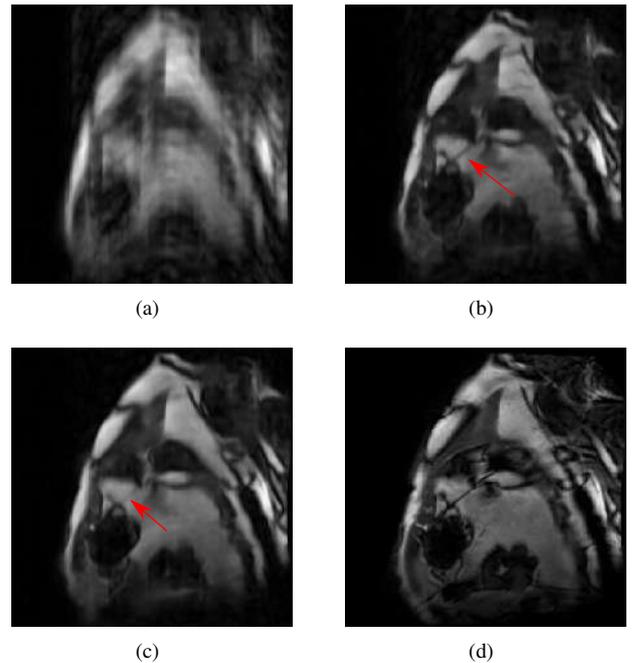


Figura 2. Reconstrucción para distintos modelos con $K = 3$. (a) Datos submuestreados, (b) MoDLMC, (c) MoDL, (d) datos originales.

importancia de resaltar diferentes detalles. El hecho de disponer de la imagen patrón (la que se obtiene sin submuestreo) facilita esta labor sin necesidad de acudir a expertos. Para resaltar estas diferencias hemos superpuesto unas flechas en las imágenes que mostramos como resultados. Puede apreciarse que nuestras reconstrucciones tienden a mantener los detalles señalados mientras que en MoDL tienden a perderse. Entendemos que ello se debe a la presencia de términos en normas ℓ_1 y ℓ_2 , un efecto que ya se ha descrito en otras propuestas [8]. Por otra parte, nuestro esquema de

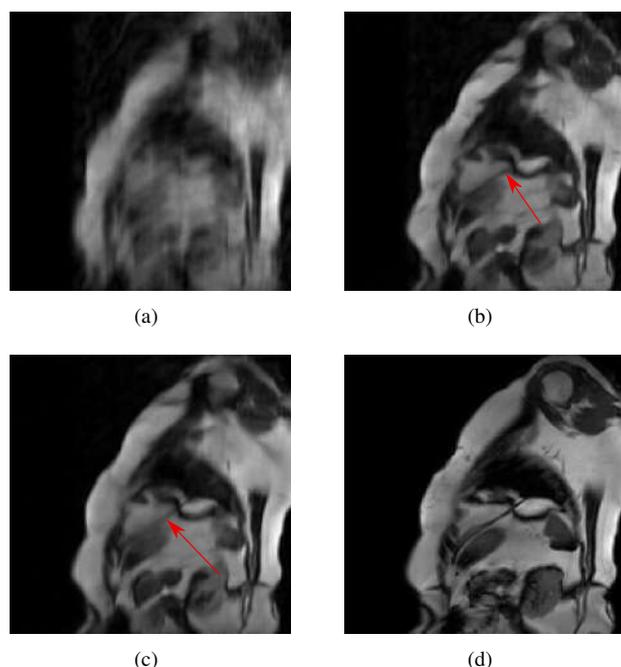


Figura 3. Reconstrucción para distintos modelos con $K = 3$. (a) Datos submuestreados, (b) MoDLMC, (c) MoDL, (d) datos originales.

be estimar a partir de los datos dos parámetros adicionales, tanto el parámetro λ_2 de regularización como el parámetro μ de aproximación del valor absoluto, lo cual para una base de datos relativamente reducida como es la empleada en este trabajo, puede traer consigo una menor capacidad de aprendizaje para nuestra solución.

Los resultados obtenidos, en consecuencia, parecen mostrar la conveniencia de incluir un esquema que tenga en consideración la redundancia temporal; ello hace pensar que la compensación de movimiento es una práctica adecuada para una solución que emplee tanto optimización de modelos como aprendizaje. No obstante, cabe plantearse si la solución basada en aprendizaje supervisado que aquí hemos extendido es la mejor opción para bases de datos relativamente reducidas o si su calidad depende en exceso del punto de partida de los parámetros de regularización de la función de optimización. Probablemente otros esquemas, no necesariamente supervisados, que tengan un menor número de parámetros de regularización pueda aunar las ventajas de la compensación de movimiento con una mejor estimación de los parámetros de la red para una base de datos dada.

5. Conclusiones y líneas futuras

En este artículo hemos presentado una propuesta que añade compensación de movimiento a la arquitectura MoDL. Los resultados de calidad subjetiva muestran una mejora en la capacidad de mantener detalles finos, si bien los parámetros cuantitativos reflejan calidades comparables.

Como las líneas futuras de este trabajo podemos mencionar la necesidad de encontrar valores iniciales de los parámetros de regularización así como la conveniencia de reducir el número de estos términos mediante la exploración de

esquemas alternativos de aprendizaje.

6. Agradecimientos

Referencias

- [1] Zhi-Pei Liang and Paul C Lauterbur. *Principles of magnetic resonance imaging: a signal processing perspective*. IEEE Press, 2000.
- [2] Klaas P Pruessmann, Markus Weiger, Markus B Scheidegger, and Peter Boesiger. Sense: sensitivity encoding for fast mri. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 42(5):952–962, 1999.
- [3] M. Zibulevsky and M. Elad. L1-l2 optimization in signal and image processing. *IEEE Signal Processing Magazine*, 27(3):76–88, 2010.
- [4] Jo Schlemper, Jose Caballero, Joseph V Hajnal, Anthony N Price, and Daniel Rueckert. A deep cascade of convolutional neural networks for dynamic mr image reconstruction. *IEEE transactions on Medical Imaging*, 37(2):491–503, 2017.
- [5] Shanshan Wang, Zhenghang Su, Leslie Ying, Xi Peng, Shun Zhu, Feng Liang, Dagan Feng, and Dong Liang. Accelerating magnetic resonance imaging via deep learning. In *2016 IEEE 13th international symposium on biomedical imaging (ISBI)*, pages 514–517. IEEE, 2016.
- [6] Hemant K. Aggarwal, Merry P. Mani, and Mathews Jacob. Modl: Model-based deep learning architecture for inverse problems. *IEEE Transactions on Medical Imaging*, 38(2):394–405, Feb 2019.
- [7] Sampurna Biswas, Hemant K Aggarwal, and Mathews Jacob. Dynamic mri using model-based deep learning and storm priors: Modl-storm. *Magnetic resonance in medicine*, 82(1):485–494, 2019.
- [8] Elena Martín-González, Teresa Sevilla, Ana Revilla-Orodea, Pablo Casaseca-de-la Higuera, and Carlos Alberola-López. Groupwise non-rigid registration with deep learning: An affordable solution applied to 2d cardiac cine mri reconstruction. *Entropy*, 22(6):687, Jun 2020.
- [9] Javier Royuela-del Val, Lucilio Cordero-Grande, Federico Simmross-Wattenberg, Marcos Martín-Fernández, and Carlos Alberola-López. Jacobian weighted temporal total variation for motion compensated compressed sensing reconstruction of dynamic mri. *Magnetic resonance in medicine*, 77(3):1208–1215, 2017.
- [10] Zhou Wang. The ssim index for image quality assessment. <https://ece.uwaterloo.ca/~z70wang/research/ssim>, 2003.

Reconstrucción de imágenes para tomosíntesis de tórax mediante aprendizaje profundo

C. F. Del Cerro^{1,2}, A. Galán^{1,2}, J. F. Urrea², J.L. Cano-Manuel², J. García-Blas³, M. Desco^{1,2,4,5}, M. Abella^{1,2,4}

¹ Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, España

² Dept. Bioingeniería e Ingeniería Aeroespacial, Universidad Carlos III de Madrid, Madrid, España

³ Arquitectura de Computadores, Comunicaciones y Sistemas, Universidad Carlos III de Madrid, Madrid, España.

⁴ Centro Nacional Investigaciones Cardiovasculares Carlos III (CNIC), Madrid, España

⁵ Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Madrid, España

Resumen

La tomosíntesis es una técnica que ofrece una alternativa a la tomografía computarizada (TAC) y a la radiografía convencional para el diagnóstico clínico. Su principal ventaja es que permite obtener información pseudo-tomográfica con una menor dosis de radiación ionizante que el TAC. Sin embargo, con técnicas de reconstrucción clásicas no se consigue recuperar completamente la información de profundidad ni se trata de una técnica cuantitativa, lo que dificulta el diagnóstico clínico. En este trabajo se propone un método de reconstrucción que utiliza una reconstrucción preliminar seguida de un postproceso basado en aprendizaje profundo (Deep Learning) para obtener una imagen más cercana al TAC. Los resultados preliminares muestran el potencial de la metodología propuesta para obtener información realmente tomográfica a partir de imágenes de tomosíntesis, lo que permitiría sustituir al TAC en aplicaciones donde la dosis de radiación sea un problema crítico.

1. Introducción

La radiografía se utiliza ampliamente en aplicaciones clínicas con fines de diagnóstico. Tiene una alta resolución espacial pero un bajo contraste porque muestra las estructuras internas superpuestas del paciente. La tomografía axial computarizada (TAC) permite quitar esa superposición de estructuras, aunque con una mayor dosis de radiación ionizante.

La tomosíntesis es una técnica pseudo-tomográfica que utiliza menos proyecciones del paciente en comparación con un TAC, lo que implica una menor dosis de radiación ionizante, pero no es capaz de separar completamente los cortes. Por tanto, al observar un corte, las estructuras de otros cortes aparecen superpuestas, lo que dificulta el diagnóstico clínico.

La tomosíntesis ha sido utilizada principalmente en imagen de mama para la detección y el diagnóstico de tumores. Se han investigado varias estrategias de reconstrucción. Por ejemplo, el algoritmo de retroproyección filtrada (FBP) con filtro de rampa modificado [1, 2] o algoritmos de reconstrucción iterativa (IR) [3-5]. Sin embargo, para los algoritmos IR, es necesario diseñar y optimizar una determinada función de coste que contenga un término de fidelidad de los datos y un término de regularización, el cuál suele ser difícil de definir manualmente. Además, en

los IR, la calidad de las imágenes reconstruidas depende en gran medida de la selección de los parámetros iterativos, que deben ajustarse empíricamente o determinarse mediante algunas estrategias que consumen mucho tiempo de cómputo.

En los últimos años, el aprendizaje profundo ha suscitado un gran interés de investigación en el campo de las imágenes médicas, como la reconstrucción de imágenes de tomografía computarizada de baja dosis (LDCT) [6] o para acelerar la reconstrucción de imágenes de resonancia magnética (MRI) [7, 8]. Hasta ahora, se pueden definir dos enfoques diferentes para utilizar el aprendizaje profundo para mejorar la calidad de las imágenes médicas. Por un lado, se obtienen las imágenes utilizando algoritmos de reconstrucción convencionales y, a continuación, se utiliza una red neuronal convolucional (CNN) para refinar estas imágenes y eliminar los artefactos. Por otro lado, en lugar de aplicarse únicamente en el dominio de la imagen, la CNN también puede utilizarse para trabajar con los datos de proyección. Como un ejemplo de esto, se puede utilizar un algoritmo IR combinado con aprendizaje profundo en [6, 9] y así trabajar simultáneamente con los datos de proyección adquiridos y la imagen reconstruida. Para la reconstrucción de imágenes de mama, en [10] se ha demostrado la viabilidad de utilizar un algoritmo IR basado en aprendizaje profundo, mediante el cual se han obtenido imágenes con una alta precisión de la densidad mamaria y un artefacto de imagen reducido.

Este trabajo propone una metodología de reconstrucción para obtener imágenes más cercanas a un TAC a partir de datos de tomosíntesis. El proceso consiste en un paso de retroproyección filtrada seguido de un postproceso basado en aprendizaje profundo.

2. Construcción de la base de datos

La base de datos se ha construido a partir de 43 estudios de TAC de tórax extraídos del portal *Medical Imaging and Data Resource Center (MIDRC)* procedentes de tres colecciones distintas. Los TAC seleccionados tenían un tamaño máximo de vóxel en el eje de profundidad menor o igual a 1 mm para evitar que las proyecciones simuladas tuvieran una baja resolución.

Debido a su diversa procedencia, estos TAC presentan distinto tamaño de vóxel. Para simplificar el problema se decidió convertir todos los estudios al mismo tamaño vóxel haciéndolos a su vez isotrópicos, resultando en volúmenes de 1.3x1.3x1.3 mm de tamaño de vóxel.

Dado que en una adquisición de tomosíntesis de tórax, la espalda del paciente se sitúa pegada al *bucky* que contiene el detector, se recortó el TAC para eliminar el aire sobrante del campo de visión. Así mismo, se eliminó la cama y demás soportes del paciente presentes en el TAC. Para evitar el artefacto causado por la limitación espacial de los TAC, que incluyen sólo la zona del tórax, se realizó una extensión de estos, como se muestra en la Figura 1.

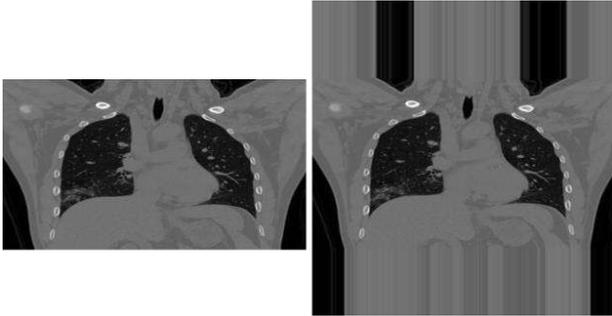


Figura 1. TAC original (izquierda) y TAC extendido a partir de la primera y la última rodaja axial (derecha)

Para simular las proyecciones y realizar las reconstrucciones se usó FuxSim [11], un software que permite simular diferentes geometrías de adquisición a partir de un mapa de densidad y unos parámetros del sistema. La simulación se basó en un protocolo estándar de tomosíntesis de tórax: 61 proyecciones con una distancia total entre la fuente y el detector de 1800 mm y un ángulo de barrido de 30 grados. Las proyecciones resultantes tenían un tamaño de 1072x1072 píxeles con un tamaño de píxel de 0.4x0.4 mm.

La reconstrucción preliminar de los volúmenes se realizó con un algoritmo basado en la retroproyección filtrada, modificando el filtro rampa para evitar la pérdida de valor medio. Para cada estudio se reconstruyó un volumen de 265x185x80, correspondiente a la parte central del tórax, que es la parte que se evalúa para el diagnóstico. La Figura 2 muestra el corte coronal central de la reconstrucción de tomosíntesis junto con el del TAC de referencia para tres volúmenes diferentes.

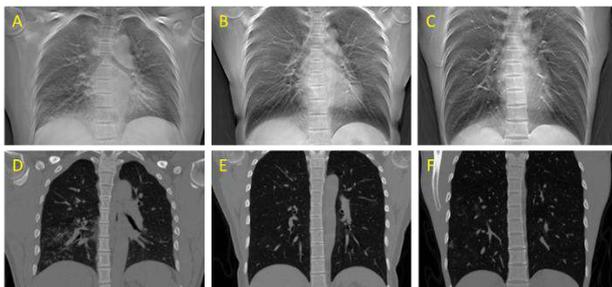


Figura 2. Corte coronal de tres reconstrucciones obtenidas con retroproyección filtrada a partir de datos de tomosíntesis (arriba) y de sus TAC correspondientes (abajo). A y D corresponden al conjunto de entrenamiento, B y E al de validación y C y F al de test.

La base de datos, compuesta por estas reconstrucciones de tomosíntesis como entradas y los TAC como referencia, se dividió en los siguientes conjuntos:

- Conjunto de entrenamiento: Compuesto por 31 volúmenes, dando un total de 2480 imágenes.
- Conjunto de validación: Compuesto por 7 volúmenes, dando un total de 560 imágenes.
- Conjunto de test: Compuesto por 5 volúmenes, dando un total de 400 imágenes.

3. Paso de postproceso

El paso de postproceso tiene como entrada una reconstrucción preliminar obtenida con retroproyección filtrada modificada y como salida un volumen con información tomográfica realzada. Para ello se utiliza una versión actualizada de la arquitectura U-net original [12]. La red se divide en dos partes, codificador y decodificador, como se muestra en la Figura 3.

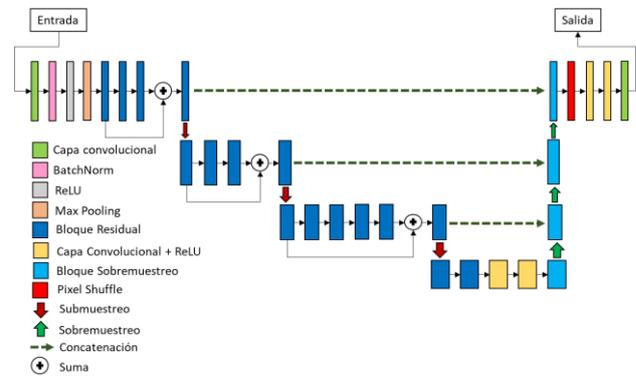


Figura 3. Diagrama de la arquitectura de red U-net utilizada.

El codificador está formado por la red neuronal convolucional ResNet-34 [13], debido a su rendimiento y la gran cantidad de documentación disponible online, entrenada previamente con *Imagenet* [14]. El decodificador está formado principalmente por bloques de sobremuestreo. Estos bloques están compuestos por una capa convolucional, a cuya salida se aplica la operación de *pixel shuffle* [15], para reducir el coste computacional, seguida de una capa de *pooling* de media, una capa de normalización por lotes, una *ReLU* y dos capas de convolución seguidas de una *ReLU* cada una. El hecho de que el codificador formado principalmente por ResNet-34 estuviese entrenado con *Imagenet* permitió realizar aprendizaje por transferencia, una técnica que utiliza un modelo entrenado con un conjunto de datos muy grande para adaptar ese conocimiento a otro conjunto de datos diferente y que según se demuestra en [16] da mejores resultados en una amplia gama de tareas. El decodificador se inicializó aleatoriamente. El entrenamiento se llevó a cabo en dos fases: primero se congeló el codificador y solo se entrenó el decodificador y después, se descongelaron las capas del codificador para terminar de ajustar todo el modelo entrenando la red de principio a fin.

Se han probado tres funciones de coste diferentes: el error cuadrático medio (*RMSE*), para generar un modelo de base,

y el índice de similitud estructural (*SSIM*), y *perceptual loss* (*PL*), ambas métricas basadas en la percepción, que han demostrado buenos resultados en una gran variedad de problemas.

Como estrategia de regularización se ha utilizado *weight decay* igual a 0.001, para que su peso en la función de coste no provoque la no convergencia del modelo. Como optimizador se eligió Adam [17], debido a su mayor velocidad de convergencia. Para determinar la tasa de aprendizaje se realizó el test presentado por Leslie N. Smith en [18], que dio como resultado una tasa de aprendizaje óptima de 0.001 para el entrenamiento del decodificador y de 0.00001 para el entrenamiento de la red completa para las tres funciones de coste diferentes.

4. Evaluación y Resultados

La Figura 4 muestra el corte coronal central de uno de los estudios de test para las diferentes funciones de coste utilizadas, junto con la referencia.

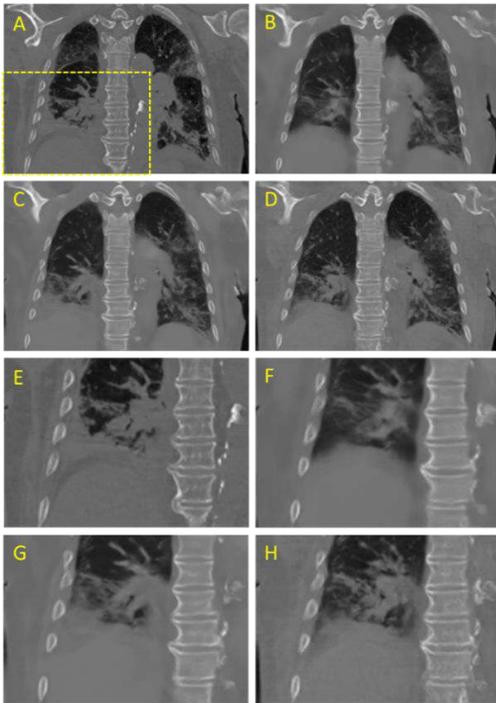


Figura 4. Corte coronal central y zoom de: TAC (A, E), RMSE (B, F), SSIM (C, G) y PL (D, H).

Se aprecia que los tres modelos consiguen una buena estimación en el hueso, sobre todo en la zona de la columna. Sin embargo, se observa como los modelos obtenidos con *RMSE* y *SSIM* pierden resolución de contraste en los tejidos blandos, generando una imagen más suavizada.

En la Figura 5 se puede ver otro corte coronal del mismo volumen de test donde se observa como los tres modelos pierden resolución espacial, en las estructuras más pequeñas, como, por ejemplo, los vasos sanguíneos del pulmón, donde las estimaciones se vuelven significativamente peores.

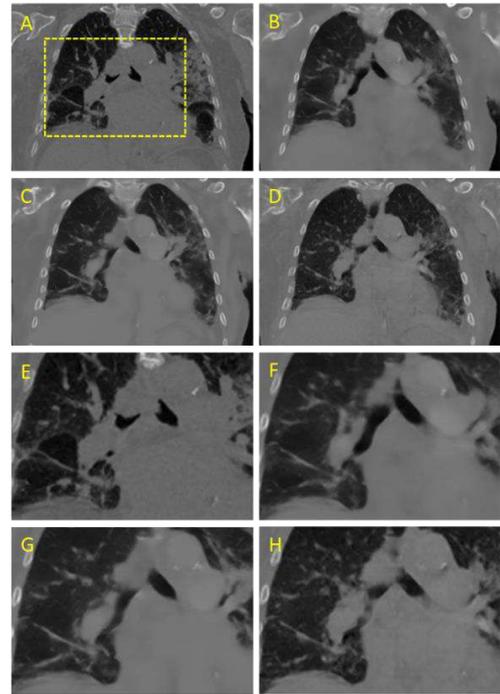


Figura 5. Corte coronal y zoom de: TAC (A, E), RMSE (B, F), SSIM (C, G) y PL (D, H).

La Figura 6 muestra el error cuadrático medio con respecto al TAC de referencia de los cinco volúmenes de test para cada modelo, junto con el correspondiente a la reconstrucción preliminar. Se puede ver como para los tres modelos el error disminuye en los cortes coronales centrales. Esto puede ser debido a que la reconstrucción preliminar está más cerca de la referencia en esos cortes. Además, se observa como los tres modelos tienen un error muy similar, siendo el modelo con *RMSE* el que obtiene menor error. Esto puede deberse a que, a pesar de que con *RMSE* y *SSIM* se pierde más resolución de contraste en tejido blando que con *PL*, el valor medio de los píxeles es más cercano a la referencia.

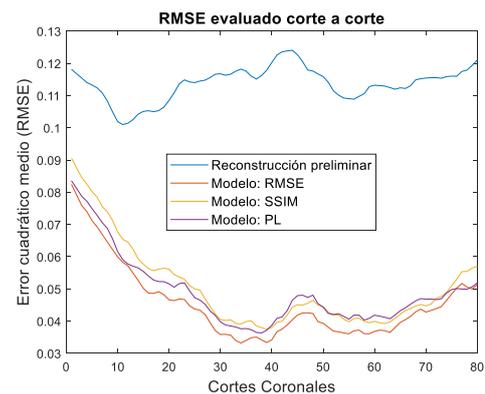


Figura 6. Evolución del error cuadrático medio a lo largo de los cortes coronales para los 5 volúmenes de test.

5. Discusión y Conclusiones

Este trabajo presenta una metodología de reconstrucción para tomosíntesis de tórax basada en un paso de retroproyección filtrada seguido de un postproceso que utiliza aprendizaje profundo. Se evaluó la arquitectura

U-net con ResNet-34 en el codificador, debido a su rendimiento en una gran variedad de tareas, con tres funciones de coste diferentes, *RMSE*, *SSIM* y *PL*, debido a que han sido ampliamente usada en problemas con fines similares a este dando buenos resultados. Los tres modelos pierden resolución espacial, sobre todo en las pequeñas estructuras. Sin embargo, se puede comprobar como *PL* consigue una mejor resolución de contraste en los tejidos blandos que *RMSE* y *SSIM*. A pesar de los pocos datos disponibles, se puede concluir que el aprendizaje profundo tiene un gran potencial para la tarea de recuperación de información tomográfica a partir de datos limitados, es decir, con pocas proyecciones y poco ángulo de barrido. Es importante mencionar que la base de datos actual está compuesta por imágenes de pacientes tanto sanos como con patologías, lo que añade una gran variabilidad a los mismos y, por tanto, una gran complejidad al problema.

Además de ampliar la base de datos, el trabajo futuro se basará en dos líneas distintas. Por un lado, se planteará una estrategia de entrenamiento multi-escala, entrenando con conjuntos de datos de diferente resolución para permitir a la red pueda aprender progresivamente las características de la imagen a diferentes resoluciones. Y, por otro lado, se abordará el problema incluyendo información 3D. Para ello, se plantean dos estrategias: por un lado, realizar entrenamientos intercalados con cortes coronales y axiales, y, por otro utilizar una red neuronal convolucional que permita procesar los datos directamente en 3D.

Agradecimientos

Este trabajo ha sido financiado por el Programa de apoyo a la realización de proyectos interdisciplinares de I+D para jóvenes investigadores de la Universidad Carlos III de Madrid 2019-2020 en el marco del Convenio Plurianual Comunidad de Madrid- Universidad Carlos III de Madrid (proyecto DEEPCT-CM-UC3M) y por CRUE Universidades, CSIC y el Banco Santander (Fondo Supera, proyecto RADCOV19). El CNIC está financiado por el Ministerio de Ciencia, Innovación y Universidades y la fundación PRO-CNIC y es un centro de excelencia Severo Ochoa (SEV-2015-0505). Las imágenes y los datos clínicos asociados que fueron descargados de MIDRC (The Medical Imaging Data Resource Center) y utilizados para la investigación en esta publicación, fueron posibles gracias al National Institute of Biomedical Imaging and Bioengineering (NIBIB) de los National Institutes of Health bajo los contratos 75N92020C00008 y 75N92020C00021.

Referencias

- [1] I. Sechopoulos, "A review of breast tomosynthesis. Part II. Image reconstruction, processing and analysis, and advanced applications," *Medical physics*, vol. 40, p. 014302, 2013.
- [2] S. D. Rose, E. Y. Sidky, I. S. Reiser, and X. Pan, "Filtered back-projection for digital breast tomosynthesis with 2D filtering," in *Medical Imaging 2019: Physics of Medical Imaging*, 2019, p. 1094851.
- [3] S. Xu, J. Lu, O. Zhou, and Y. Chen, "Statistical iterative reconstruction to improve image quality for digital breast tomosynthesis," *Medical physics*, vol. 42, pp. 5377-5390, 2015.
- [4] T. Wu, R. H. Moore, E. A. Rafferty, and D. B. Kopans, "A comparison of reconstruction algorithms for breast tomosynthesis," *Medical physics*, vol. 31, pp. 2636-2647, 2004.
- [5] S. D. Rose, A. A. Sanchez, E. Y. Sidky, and X. Pan, "Investigating simulation-based metrics for characterizing linear iterative reconstruction in digital breast tomosynthesis," *Medical physics*, vol. 44, pp. e279-e296, 2017.
- [6] J. He, Y. Yang, Y. Wang, D. Zeng, Z. Bian, H. Zhang, et al., "Optimizing a parameterized plug-and-play ADMM for iterative low-dose CT reconstruction," *IEEE transactions on medical imaging*, vol. 38, pp. 371-382, 2018.
- [7] Y. Yang, J. Sun, H. Li, and Z. Xu, "ADMM-CSNet: A deep learning approach for image compressive sensing," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, pp. 521-538, 2018.
- [8] S. Wang, Z. Su, L. Ying, X. Peng, S. Zhu, F. Liang, et al., "Accelerating magnetic resonance imaging via deep learning," in *2016 IEEE 13th international symposium on biomedical imaging (ISBI)*, 2016, pp. 514-517.
- [9] J. Cheng, H. Wang, Y. Zhu, Q. Liu, Q. Zhang, T. Su, et al., "Model-based deep medical imaging: the roadmap of generalizing iterative reconstruction model using deep learning," *arXiv preprint arXiv:1906.08143*, 2019.
- [10] N. Moriakov, K. Michielsen, J. Adler, R. Mann, I. Sechopoulos, and J. Teuwen, "Deep learning framework for digital breast tomosynthesis reconstruction," in *Medical Imaging 2019: Physics of Medical Imaging*, 2019, p. 1094804.
- [11] M. Abella, E. Serrano, J. Garcia-Blas, I. García, C. De Molina, J. Carretero, et al., "FUX-Sim: Implementation of a fast universal simulation/reconstruction framework for X-ray systems," *PloS one*, vol. 12, p. e0180363, 2017.
- [12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234-241.
- [13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
- [14] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, 2009, pp. 248-255.
- [15] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, et al., "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1874-1883.
- [16] S. Kornblith, J. Shlens, and Q. V. Le, "Do better imagenet models transfer better?," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2661-2671.
- [17] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [18] L. N. Smith, "Cyclical learning rates for training neural networks," in *2017 IEEE winter conference on applications of computer vision (WACV)*, 2017, pp. 464-472.

A simplified methodology to measure gait spatio-temporal parameters using low-cost inertial sensors.

K.M. Reyes Leiva^{1,2}, M.A. Cuba Gato¹, J.J. Serrano Olmedo^{1,3}

¹ Center for Biomedical Technology (CTB), Universidad Politécnica de Madrid, Madrid, Spain, {karla.reyes, josejavier.serrano}@ctb.upm.es, macuba@alumnos.upm.es

² Engineering Faculty, Universidad Tecnológica Centroamericana UNITEC, San Pedro Sula, Honduras

³ Networking Center of Biomedical Research for Bioengineering, Biomaterials and Nanomedicine, Instituto de Salud Carlos III, Madrid, Spain.

Abstract

*In the Orientation and Mobility (O&M) training for visually impaired people (VIP), the spatio-temporal gait parameters are of specific interest for rehabilitators to assess performance and improvements. Thus, the objective of this research was to propose a simple architecture based on the use of inertial sensors for quantitative estimation of distance traveled, step detection, gait velocity & step length. These parameters were calculated using absolute orientation angles with three different algorithms according to the selected biomechanical model. The methods include the use of one sensor placed either in the thigh or shank or a third method using both sensors to obtain the measurement of the rotations experienced by both leg segments in the sagittal plane during the gait cycle. The validation tests were performed with real-time acquisitions under experimental conditions walking indoor distances at regular, high, and low gait velocity. Two of the proposed methods were selected for presenting the lowest percentage of relative error in all the traveling experimentations (4-15,8%). The difference from the measured value to the actual value of each step varies between 2-7 cm for the regular and high-velocity traveling, and, between 10-31 cm for slow traveling & small stepping for the selected methods. The results suggest that the proposed architecture could be an assistive technology used for O&M training. Further research including the validation of this proposed methodology with visually impaired volunteers and rehabilitators will be done as the next steps in this research. For this purpose, some hardware improvements must be done. **Keywords: Gait parameters, Inertial Sensors, Orientation and Mobility, Visually Impaired People.***

1. Introduction

The gait analysis is an important research field for a large and ever-growing number of health applications. As evidenced in most of the gait studies, the use of optical motion analysis systems is often referred to measure and analyze the spatio-temporal gait parameters [1]. However, the characteristics of this systems (expensive, robust & immobile, resource intensive) limits its accessibility and applications. These gait parameters have been measured with success using inertial measurement unit sensors (IMUs). IMUs are motion sensors since they are designed to measure acceleration (the specific force in the sensor itself during movement), and angular velocity (the time rate in which an object rotates) with respect to the inertial frame. In some cases, IMUs are complemented with magnetometers, that provide the measurement of the local magnetic field to obtain orientation angles (roll, pitch and heading). By this sensor

input, it has been possible to obtain real gait measurements [2] to support different applications such as navigation [3].

Navigation electronic aids for visually impaired have been developed for many years [4], [5], different solutions with high accuracy have been proposed, however most of them rely on excessive architecture due to all the requirements needed for environmental sensing. This can lead rejection of the technologies by the VIP.

Some authors highlight the importance of development of navigation solutions specifically designed for VIP, due to the differences in gait patterns in comparison with people without visual impairment [6].

A recent review focusing on the use of inertial sensors for assistive technologies for VIP showed that there is not inertial sensor dedicated systems for VIP, and that there is a gap in the literature regarding biomechanical analysis focused on VIP-oriented application and using IMU's as the sensing input [7].

Moreover, most of the spatiotemporal gait parameters analyses for VIP have been developed using Motion Tracking systems. However, in this article three IMU-based methods for measuring step length (SL), gait velocity (GV), step count (SC) and total displacement(D) are presented.

The novelty of the methods results from the combination of

1. The use of absolute orientation angle values (leg rotation) during the gait cycle to calculate the SL and GV instead of accelerometry or angular velocity values & the reliance on the gait biomechanical model instead of abstraction model or direct integration model to avoid the accumulated drift error to calculate each SL and D traveled. This allows maintaining the advantage of having a simple system set up with low computational processing [8].
2. The precision of the real-time measurement of the spatio-temporal gait parameters despite the simplicity of the method which embeds only one or two wearable low-cost inertial sensors. This avoids the after acquisition Gait Analysis software's used to calculate the parameters and process data in clinical approaches as in [9].
3. The proposed user-oriented design to be implemented as an assistive device for O&M training and the conditions settled in the measurement algorithms. In the current literature, there is a wide interest for clinical research of gait parameters using wearable inertial

sensors for instance, for neurological diseases such as Stroke, Parkinson or Multiple Sclerosis [10], also, the wearable inertial sensors have been used to monitor clinical gait parameters of interest in the elder population [11]. However, for VIP gait analysis, non-wearable systems have been designed [1], [12], leading to a gap in user-oriented wearable systems design.

Experimental tests were carried out with sighted volunteers to determine which of the methodologies presented the most accurate results, resulting into two final methods. Sighted volunteers were included for this stage of validation since the proposed methods can be employed for other fields of applications; clinical or non-clinical. As a next stage in this research, the validation of the selected methods with VIP must be done.

2. Methodology

The measuring system including two BNO055 Bosch 9DOF inertial sensors was developed previously and is detailed in [9].

2.1. Biomechanical principles

In a gait cycle formed by two consecutive steps, θ (Figure 1 - A) refers to the inclination angle with respect to the vertical measured by the inertial sensor, which can be placed on either of the two legs. For its part, γ is the rotation angle between heel strike and heel off, during the stance phase of the gait cycle of both legs.

This rotation can be calculated in the case of the leg wearing the device and can be approximated in the case of the opposite leg, from the values of θ .

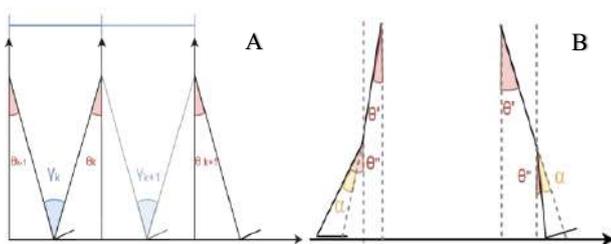


Figura 1. Biomechanical models (A) thigh sensor, (B) Two-sensors & shank sensor.

Moreover, unlike the simplified schema in (Figure 1 - A) suggests, the knee doesn't remain stiff during gait. This means that γ should be, more accurately, the rotation of the virtual segment connecting the hip with the heel. To determine this segment, the inclination, and the length of both main subsegments of the leg (thigh and shank) must be known. This adjusted model is the one shown in Figure 1- B. In the case of the two-sensor (TWS) configuration, both inclinations were measured by the IMUs. For its part, in the case of one-sensor-only configurations (named "thigh sensor (THS)" and "shank sensor (SHS)" respectively), the sagittal flexion-extension angles of the knee at heel strike and heel off were extracted from bibliographic references [13], [14], and the unknown inclination was subsequently calculated relying on the fact that $\alpha = \|\theta' - \theta''\|$, where θ' is the thigh inclination, θ'' is the shank inclination, and α is the flexion-extension angle.

In any case, the rotations denoted by γ lead to a displacement of the center of mass in the x axis of the navigation coordinate frames which matches the direction of the current step (k). Thus, this displacement becomes a good approximation of the step length.

2.2. Experimental procedure

Ten healthy volunteers were asked to participate in the acquisitions. The study was previously approved by the ethics committee of Universidad Politécnica de Madrid in October 2020 (Ref. ID 2020000224). The subjects were instructed to walk a straight line in five different conditions: Exp1: Walk for 10 meters in regular velocity. Exp2: Walk for 20 meters in regular velocity. Exp3: Walk for 20 meters in fast velocity. Exp4: Walk for 20 meters in slow velocity and small stepping. Exp5: Walk for 20 meters with a 5 second stop in halfway (10 meters). Since in this proposed validation there is no participation of visually impaired volunteers, Exp 4 and Exp5 were included to have an approximation of the gait characteristics of the final users.

3. Results and Discussion

3.1. Step count & Step Length

Two different SC algorithms were used, one for both TWS and THS methods, and one for the SHS method. The results from the experimental conditions showed that the SHS was more accurate for SC. The comparison was made with the real SC registered visually by two volunteers during the acquisitions. Total errors (%Error) of 4,69% and 2,73% were obtained from the SC methods, averaged for all experimental conditions and 10 subjects. The algorithms were designed to detect a new step whenever the leg passes through the vertical position and, to detect the local maximum and local minimum in each inflexion point from the rotation angles patterns. Also, a simple activity recognition algorithm based on the angular velocity was developed. Its purpose was to determine if the user is moving or not, to avoid the detection of "false steps".

Other conditions were added. For instance, to count a new step, each local maximum must be followed by a local minimum and vice versa. This can produce a variation in the pattern due to the variation in the rotation angles in the heel strikes at this moment. However, the %Error obtained for the SC is the lowest compared with the %Error obtained for D, SL and GV measurements. Figure 2 shows a representation of the rotation angles patterns in Exp2 and Exp5 for S05 acquisitions. In the first case the SHS method counted 30/29 steps while the THS counted 28/29. Second, the SHS method counted 29/29 steps while the THS counted 29/29. In this case, even with a stand stop (which means no rotation angles during some seconds), with light variation in the pattern for the heel strike and as shown in Fig2-b, the step count algorithms worked with accuracy in both methods.

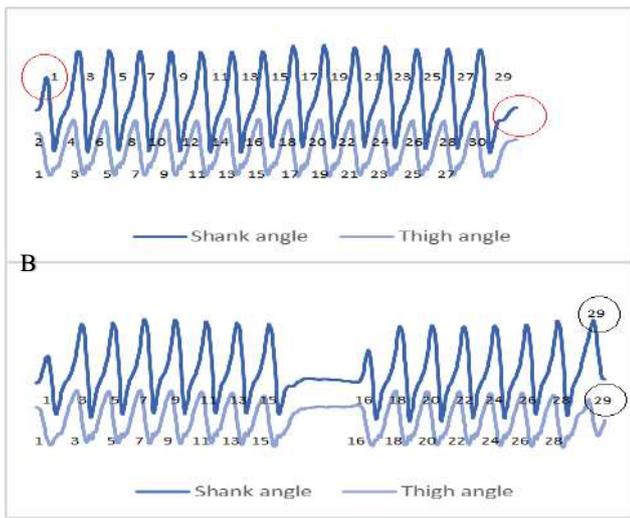


Figura 2. Comparison of the rotation angles measured from the Shank and Thigh in the Exp2 (A) Exp5 (B).

Regarding SL, the calculated mean accepted value in order to obtain the correct %error, the register SC by the volunteers and the defined total displacement was used. From the values of the rotation γ , the SL was obtained with the following equation, which implements the cosine law:

$$SL_k = \sqrt{2 * hh_i^2 - 2 * hh_{i+1} * cos \gamma_k}$$

Where, SL_k is the length of the step number k and hh_i are the lengths of the hip-heel segments for each leg. The mean %Error of the measured SL for Epx1, Epx2, Exp3 and Exp5 for all 10 subjects was 11,17% for the TWS, 37,22% for the SHS method and 10,05% for the THS method. Table 1 summarizes the mean Error for TWS and THS only, -note that for this table, only Exp1 and Exp2 were considered as normal velocity gait experimental conditions-. TWS presents higher Error than THS, (7,35 cm vs. 5,58cm), however THS presents more SD than TWS (3,74 cm vs. 2,05 cm).

	SS (cm)	TWS (cm)	Error (cm) (mean±SD)	THS (cm)	Error (cm) (mean±SD)
S01	83,43	88,33	4,90±2,08	76,00	7,43±5,65
S02	63,89	70,67	6,78±2,31	67,00	3,11±18,25
S03	68,25	74,33	6,08±2,08	76,00	7,75±1,00
S04	68,25	82,00	13,75±1,73	79,33	11,08±1,15
S05	68,20	79,67	11,47±0,58	82,67	14,47±1,53
S06	60,01	66,00	5,99±3,46	62,33	2,32±4,16
S07	68,25	76,33	8,08±1,15	83,00	14,75±2,00
S08	60,19	70,00	9,81±1,00	67,33	7,15±0,58
S09	63,89	68,33	4,44±2,08	67,00	3,11±2,65
S10	84,54	82,33	2,21±4,04	85,00	0,46±3,46
	Mean±SD		7,35±2,05	Mean±SD	5,58±3,74

Tabla 1. Actual step size versus estimated step size, error, and standard deviation in centimeters for TWS and THS.

3.2. Distance & Gait Velocity

Finally, the D measurement refers to the addition of the measured SL for every detected step as shown in Fig.3. This means that, the %Error from the SL and SC measurements

directly affects the results for the measured distance. Also, the gait velocity for each method was obtained by dividing the measured distance by the total walk time. Also, to estimate walking time and have a mean accepted value to obtain the correct %error in GV two volunteers register every EXP time using a chronometer. In the case of GV, the velocity error can be affected by the SL and SC measurements as well as by the addition of a time counter. As shown in Figure 3, from Exp where the total displacement was 20 meters, the more accurate methods are THS and TWS.

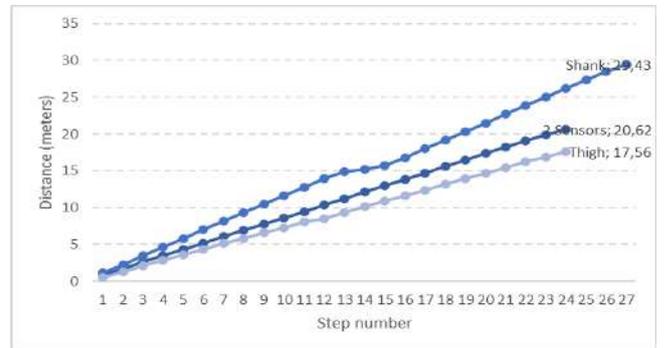


Figura 3. (A) Step length measurements in each step count for the Exp2 in S05.

In fact, the total %Error for D measurement for Epx1, Epx2, Exp3 and Exp4 for all 10 Subjects was 8,22% for the TWS, 38,69% for SHS method and 9,83% for the THS method. Note that if the Exp 4 experimental condition is added to the mean values, there is an increase of the %Error for THS and SHS to 13,02% and 42,89%, respectively, while the TWS %Error remains the same. This suggests that this method can be more accurate to measure D and SL, SC for small stepping and low GV.

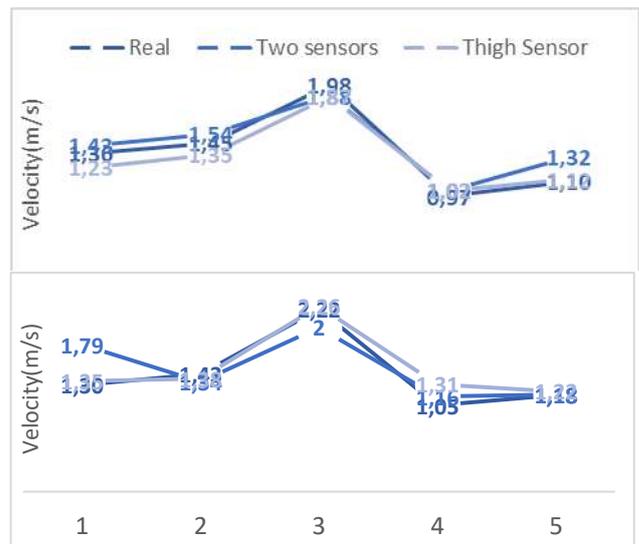


Figura 4. Gait Velocity measurement in two random subjects for Exp1 (1), Exp2 (2), Exp3(3), Exp4 (4) and Exp5 (5).

Finally, for GV measurements, there is a very small variation in %Error between THS and TWS methods in all the experimental conditions. The mean %Error value was 11,93% for TWS and 12,90% for THS, -these methods shared the same methodology for time counting during gait cycles. Figure 4 shows the measurement values of two

random subjects in all the experimental conditions. The figure demonstrates how both methods can detect the changes in velocity for Exp3 and Exp4 and that the normal velocity measured for Exp1,2 and 3 in both methods are almost the same.

In general, TWS presented a higher accuracy with lower SD in the summary of all experimental conditions and for all spatio-temporal gait parameters measured.

Since a great percentage of the VI population is elder, due to the characteristics of their gait, the characteristics of the GV for instance can vary compared to healthy adults. According to recent research, healthy adults can have a mean GV from 1,39 to 1,49 m/s [9] while blind or VI people, present a mean GV from 0,86-1,11 m/s [12]. As stated by the results of Exp4 and Exp5, both TWS and THS methods could be used in O&M training of VIP. However, the TWS is more suitable due to accuracy. Also, if the SC method of SHS is used in the TWS configuration, there is a possibility to reduce the main %Error in D and GV.

Finally, with the inputs of the system (SC, SL, GV, D) it is possible to estimate other spatio-temporal parameters that can be of interest for rehabilitator of O&M such as "Candence" or "Stride rate" which is the rate that measures the steps per minute or double stance [15]. Or by implementing the same condition that determinates if the person is walking or not for SC (validated in Exp5), it can be measured by instance the time that a user pauses between steps to move from point A to point B during a displacement task.

4. Conclusions

Three different methods to evaluate spatio-temporal parameters were designed and tested in the following experimental conditions: 10- and 20-meters travel distances and normal, low, and high velocities. The more accurate values were presented by the Two-sensors configuration for D (8,82%Error), SL (10,59%Error) and GV(4,69%Error), however very close values with slightly lower accuracy, higher SD and lower Error were obtained by the one-sensor-only THS configuration, with differences of +4,21%Error for D, +5,20%Error for SL and +0,97%Error for GV. The SHS resulted more accurate to measure SC with a 2,73%Error. The architectural differences between the more accurate methods (TWS and THS) are significative since adding a second sensor means adding extra computational cost and adding hardware cost. Nevertheless, the Exp4 results indicate that TWS can be more accurate for gait evaluation of elder VIP by instance, who have a slower gait velocity and smaller step length. As a next stage in this research, the validation of the selected two methods in different experimental conditions with the participation of VIP must be done to determine which is the more appropriate method for the assistive technology as an O&M tool.

Acknowledgment

The author Karla Miriam Reyes Leiva acknowledges scholarship support from the Fundación Carolina FC and the Universidad Tecnológica Centroamericana UNITEC.

References

- [1] A. N. Sankako, P. Marília, P. R. Garcia Lucareli, S. M. Ribeiro de Carvalho, and L. M. Presumido Braccialli, "Temporal spatial parameters analysis of the gait in children with vision impairment," *Int. J. Orientat. Mobil.*, vol. 8, no. 1, pp. 90–100, 2016.
- [2] M. Patel, A. Pavic, and V. A. Goodwin, "Wearable inertial sensors to measure gait and posture characteristic differences in older adult fallers and non-fallers: A scoping review," *Gait and Posture*. 2020.
- [3] S. Basso, G. Frigo, and G. Giorgi, "A smartphone-based indoor localization system for visually impaired people," *2015 IEEE Int. Symp. Med. Meas. Appl. MeMeA 2015 - Proc.*, pp. 543–548, 2015.
- [4] M. M. Islam, M. S. Sadi, K. Z. Zamli, and M. M. Ahmed, "Developing Walking Assistants for Visually Impaired People: A Review," *IEEE Sens. J.*, vol. 19, no. 8, pp. 2814–2828, Apr. 2019.
- [5] R. N. Kandalan and K. Namuduri, "Techniques for Constructing Indoor Navigation Systems for the Visually Impaired: A Review," *IEEE Trans. Human-Machine Syst.*, vol. 50, no. 6, pp. 492–506, 2020.
- [6] G. H. Flores and R. Manduchi, "WeAllWalk," *ACM Trans. Access. Comput.*, vol. 11, no. 1, pp. 1–28, 2018.
- [7] K. M. Reyes Leiva, M. Jaen, B. Codina, and J. J. Serrano Olmedo, "The role of Inertial Measurement Unit sensors in assistive technologies for visually impaired people: A review.," *Sensors*, vol. 21, no. 4767, pp. 1–26, 2021.
- [8] S. Yang and Q. Li, "Inertial Sensor-Based Methods in Walking Speed Estimation: A Systematic Review," *Sensors*, vol. 12, pp. 6102–6116, 2012.
- [9] J. Soulard, J. Vaillant, R. Balaguier, and N. Vuillerme, "Spatio-temporal gait parameters obtained from foot-worn inertial sensors are reliable in healthy adults in single- and dual-task conditions," *Sci. Rep.*, vol. 11, no. 1, pp. 1–15, 2021.
- [10] T. Bowman *et al.*, "Wearable devices for biofeedback rehabilitation: A systematic review and meta-analysis to design application rules and estimate the effectiveness on balance and gait outcomes in neurological diseases," *Sensors*, vol. 21, no. 10, 2021.
- [11] L. Ruiz-Ruiz, A. R. Jimenez, G. Garcia-Villamil, and F. Seco, "Detecting fall risk and frailty in elders with inertial motion sensors: A survey of significant gait parameters," *Sensors*, vol. 21, no. 20, 2021.
- [12] E. S. da Silva *et al.*, "Gait and functionality of individuals with visual impairment who participate in sports," *Gait Posture*, vol. 62, no. December 2017, pp. 355–358, 2018.
- [13] J. Perry, "In Atlas of Limb Prosthetics: Surgical, Prosthetic, and Rehabilitation Principles," in *American Academy of Orthopedic Surgeons*, Rosemont, IL, 1992.
- [14] A. Almandari and V. N. Krovi, "A Review of Computational Musculoskeletal Analysis of Human Lower Extremities," in *In Human Modeling for Bio-Inspired Robotics: Mechanical Engineering in Assistive Technologies*, Elsevier Inc, 2017, pp. 37–73.
- [15] V. K. Ramsey, B. B. Blasch, and A. Kita, "Effects of Mobility Training on Gait and Balance," *J. Vis. Impair. Blind.*, vol. 97, no. 11, pp. 720–726, 2003.

Wearable technology for predicting the quality of surgical performance and workload in laparoscopic surgery

J. Castillo Rabazo¹, J. A. Sánchez-Margallo¹, D. Durán Rey¹, M. R. González Portillo¹, I. López-Agudelo¹, F. M. Sánchez-Margallo¹, P. A. Gloor².

¹ Jesús Usón Minimally Invasive Surgery Centre, Cáceres, Spain, {jcastillo, jasanchez, dduran, mrgonzalez, ilopez, msanchez}@ccmijesususon.com

² MIT Centre for Collective Intelligence, MA, USA, pgloor@mit.edu

Abstract

This study presents the first steps towards the development of a machine learning algorithm with the purpose of predicting surgical performance and workload, both physical and mental, during laparoscopic surgical training by means of hand accelerations measured with a smartwatch.

1. Introduction

Artificial Intelligence (AI) has quickly gained a place in virtually every branch of scientific research. Healthcare in general, and medical technology in particular, are no exceptions with an increment in published papers of more than a 1000% in ten years [1]. Despite the recent advances in AI algorithms and in the technology needed to support and to run them, data collection and processing remain one of the biggest challenges for every AI application, as for the majority of them, a huge quantity of data is required in order for those algorithms to work properly [2]. Besides, when it comes to applications in surgery, where data coming from surgeons with different levels of expertise is necessary, the challenge is even bigger, as this type of data is arduous and expensive to obtain.

Surgical training is crucial for surgeons in order to build the skillset required in today's industry standards. Objective evaluation of certain surgical skills is mostly difficult, costly, and time consuming. Wearable technology and AI offer a promising solution to this problem, as advances in sensor technology have allowed for smaller and more comfortable devices that could be used during surgery with little to no impact to the surgeon's performance. In this context, we propose the use of smartwatches during surgery to acquire motion data coming from the surgeon's hand accelerations. In particular, this study aims to present the first steps and results towards the development of a machine learning (ML) algorithm to predict surgical performance and workload, both physical and mental, during laparoscopic surgery training.

2. Materials and Methods

This study was conducted during two training courses in laparoscopic gynaecology and one training course in urological laparoscopic surgery. Tubal ligation,

oophorectomies and prostatectomies were performed during the courses. Surgeons wore a Ticwatch E2 smartwatch (Mobvoi; Kowloon, Hong Kong) on their dominant hand with a custom application running on it with the purpose of acquiring motion data (Fig. 1). At the end of each surgical procedure, participants were asked to complete the Surgery Task Load Index (SURG-TLX) [3], which is a subjective questionnaire to evaluate the surgical workload based on the NASA-TLX questionnaire. Surgical procedures were also rated by an expert surgeon using the Global Operational Assessment of Laparoscopic Skills (GOALS) method [4], a tool for objective evaluation of surgical performance.

For the statistical analysis, correlations between the surgeons' physiological and kinematic parameters, surgical workload and quality of surgical performance were analysed. Exploratory data analysis was conducted in order to gain insight about collected data. When it comes to predictive ML model design, two possible objectives were taken into account: a classification model that could predict the specialty of a surgeon or the procedure they were performing based on hand accelerations during surgery, and a regression algorithm that could give the surgeon's a GOALS score also based on hand accelerations during surgery without the need of an expert surgeon's evaluation. Finally, a principal component analysis (PCA) was conducted in order to reduce the number of features to be considered in the design of the algorithm and to avoid the training of a predictive model with correlated predictors.



Figure 1. Experimental setup: Surgeons wearing a smartwatch on their dominant hand in order to register acceleration data from the smartwatch's IMU

3. Results and discussion

Thirteen novel gynaecologists and six novel urologists participated in this study with a total of 25 procedures performed. Surgeons showed a negative correlation between the depth perception and hand accelerations on the Z axis (orthogonal to the sphere of the smartwatch) and a positive correlation between the experience and the GOALS score.

When it comes to differences between surgical specialties, no statistically significant differences were found in the GOALS score or in the SURG-TLX between urologists and gynaecologists (Fig. 2). The first seven principal components obtained in the PCA explain 80.36% of the variability of the data. Besides, when representing each subject as a function of the variability of the first two principal component, urologists tend to vary in the direction of the second principal component and gynaecologists in the direction of the first and the second principal component (Fig. 3).

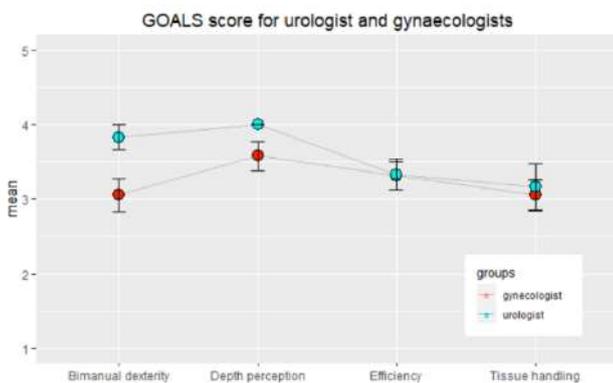


Figure 2. Mean GOALS score for each item for both groups

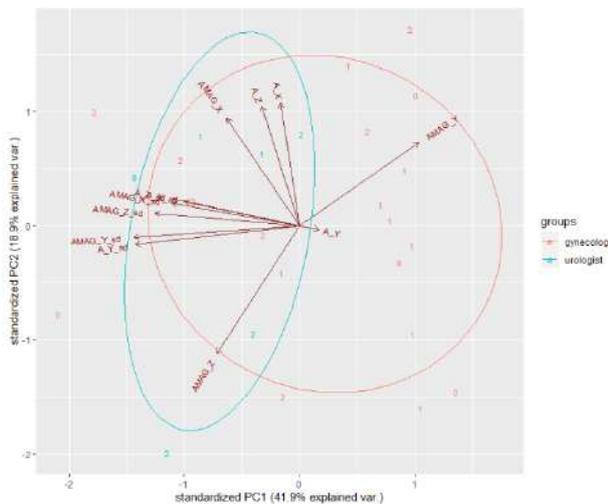


Figure 3. Subjects represented as a function of the first two principal components

One of the biggest challenges that this study presents is the data acquisition. Creating a big enough database with hand accelerations from surgeons with different levels of expertise performing real surgery is not only time consuming but also expensive, and it also requires a lot of planning and the effort from professionals from very diverse backgrounds. Having the amount of data to build a successful ML model might take years and many resources. Some of the data collection aspects such as data labeling, data cleaning and storage are not often taken into account and they are not only crucial but also the most time-consuming tasks.

4. Conclusions

Preliminary results showed that surgical workload and performance within novel gynaecologist is independent of the surgical task or procedure performed. Surgeons with better depth perception scores within the GOALS scale also presented fewer hand accelerations on the Z axis overall. PCA showed a promising setup in which an ML classification algorithm could be implemented. However, this statement is based on current data trends and might need to be updated by the time the required amount of data is acquired. More data would be needed to build an accurate classification model and to be able to obtain sufficient results to support the feasibility of developing an ML algorithm that can predict a surgeon's GOALS or SURG-TLX score for a particular procedure.

Acknowledgements

Authors declare no conflict of interest and would like to thank all the surgeons that participated in this study for their time and patience. This study has been partially funded by the MISTI Global Seed Funds, “la Caixa” Foundation, Junta de Extremadura (Spain), European Social Fund, European Regional Development Fund (ERDF) "A way to make Europe", and the Spanish Ministry of Science, Innovation and Universities through ERDF funds of the Intelligent Growth Operational Program (LCF/PR/MIT18/11830006, TA18023, PD18077, GR18199, CPI-2019-33-1-TRE -14).

References

- [1] SHAH, Rushabh; CHIRCU, Alina. IoT and AI in healthcare: A systematic literature review. *Issues in Information Systems*, 2018, vol. 19, no 3.
- [2] ROH, Yuji; HEO, Geon; WHANG, Steven Euijong. A survey on data collection for machine learning: a big data-ai integration perspective. *IEEE Transactions on Knowledge and Data Engineering*, 2019.
- [3] WILSON, Mark R., et al. Development and validation of a surgical workload measure: the surgery task load index (SURG-TLX). *World journal of surgery*, 2011, vol. 35, no 9, p. 1961-1969.
- [4] HOGLE, Nancy J., et al. Evaluation of surgical fellows' laparoscopic performance using Global Operative Assessment of Laparoscopic Skills (GOALS). *Surgical endoscopy*, 2014, vol. 28, no 4, p. 1284-1290.

Citometría de microflujo para estudios bacterianos

I. Rico-Gómez¹, M. Cañadas¹, S. Laguna¹, C. Gómez-Cruz¹, A. Muñoz-Barrutia^{1,2}

¹ Departamento de Bioingeniería e Ingeniería Aeroespacial, Universidad Carlos III de Madrid, Leganés, España

² Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, España

Resumen

Evaluar la eficacia de los fármacos para tratar una enfermedad bacteriana exige como paso final el cultivo de bacterias, seguido del recuento de unidades formadoras de colonias. El objetivo de este proyecto es acelerar este proceso mediante el desarrollo de una plataforma de microflujo para el conteo directo de bacterias. Dicha plataforma fusiona las ventajas de la citometría de flujo, la microfluídica y el procesamiento de imágenes. El software de aprendizaje profundo utilizado para analizar las imágenes genera imágenes sintéticas para el entrenamiento. Los resultados, con una precisión del 91.7%, demuestran el éxito de la plataforma en la focalización, detección y cuantificación de bacterias.

1. Introducción

Para determinar la eficacia de un compuesto es necesario contabilizar el número de bacterias restantes en la muestra tras la aplicación de dicho agregado. Un procedimiento común utilizando modelos animales, implica el cultivo de una muestra con un compuesto específico dentro de un biorreactor durante dos semanas, seguido de la cuantificación de las Unidades Formadoras de Colonias (UFC). Sin embargo, este proceso lleva mucho tiempo y las UFC no siempre son representativas de la población total de bacterias. En consecuencia, la detección de bacterias es compleja, lenta e inexacta, lo que ralentiza el desarrollo y la evaluación de nuevos fármacos.

Este proyecto busca desarrollar un método más rápido y preciso para el recuento de bacterias. Mediante la fusión de las ventajas de la citometría de flujo, la microfluídica y el procesamiento de imágenes de microscopía, se ha desarrollado una plataforma de citometría de microflujo. Esta plataforma permitiría la evaluación de la carga bacteriana de una manera mucho más rápida, ya que las bacterias se detectan y cuentan simultáneamente. Trabajos previos que planean plataformas similares para el conteo de partículas, por ejemplo, Nawaz *et al.* [1] que enfoca y analiza en 3D partículas de un tamaño de 2.5 μm utilizando un microscopio confocal o el un trabajo previo de nuestro grupo, Laguna *et al.* [2] que detecta partículas fluorescentes de 1 μm utilizando un microscopio de hoja plana.

En este proyecto se emplean dos diseños de microdispositivos diferentes para lograr el enfoque de las bacterias. La adquisición de imágenes se realiza mediante un microscopio invertido, y el procesamiento de imágenes se realiza con un método de aprendizaje profundo que permite el etiquetado de todas las partículas presentes.

El principal objetivo de este trabajo es contribuir en el tratamiento de las enfermedades bacterianas, ofreciendo una plataforma novedosa capaz de reducir el tiempo y el coste necesarios para la estimación de la carga bacteriana.

2. Materiales y Métodos

2.1. Diseño de los Microdispositivos

El objetivo principal es la cuantificación de las bacterias de manera individual, pero debido a su pequeño tamaño, un solapamiento en

el plano x-y-z es difícil de evitar. Durante este proyecto se han utilizado dos diseños de microdispositivos. Un diseño se utiliza para la realización de los primeros experimentos y consiste en un diseño sin enfoque, el chip en Y, mientras que el otro es un modelo de enfoque 2D capaz de enfocar el flujo en el plano x-y. La discretización completa de las partículas debe obtenerse escaneando la muestra en la dirección z.

El principal motivo para utilizar el chip sin enfoque es el hecho de que aunque las partículas no pueden ser discretizadas, es más fácil determinar si las partículas atraviesan el canal, porque este modelo tiene canales más anchos. El diseño del chip con enfoque 2D fue propuesto por Inatomi *et al.* [3] y permite la orientación de las partículas utilizando dos flujos laterales que hacen que la suspensión se mueva por el centro del canal. Ambos diseños fueron fabricados en PDMS mediante técnicas de litografía suave. La Figura 1 muestra ambos diseños con sus respectivas dimensiones.

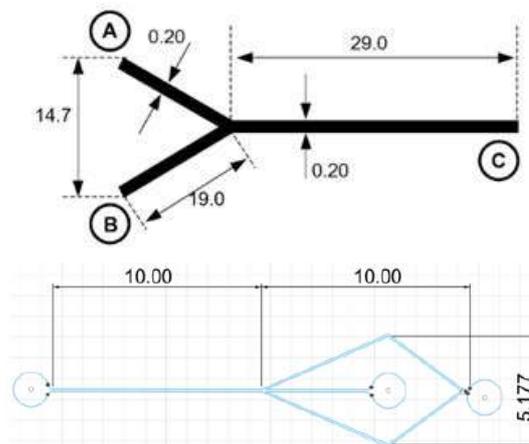


Figura 1. Chip en Y con dimensiones en milímetros (Arriba). Chip de enfoque 2D, con dimensiones en milímetros (Abajo). Adaptación del diseño de Inatomi *et al.* [3].

2.2. Simulaciones de flujo

Los flujos óptimos se determinan realizando simulaciones de flujo con el software ANSYS [4]. Las simulaciones se obtienen para el chip en Y y el chip de enfoque 2D. Todos los flujos cumplen con las condiciones necesarias para ser flujos laminares, con un número de Reynolds inferior a 2000.

El flujo mínimo está limitado por las bombas utilizadas, y es igual a 0,73 $\mu\text{l/hora}$ para la bomba BS-800 y 1,26 pl/min para la bomba Elite 11 de Harvard Apparatus. Estos flujos corresponden a un radio de jeringa igual a 2,35 mm. La velocidad máxima, V_{max} , está determinada por la altura, h_{max} , del campo de visión, que corresponde al rectángulo alrededor del canal principal, y los fotogramas por segundo, fr. La h_{max} es de 1,57 $\mu\text{m/píxel}$ * 1024 píxeles, y la tasa de fotogramas proporcionada por la cámara del microscopio es de 40 fotogramas/segundo. Los flujos máximos, F_{max} se obtienen a partir de dicha Velocidad Máxima, V_{max} , y teniendo en cuenta el

área de sección transversal del microdispositivo, A_{st} . La Tabla 1 muestra el F_{max} para cada microdispositivo, que ha sido obtenido con las siguientes fórmulas:

$$\begin{aligned} V_{max} &= h_{max} * fr, \\ F_{max} &= V_{max} * A_{st}, \end{aligned}$$

Cuando se trabaja con bacterias, hay que tener en cuenta que el medio del Caldo de Lisogenia (LB) donde se cultivan las bacterias tiene una viscosidad mayor que la del agua. La relación entre la viscosidad y el caudal volumétrico viene determinada por la Ley de Poiseuille. Para un flujo laminar ($Re < 2000$), el caudal volumétrico, V_{fr} , viene dado por la diferencia de presión (AP) dividida por la resistencia viscosa (R), según la ecuación:

$$V_{fr} = \frac{AP}{R}. \quad (1)$$

Microdispositivo	V_{max}	A_{st}	F_{max}
Chip en Y	0.064 m/s	0.02 mm ²	76.8 µl/min
Chip enfoque 2D	0.064 m/s	0.001 mm ²	3.84 µl/min

Tabla 1. Máximo flujo para cada microdispositivo.

Esta resistencia depende linealmente de la viscosidad cinemática Kv y de la longitud L . Cuanto mayor sea la viscosidad, mayor será la resistencia y, en consecuencia, menor será el caudal volumétrico. La resistencia para canales rectangulares, con una altura h mucho menor que la anchura w se describe mediante la siguiente ecuación:

$$R = \frac{12 * Kv * L}{w * h^3} (h \ll w). \quad (2)$$

Utilizando el software ANSYS, se calcularon simulaciones con distintos flujos en estos rangos para obtener los perfiles de velocidad y presión. Teniendo en cuenta que la viscosidad del fluido corresponde a la del medio bacteriano y no a la del agua. Este valor se encuentra en el estado máximo de crecimiento de las bacterias alrededor de 0,02 m²/s [5]. En base a los resultados de dichas simulaciones se seleccionaron los flujos óptimos para probar experimentalmente.

2.3. Preparación de las soluciones

El objetivo es utilizar la plataforma para el recuento de bacterias, pero al principio se utilizan distintas soluciones con microesferas fluorescentes, con un tamaño de 1 µm y un máximo de excitación de 468 nm con luz azul.

La concentración utilizada para el microdispositivo en Y fue de 0,001 g/ml, mientras que para el microdispositivo 2D se utilizó una concentración de 0,002 g/ml. El principal objetivo es encontrar las concentraciones que permiten la visualización de las partículas de la manera más individualizada posible.

Después de caracterizar los flujos con soluciones de microesferas, se repitieron los mismos experimentos con bacterias *E.Coli*, con un tamaño entre 1 y 2 µm. La tinción de las bacterias se realizó con SYTO 9, con un máximo de excitación azul de 485 nm. Para este proyecto se prepararon eppendorfs que contenían 1,5 ml de solución de bacterias y 4,5 µl de SYTO 9, que se incubaron durante 30 minutos a 37°. Además de la solución bacteriana inicial, se obtuvieron diluciones a 10X, 100X y 1000X diluyendo la solución original con la cantidad adecuada de medio LB.

El siguiente paso tras la preparación de las soluciones consiste en conectar las entradas de los microdispositivos a las jeringas, previamente llenadas con las soluciones y colocadas en las bombas.

3. Adquisición de imágenes

El microscopio utilizado para la adquisición de imágenes es el LEICA DMI8, un microscopio invertido. La fuente de luz y el condensador están en la parte superior apuntando hacia abajo, y los objetivos y la torreta apuntan hacia arriba. Los componentes de un microscopio invertido son los mismos que los de otros microscopios, pero con una disposición diferente. El software empleado es LAS x (Leica Application Suite X). Este software permite la adquisición a alta velocidad y el análisis de datos durante la adquisición. LAS X cuenta con una serie de módulos específicos que permiten configurar el microscopio para casi cualquier aplicación.

El objetivo utilizado es el HC PL FLUOTAR, un objetivo semiapocromático, con un aumento de 10x. La Apertura Numérica es de 0,32 y la distancia de trabajo es de 11,33 mm. El ocular empleado es el HC PLAN 10x, con un aumento de la imagen de 10X. El campo de visión es de 20 mm.

La cámara instalada es una Hamamatsu ORCA FLASH 4.0 V3. Esta cámara tiene un sensor CMOS científico (sCMOS). La velocidad de adquisición de fotogramas está limitada por la interfaz USB 3.0, con un máximo de 30 campos por segundo (fps). Sin embargo, un Camera Link puede acelerar la lectura hasta 100 fps. La frecuencia de imagen seleccionada para la adquisición fue de 40 fps. La cámara también ofrece la posibilidad de binning, transformando imágenes de 2048 píxeles x 2048 píxeles en imágenes de 1024 píxeles x 1024 píxeles. Durante este proyecto se han adquirido imágenes utilizando un binning de 2 píxeles x 2 píxeles. El tiempo de exposición, tiempo que el sensor de la cámara está expuesto a la luz adquiriendo una imagen, es igual a 0,02 ms.

4. Detección de las partículas en las imágenes

Para marcar todas las partículas presentes en las imágenes, tanto microesferas como bacterias, se seleccionó un software de aprendizaje profundo, DeepTrack 2.0 [6]. Este software proporciona una interfaz gráfica fácil de usar, junto con scripts en cuadernos Jupiter, diseñados para distintos fines.

El primer paso para el análisis de las imágenes con este software consiste en un preprocesamiento de las imágenes con ImageJ. Para eso se aplica un filtro medio de radio 5 para suavizar las imágenes, eliminando el ruido. A continuación, el script DeepTrack 2.0 para el seguimiento de múltiples partículas fue seleccionado, ya que el objetivo es localizar todas las partículas fluorescentes en las imágenes. Para esta aplicación, se utiliza una red en U [6].

Los principales pasos para realizar la detección de partículas utilizando DeepTrack2.0 están ilustrados en la Figura 2 y el procedimiento completo se explica a continuación:

- 1) **Definición de las partículas y sistema óptico:** Se crean partículas puntuales sintéticas, definidas por su intensidad y su posición. La definición del sistema óptico toma un conjunto de partículas y las convolucionan con la función pupilar del sistema óptico, en este caso el seleccionado es fluorescencia. También se seleccionan parámetros como la apertura numérica, la resolución, el aumento y la longitud de onda del rayo láser para crear un sistema con los mismos parámetros que el microscopio utilizado para la adquisición de imágenes.
- 2) **Normalización y definición del ruido:** La imagen se normaliza reescalándola para que esté contenida entre dos números aleatorios en el rango (0,1). También se añade ruido de Poisson con una relación señal a ruido (SNR) aleatoria entre 3 y 4.
- 3) **Imágenes de entrenamiento:** Se crean imágenes sintéticas con un número aleatorio de partículas fluorescentes, añadiendo ruido y normalizándolas. Se

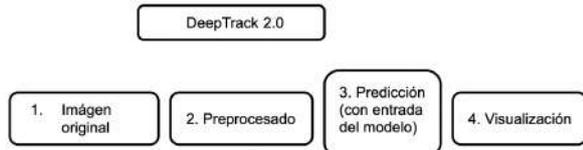


Figura 2. Pipeline de los pasos principales del algoritmo de aprendizaje profundo para la detección de partículas: (1) Imagen Original, (2) Preprocesado (Filtro medio de radio 5), (3) Predicción (con entrada del modelo), (4) Visualización

utiliza un bucle para generar en cada repetición una imagen con características aleatorias.

- 4) **Imágenes de destino:** Se crean las imágenes sintéticas objetivo que se utilizarán en el entrenamiento. Estas imágenes contienen partículas de radio 2 en la misma posición que las partículas de las imágenes de entrada.
- 5) **Definición del modelo de red neuronal:** El modelo de red seleccionado en esta aplicación es una red U, creada con la función U-net. Los falsos negativos se penalizan diez veces más que los falsos positivos, ya que en la imagen hay muchos más 0s que 1s.
- 6) **Entrenamiento del modelo:** La red se entrena con las imágenes generadas.

Finalmente, el modelo se aplica sobre la imagen preprocesada para obtener la imagen de predicción (partículas detectadas). Dicha imagen se muestra junto con la imagen de entrada.

5. Resultados

En esta sección se muestran los resultados obtenidos en las simulaciones de flujo. Así como el resultado de la detección de las partículas fluorescentes y las bacterias sobre los videos de microscopía adquiridos sobre los microdispositivos estudiados.

5.1. Simulaciones de flujo

Para cada microdispositivo, se obtuvieron perfiles de distribución de velocidad y de presión para los diferentes flujos antes de ser probados experimentalmente.

Para el chip en Y la configuración de flujo que mejor se adapta al propósito principal, es decir, la visualización de partículas individuales a través de los canales de los microdispositivos, resultó ser de 7,5 $\mu\text{l}/\text{min}$, con un flujo de entrada de 2,5 $\mu\text{l}/\text{min}$, y un flujo lateral de 5 $\mu\text{l}/\text{min}$. La velocidad máxima alcanzada para este caudal fue de 0,012 m/s, y la presión máxima de 9.790 Pa.

La Figura 3 muestra la distribución de la velocidad y de la presión para este caudal.

En el caso de el chip 2D con enfoque, el flujo óptimo resultó ser un flujo de entrada de 0,65 $\mu\text{l}/\text{min}$ y un flujo lateral de 2 $\mu\text{l}/\text{min}$. La presión máxima para este flujo es de 26.981 Pa.

La figura 4 representa los perfiles de velocidad y presión para este flujo óptimo. La velocidad máxima alcanzada para esta configuración de flujo es de 0,047 m/s.

El enfoque es de 5 μm , ya que debido a las limitaciones de la versión de estudiante del software ANSYS el tamaño de la celda no puede ser inferior a ese valor.

La Figura 5 muestra la distribución de partículas en este chip para el flujo seleccionado.

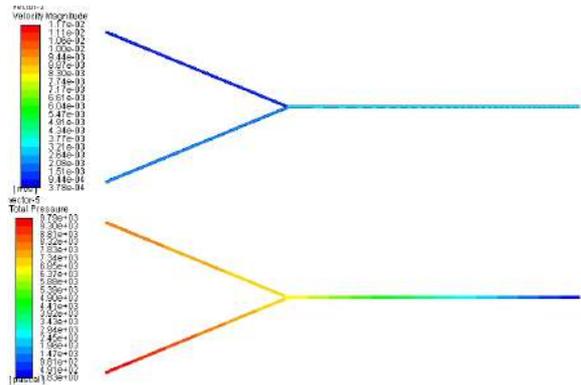


Figura 3. Distribución de la velocidad en el microdispositivo en Y (Arriba), y distribución de la presión (Abajo), con un flujo de entrada de 2,5 $\mu\text{l}/\text{min}$, y un flujo lateral de 5 $\mu\text{l}/\text{min}$.

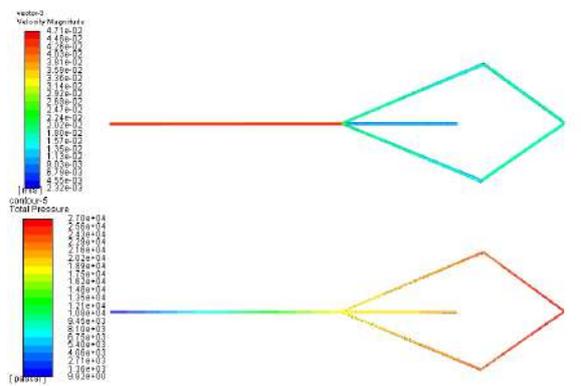


Figura 4. Distribución de la velocidad en los canales de enfoque 2D para un fluido con características de agua y una configuración de flujo dada por un flujo de entrada de 0,65 $\mu\text{l}/\text{min}$ y uno lateral de 2 $\mu\text{l}/\text{min}$ (Arriba). Distribución de la presión total para la misma configuración de flujo (Abajo).

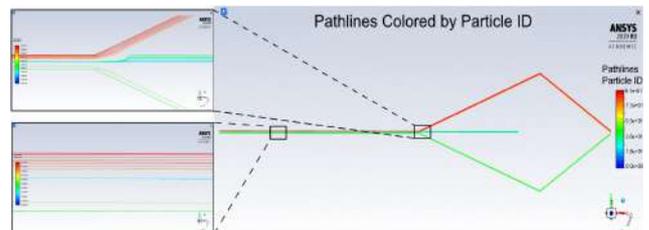


Figura 5. Distribución de las partículas del modelo de enfoque 2D, mostrando las regiones de enfoque y focalización. Las bacterias corresponden a las líneas de corriente en azul.

5.2. Imágenes adquiridas y procesadas

Las imágenes obtenidas para ambos microdispositivos se adquirieron con el microscopio LEICA DMI8. Después de probar los distintos flujos considerados óptimos, las imágenes que permitían una mejor visualización de las microesferas o bacterias fueron seleccionadas para su posterior procesado.

Las Figura 6 y 7 muestran un ejemplo de imágenes consecutivas para una solución de microesferas utilizando microdispositivos en Y y de enfoque 2D, respectivamente.

Utilizando el software DeepTrack 2.0 las partículas presentes en cada imagen son detectadas y marcadas. La Figura 8 muestra un ejemplo de imagen adquirida utilizando flujo de entrada de 2,5 $\mu\text{l}/\text{min}$ y un flujo lateral de 5 $\mu\text{l}/\text{min}$, utilizando una solución de microesferas de concentración 0,001 g/ml.

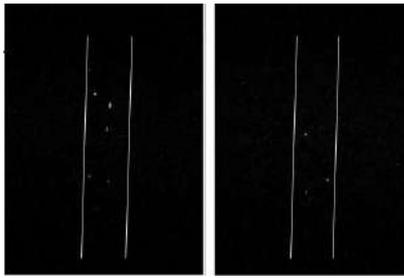


Figura 6. Ejemplo de fotogramas consecutivos obtenidos para una solución de microesferas con el microdispositivo en Y, con un flujo lateral de $5 \mu\text{l}/\text{min}$, y un flujo de entrada de $2,5 \mu\text{l}/\text{min}$, concentración de $0,001 \text{ g}/\text{ml}$. Los canales están dibujados para su visualización.

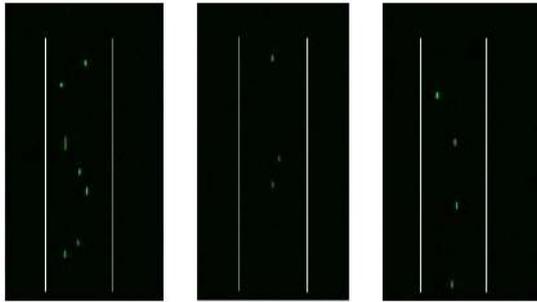


Figura 7. Ejemplo de imágenes obtenidas para una solución de microesferas con el microdispositivo de enfoque 2D. Las imágenes están en el mismo orden de adquisición, de izquierda a derecha, con un tiempo entre cada una de $0,025 \text{ s}$. Las partículas están mejor enfocadas en la última imagen (derecha). Los canales están dibujados para su mejor visualización.

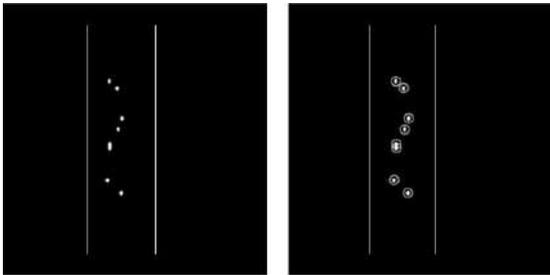


Figura 8. Ejemplo de imagen procesada con Deep Track 2.0, utilizando un microdispositivo en Y, y una solución de microesferas con un flujo de entrada de $2,5 \mu\text{l}/\text{min}$ y un flujo lateral de $5 \mu\text{l}/\text{min}$, con una concentración de $0,001 \text{ g}/\text{ml}$. Imagen preprocesada (izquierda), e imagen de salida con las partículas detectadas (derecha). Los canales están dibujados para su mejor visualización.

Con el fin de obtener la precisión del software de procesado, las partículas presentes en 100 imágenes consecutivas adquiridas con el microdispositivo en Y, y 100 con el microdispositivo de enfoque 2D, son analizadas manualmente. El objetivo es comparar los resultados obtenidos con la verdad fundamental. Se obtuvo el número de verdaderos positivos (V_P), verdaderos negativos (V_N), falsos positivos (F_P), y falsos negativos (F_N), y la precisión, P_{cs} , se calculó con la siguiente fórmula:

$$P_{cs} = \frac{V_P + V_N}{V_P + V_N + F_P + F_N} \quad (3)$$

El resultado fue igual a 91.7%. Este valor podría ser mayor en un futuro con una mejor adaptación del software DeepTrack 2.0, generando imágenes de entrenamiento más similares a las reales.

6. Discusión y conclusiones

El objetivo principal del proyecto es acelerar el desarrollo de nuevos fármacos reduciendo el tiempo necesario para probar su eficacia. La plataforma desarrollada cumple este objetivo al evitar el conteo de las unidades formadoras de colonias (UFC).

La plataforma se ha probado con dos diseños de microdispositivos, el chip en Y que no permite el enfoque de las partículas y el chip 2D que sí genera una corriente enfocada. Las simulaciones obtenidas para caracterizar los flujos óptimos antes de las pruebas experimentales también demostraron un enfoque igual a $5 \mu\text{m}$ utilizando el chip de enfoque 2D. Dado que garantizar el enfoque de las partículas es crucial para una cuantificación precisa, futuras investigaciones deben orientarse a la mejora de la caracterización del modelo de enfoque.

Los primeros experimentos se llevaron a cabo utilizando suspensiones de microesferas, pero también se ha probado la plataforma con soluciones bacterianas, siendo los resultados satisfactorios. No obstante, en trabajos futuros se debería controlar el cultivo de las bacterias midiendo la densidad óptica (DO) de las soluciones. De este modo, se podrían determinar las concentraciones que conducen a la detección precisa de las partículas.

La adquisición de imágenes se ha llevado a cabo con un microscopio invertido, pero otro tipo de microscopio podría adaptarse a la plataforma, dada su versatilidad. El procesamiento de las imágenes utilizando el software de aprendizaje profundo DeepTrack 2.0 ha permitido la detección de partículas presentes en cada fotograma con un 91.7% de precisión. No obstante, los resultados pueden mejorar generando imágenes de entrenamiento más similares a las imágenes obtenidas experimentalmente.

El proyecto se ha llevado a cabo satisfactoriamente, demostrando que la plataforma podría utilizarse en un futuro para la detección y cuantificación de bacterias. Esto contribuiría a una disminución en el tiempo necesario para evaluar la eficacia de un compuesto ante una enfermedad bacteriana.

7. Agradecimientos

Este proyecto ha sido parcialmente financiado por el Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación, con el proyecto de referencia PID2019-109820RBI00 MINECO/FEDER, UE, cofinanced by European Regional Development Fund (ERDF) 'A way of making Europe.'

Bibliografía

- [1] A. A. Nawaz, X. Zhang, X. Mao, J. Rufo, S.-C. S. Lin, F. Guo, Y. Zhao, M. Lapsley, P. Li, J. P. McCoy *et al.*, "Sub-micrometer-precision, three-dimensional (3d) hydrodynamic focusing via "microfluidic drifting"," *Lab on a Chip*, vol. 14, no. 2, pp. 415–423, 2014.
- [2] S. Laguna, C. Gomez-Cruz, I. Albert-Smet, J. J. Vaquero, J. Ripoll, and A. Muñoz-Barrutia, "Citometría de flujo con hoja plana para el conteo de bacterias de tuberculosis", booktitle = "libro de actas del xxxviii congreso anual de la sociedad española de ingeniería biomédica (caseib'20)," pp. 265–268, 2020.
- [3] K.-I. Inatomi, S.-I. Izuo, and S.-S. Lee, "Application of a microfluidic device for counting of bacteria," *Letters in applied microbiology*, vol. 43, no. 3, pp. 296–300, 2006.
- [4] A. Inc., "Ansys fluent," same <https://www.ansys.com/products/fluids/ansys-fluent>.
- [5] K. Sklodowska, P. R. Debski, and M. et al., "Simultaneous measurement of viscosity and optical density of bacterial growth and death in a microdroplet," *Micromachines*, vol. 9, no. 5, p. 251, 2018.
- [6] B. Midtvedt, S. Helgadottir, A. Argun, J. Pineda, D. Midtvedt, and G. Volpe, "Quantitative digital microscopy with deep learning," *arXiv preprint arXiv:2010.08260*, 2020.

Información a priori obtenida con aprendizaje profundo para su uso en algoritmo de reconstrucción iterativo

A. Piol^{1,2}, C. F. del Cerro^{1,2}, J. García-Blas³, M. Desco^{1,2,4,5}, M. Abella^{1,2,5}

¹ Departamento de Bioingeniería e Ingeniería Aeroespacial, Universidad Carlos III, Madrid, Spain

² Instituto de Investigación Sanitaria Gregorio Marañón (IiSGM), Madrid, Spain

³ Departamento de Informática, Universidad Carlos III de Madrid, Madrid, España

⁴ Centro de Investigación en Red de Salud Mental (CIBERSAM, CIBER CB07/09/0031), 28007 Madrid, Spain

⁵ Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), Madrid, Spain

Resumen

En sistemas de tomografía computarizada tradicionales, los datos se adquieren girando fuente y detector alrededor de la muestra completando un rango angular de 360 grados y obteniendo, generalmente, más de 360 proyecciones. Sin embargo, existen situaciones donde solo es posible obtener un número limitado de proyecciones. Un ejemplo es el gating respiratorio, donde las proyecciones clasificadas en una determinada fase del ciclo pueden no ser en número suficiente para obtener una reconstrucción de alta calidad. Esta limitación de datos resulta en artefactos específicos en la reconstrucción y, consecuentemente, en pérdida de detalles importantes para el diagnóstico clínico. Una solución para compensar la falta de datos es utilizar información a priori en algoritmos iterativos. El algoritmo PICCS permite la reconstrucción del volumen con un número limitado de proyecciones, incorporando como información a priori una reconstrucción con datos completos de la muestra.

En este trabajo se propone una metodología de reconstrucción de datos incompletos que incorpora en PICCS un paso de obtención de esta información a priori con metodología de aprendizaje profundo.

1. Introducción

En sistemas de tomografía computarizada, el par fuente-detector gira simultáneamente alrededor del paciente, cubriendo generalmente 360 grados, adquiriendo un conjunto de proyecciones (radiografías planas) desde diferentes posiciones angulares. Cuando el conjunto de proyecciones cumple determinados requisitos, de número y rango angular, la reconstrucción del volumen bajo análisis se puede hacer utilizando algoritmos tradicionales basados en retroproyección filtrada [1]. Sin embargo, existen diferentes situaciones en las que el número de proyecciones adquiridas no alcanza el mínimo para poder utilizar estos algoritmos. Un ejemplo es el uso de arcos en C para obtener imagen tomográfica en quirófano, que presentan limitaciones mecánicas que no permite cubrir todo el rango angular. Otro caso de limitación de datos es la selección de un paso angular grande como medio para reducir las dosis de radiación. En estos casos, el resultado de reconstrucciones con algoritmos convencionales basado en FDK está afectado por dos tipos de artefactos: streaks, en el caso de un paso angular demasiado grande, y

distorsión de los bordes, en el caso de un rango angular limitado.

Una forma de compensar la falta de datos es la incorporación de información a priori en algoritmos iterativos. Ejemplos de información a priori son la no-negatividad de la solución o la restricción del espacio de búsqueda de la solución en una determinada área del campo de visión (FOV), en la que sabemos que se encuentra la muestra. En [7], presentamos un algoritmo de reconstrucción que utiliza el contorno de la muestra para definir la región de búsqueda de la solución: el algoritmo definido support-constraint limited data (SCoLD). En el artículo, la máscara se obtiene mediante la adquisición de la superficie de la muestra con un escáner de luz 3D (Artec Eva) que se convierte en máscara binaria para poder ser utilizada en SCoLD. Sin embargo, la adquisición de esta superficie puede complicar el protocolo clínico e incluso ser imposible por la existencia tejido cubriendo al paciente. Además, el utilizo del algoritmo SCoLD soluciona los artefactos causados por un rango angular limitado, mientras no funciona bien con artefactos debidos a pasos angulares altos (streaks).

Una alternativa que se ha propuesto para la reconstrucción de datos limitados en rayos X se basa en técnicas de aprendizaje profundo [2, 3]. Estos trabajos tienen limitaciones en resolución espacial, con un tamaño máximo de imagen de 256×256 píxeles. Por otra parte, para evitar sesgos que resulten en estructuras en la imagen final que realmente no están presentes en los datos, es necesario disponer de una base de datos reales grande. Otra posible información a priori utilizada en algoritmos iterativos es la homogeneidad de los tejidos. Esta característica se puede conseguir con métodos de regularización basados en la función de Total Variation (TV) [2-4], para casos de adquisiciones con paso angular alto (artefacto de streaks). Para estos problemas se ha propuesto incorporar como información previa una reconstrucción con datos completos. En [4], se utiliza una reconstrucción con datos completos del mismo paciente en un estudio de gating respiratorio, obtenida a partir de las proyecciones de las diferentes fases respiratorias.

La Figura 1 muestra en ejemplo de resultados aplicando el algoritmo PICCS a un conjunto de datos de 120 proyecciones en un rango angular de 360 grados, que permite reducir los artefactos y alcanzar una calidad de imagen muy parecida a la de un TAC estándar (360 proyecciones con 360 grados de rango angular).

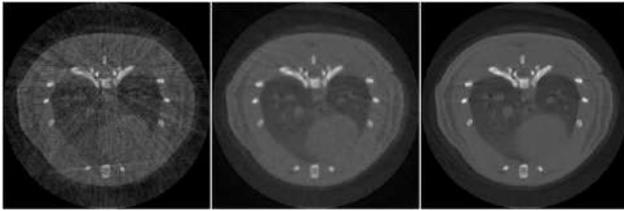


Figura 1. Corte axial de un estudio de datos limitados de un pequeño animal reconstruido con FDK (izquierda) y PICCS (centro). Reconstrucción con datos completos utilizando algoritmo FDK (derecha).

En [5] la información a priori (prior) es la reconstrucción FDK utilizando todas las proyecciones adquiridas. Esto se consigue una reconstrucción con una buena textura en los tejidos, pero siendo una imagen borrosa, sobretodo en correspondencia de pulmones y diafragma. Sin embargo, existen situaciones donde no se dispone de la imagen reconstruida con datos completos y no es posible obtenerla con el conjunto de datos disponible. En este trabajo se propone eliminar los problemas derivados de la falta de información a priori necesaria para el método PICCS, obteniéndola usando técnicas de aprendizaje profundo.

2. Método propuesto

El método propuesto consiste en tres pasos: (1) reconstrucción preliminar FDK con artefactos, que será la entrada a la red, (2) obtención del prior de PICCS desde esta reconstrucción preliminar y (3) uso de este prior en PICCS para obtener la reconstrucción final sin artefactos. Como el prior es necesario para aportar información sobre la textura y no tanto para detalles, se prueban dos resoluciones: una es la resolución original, que llamaremos alta resolución y la otra obtenida submuestreando, la cual llamaremos baja resolución.

Para la obtención del prior, se utiliza U-net, red propuesta para segmentación en [6], con el Error Cuadrático Medio (MSE) como función de coste. Como parámetros de entrenamiento se utilizan una tasa de aprendizaje de 10^{-3} durante 30 épocas y el optimizador Adam [7]. Los datos de entrada de la red son cortes 2D de dos tipos de tamaño: $320 \times 320 \times 1$ (baja resolución) y $656 \times 656 \times 1$ (alta resolución). Para el proceso de entrenamiento y validación del modelo se ha utilizado Python (versión 3.8).

La red entrenada se basa en la arquitectura U-net con 5 niveles. Para el entrenamiento a baja resolución, en el último nivel hay un tamaño de imagen de 20×20 por 512 imágenes de características. Para el entrenamiento a alta resolución, en el último nivel hay un tamaño de imagen de 41×41 por 512 imágenes de características. La Figura 2 muestra un esquema de la arquitectura de la red utilizada, para el caso de baja resolución ($320 \times 320 \times 1$).

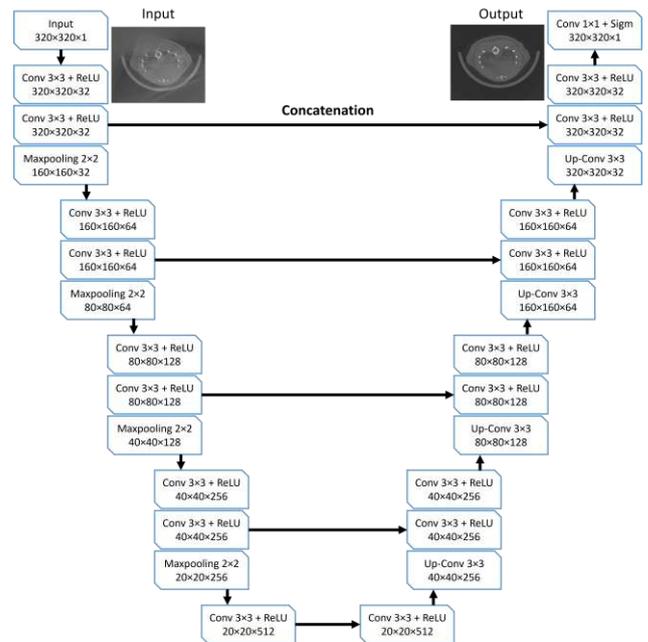


Figura 2. Arquitectura de la red utilizada con tamaño de entrada $320 \times 320 \times 1$.

Para poder utilizar la imagen de prior en el algoritmo PICCS, las imágenes en salida a la red en el caso de baja resolución se redimensionan al tamaño de la reconstrucción final.

3. Generación de la base de datos

La base de datos para el entrenamiento del modelo se ha creado a partir de cuatro estudios de roedores adquiridos con el sistema micro CT Add-on de SEDECAL [8]. Cada estudio está formado por 360 proyecciones de 768×486 pixeles obtenidas en un rango angular de 360 grados, distribuidas uniformemente. La reconstrucción de referencia se ha obtenido a partir de los datos completos con un método basado en FDK [9], dando lugar a un volumen de tamaño $656 \times 656 \times 486$.

Se han simulado con FuxSim [10], un software dedicado a la simulación de adquisición y reconstrucción de imágenes para sistemas de rayos-x, cuatro estudios eliminando proyecciones de manera aleatoria para cada uno de los escenarios de datos limitados que se muestran en la Figura 3: (A) Rango angular limitado, (B) número de proyecciones limitado y (C) rango angular y número de proyecciones limitados.

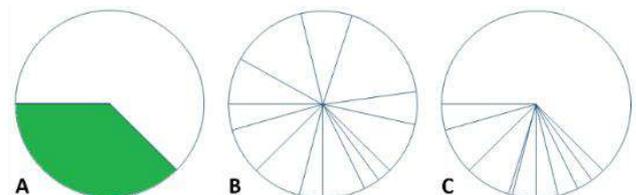


Figura 3: Esquema de selección de rango angular y proyecciones para los casos de rango angular limitado (A), número de proyecciones limitado (B) y rango angular y número de proyecciones limitados (C).

El conjunto de datos está compuesto por entre 130 y 204 cortes centrales de cada estudio, obteniendo un total de 5194 imágenes, separados en 3947 para entrenamiento y 1247 para validación. Los cortes centrales se han seleccionado en base a la tipología de adquisición hecha, con el intento de excluir posibles artefactos en los extremos de las reconstrucciones. Se han evaluado dos resoluciones, una al tamaño original (alta resolución) y otra submuestreando con un factor de 2.05, resultando en un tamaño de 320×320 (baja resolución).

El *batch size* (número de imágenes que se puede analizar en cada época) es de 1 para el caso de alta resolución y de 4 para el caso de baja resolución. Debido a limitaciones de memoria, no ha sido posible evaluar valores más grandes de *batch size* que podrían mejorar el entrenamiento de la red.

La Figura 4 muestra ejemplos de cortes para cada escenario de datos limitados junto a la reconstrucción de referencia.

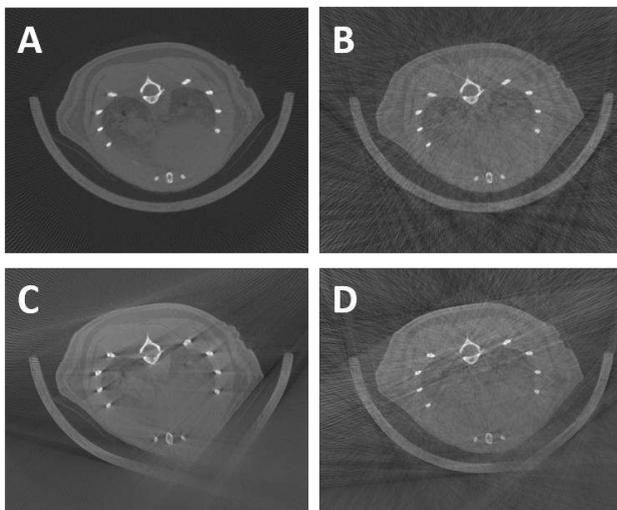


Figura 4: Corte axial central de la reconstrucción FDK con datos completos (A), datos con un número de proyecciones limitado (B), datos con rango angular limitado (C), datos con número de proyecciones y rango angular limitados (D).

4. Evaluación y resultados

Para evaluar el método propuesto se ha simulado con FuxSim una adquisición de rango angular completo con paso angular de 3 grados (120 proyecciones totales). La Figura 7A muestra el artefacto visible en la reconstrucción FDK con datos limitados: streaks en los cortes axiales.

Se ha evaluado la calidad de la reconstrucción con PICCS utilizando como prior: (1) la salida de la red en alta resolución (PICCS-prior_{AR}) y (2) la salida de la red en baja resolución redimensionada al tamaño de referencia (PICCS-prior_{BR}).

La Figura 5 muestra las imágenes prior de alta y baja resolución obtenidas con la red. La flecha blanca en la Figura 5-A indica como se pueden identificar mayormente las estructuras en la zona del diafragma y de los vasos dentro de los pulmones en el caso de entrenamiento de la red en baja resolución. En el panel inferior de la Figura 5, se enseña un detalle de las salidas de la red de alta y baja

resolución. Las flechas blancas en la Figura 5-D indican el contorno de los huesos y de la muestra, que se ven más definidos en el caso de entrenamiento de la red en alta resolución.

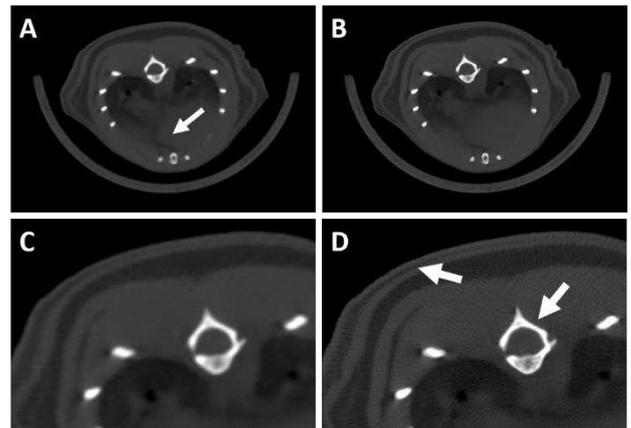


Figura 5: Panel superior: imagen en salida a la red entrenada en baja resolución (A) y en salida a la red entrenada en alta resolución (B). Panel inferior: detalle de imagen en salida a la red entrenada en baja resolución (C) y de imagen en salida a la red entrenada en alta resolución (D)

La Figura 6 muestra los resultados para el caso PICCS-prior_{BR}. Utilizando este prior baja resolución, la reconstrucción final con PICCS resulta en general borroso en el interior de la muestra, aunque se elimina el artefacto de streaks. La característica de borrosidad se traslada a la reconstrucción final desde el prior_{BR}.

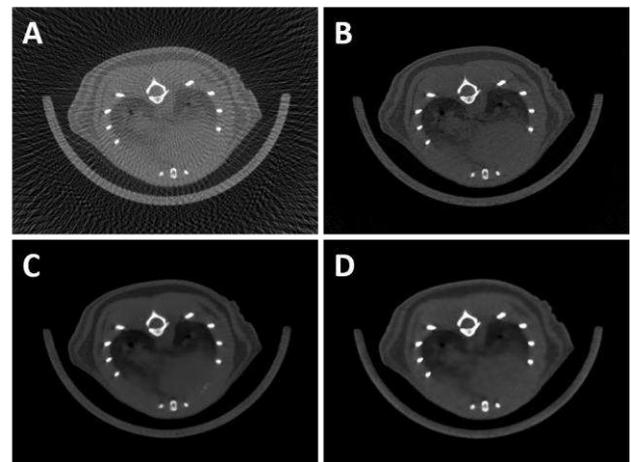


Figura 6. Corte axial de la reconstrucción preliminar con FDK con datos limitados (A), de referencia (B), en la salida a la red, baja resolución (C) y resultado en salida a PICCS (D). Rango angular de 360 grados y paso angular de 3 grados.

La Figura 7 muestra los resultados para el caso PICCS-prior_{AR}. El prior alta resolución no traslada ninguna característica de borrosidad en la reconstrucción final. Al contrario, los bordes y en los huesos están bien definidos, más que en el caso de baja resolución. Sin embargo, en la reconstrucción final, se pueden notar menos estructuras, o estructuras definidas en manera equivocada, en la zona del diafragma y de los vasos de los pulmones.

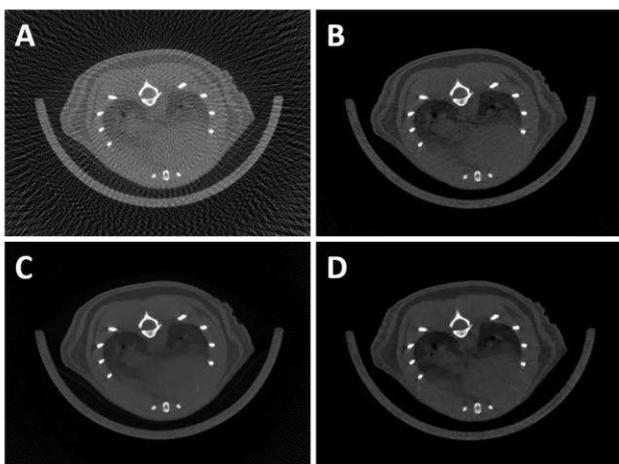


Figura 7: Corte axial de la reconstrucción preliminar con FDK con datos limitados (A), de referencia (B), en salida a la red, alta resolución (C) y resultado en salida a PICCS (D). Rango angular de 360 grados y paso angular de 3 grados.

5. Discusión y conclusiones

Este trabajo presenta una metodología de reconstrucción de datos incompletos basado en la combinación de PICCS, un algoritmo de reconstrucción que incorpora información a priori (prior), con una metodología de aprendizaje profundo que permite extraer ese prior. En estudios dinámicos, como por ejemplo estudios de gating respiratorio, el número total de proyecciones adquiridas permite la obtención de una reconstrucción con datos completos y, por consiguiente, un prior para PICCS de buena calidad. Sin embargo, en estudios estáticos esta reconstrucción no está disponible y el prior debe ser obtenido de otra manera. Este problema se encuentra sobretodo en situaciones donde se quiere reducir el número total de proyecciones adquiridas, con el intento de reducir las dosis de radiación dada al paciente. El método propuesto es adecuado para situaciones de este tipo, porque elimina la necesidad de tener estudios con datos completos de la muestra.

Aunque se ha entrenado la red para 3 tipologías de artefacto diferentes, hemos evaluado solamente el caso de los streaks. En futuro, el funcionamiento del modelo será evaluado para las otras tipologías de artefactos. Además, es posible aumentar la especificidad del modelo, entrenando la red solo para determinados artefactos (por ejemplo solo streaks o solo rango angular limitado).

Los parámetros de PICCS han sido ajustados heurísticamente, observando la calidad de la salida durante las iteraciones, con respecto a la reconstrucción de referencia. La idea futura es la de poder identificar automáticamente los valores de los parámetros en función de la tipología de datos en entrada, intentando, al final, juntar los dos problemas en un solo modelo. El modelo daría como salida el prior junto con los parámetros para algoritmo PICCS.

El estudio en baja resolución nos permite mejorar las prestaciones de la red cambiando el *batch size*, dado que los requisitos de memoria son menores. Eso permite la mejor identificación de las estructuras de la muestra. La

contra es un resultado es general más borroso con respecto a lo de alta resolución. La idea futura es solucionar el problema de las limitaciones en memoria, encontradas en el estudio en alta resolución, con un procesamiento en paralelo o utilizando parches.

El uso de métodos híbridos como el propuesto, que incluye un paso de reconstrucción iterativa, comporta tiempos de ejecución altos. La idea futura es una metodología que elimine el paso de reconstrucción iterativa, obteniendo de la red directamente la reconstrucción sin artefactos. Para ello, será necesario resolver los problemas de resolución espacial, que depende de los límites de memoria, y optimizar la fase de entrenamiento para evitar problemas de sobreentrenamiento.

Agradecimientos

Este trabajo ha sido financiado por el Instituto de Salud Carlos III, proyecto DTS17/00122. Además ha sido financiado por el Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación, proyecto "DPI2016-79075-R – AEI/FEDER, UE", cofinanciado por el Fondo Europeo de Desarrollo Regional (FEDER), "Una manera de hacer Europa". Además, ha sido financiado por la Comunidad de Madrid y la Universidad Carlos III de Madrid (proyecto DEEPCT-CM-UC3M) y por Crue-CSIC-Santander "Fondo Supera Covid19", proyecto RADC0V19 (uc3m + fibhgm). El CNIC está financiado por el Ministerio de Ciencia e Innovación (MCIN) y la fundación PRO-CNIC.

Referencias

- [1] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Amer.*, vol. 10, pp. 612–619, 1984.
- [2] H. Zhang, L. Li, K. Qiao, L. Wang, B. Yan, L. Li, et al., "Image prediction for limited-angle tomography via deep learning with convolutional neural network," *arXiv preprint arXiv:1607.08707*, 2016.
- [3] Y. Huang, S. Wang, Y. Guan, and A. Maier, "Limited angle tomography for transmission X-ray microscopy using deep learning," *Journal of Synchrotron Radiation*, vol. 27, 2020.
- [4] F. P. J. Abascal, M. Abella, E. Marinetto, J. Pascau, and M. Desco, "A novel prior- and motion-based compressed sensing method for small-animal respiratory gated CT," *PLoS One*, vol. 11, p. e0149841, 2016.
- [5] J. F. Abascal, M. Abella, A. Sisniega, J. J. Vaquero, and M. Desco, "Investigation of different sparsity transforms for the PICCS algorithm in small-animal respiratory gated CT," *PLOS one*, vol. 10, p. e0120140, 2015.
- [6] M. Buda, A. Saha, and M. A. Mazurowski, "Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm," *Computers in biology and medicine*, vol. 109, pp. 218–225, 2019.
- [7] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [8] J. J. Vaquero, S. Redondo, E. Lage, M. Abella, A. Sisniega, G. Tapias, et al., "Assessment of a New High-Performance Small-Animal X-ray Tomograph," *IEEE Trans Nucl Sci*, vol. 55, pp. 898–905, 2008.
- [9] M. Abella, J. J. Vaquero, A. Sisniega, J. Pascau, A. Udías, V. García, et al., "Software architecture for multi-bed FDK-based reconstruction in X-ray CT scanners," *Computer methods and programs in biomedicine*, vol. 107, pp. 218–232, 2012.
- [10] M. Abella, E. Serrano, J. Garcia-Blas, I. García, C. De Molina, J. Carretero, et al., "FUX-Sim: Implementation of a fast universal simulation/reconstruction framework for X-ray systems," *PloS one*, vol. 12, p. e0180363, 2017.

Método para la caracterización de la señal de presión arterial invasiva continua durante la reanimación cardiopulmonar

J. Urteaga Urizarbarrena¹, A. Elola Artano², E. Aramendi Ecenarro^{1,3}, P.O. Berve⁴, L. Wik⁴

¹ Departamento de Ingeniería de Comunicaciones, Universidad del País Vasco, 48013 Bilbao, España

² Departamento de Tecnología Electrónica, Universidad del País Vasco, 48013 Bilbao, España

³ Instituto de Investigación Sanitaria Biocruces Bizkaia, Hospital Universitario Cruces, 48903 Barakaldo, España

⁴ Norwegian National Advisory Unit on Prehospital Emergency Medicine, Hospital Universitario de Oslo, 0424 Oslo, Noruega

Resumen

La señal de presión arterial invasiva (PAI) y sus características calculadas a partir de la forma de onda son ampliamente usadas en la medicina para monitorizar el sistema cardiovascular de los pacientes. Se ha demostrado que la señal PAI puede ser de ayuda a la hora de guiar la terapia de reanimación cardiopulmonar (RCP) de pacientes en parada cardíaca. Los algoritmos actuales para la extracción de características de la señal PAI han sido diseñadas para pacientes hemodinámicamente estables y no se ha demostrado su funcionamiento en pacientes en parada cardíaca. El objetivo de este trabajo es diseñar un algoritmo para la detección de picos en la señal PAI y la extracción de características fisiológicas durante la terapia de RCP. La base de datos de este estudio cuenta con 119 segmentos de 32 pacientes con una duración total de 368 min. La señal PAI se preprocesa antes de detectar los picos de presión en la señal derivada discreta aplicando un umbral adaptativo. Después, se calculan los valores de presión sistólica y diastólica (PS y PD), presión de pulso (PP) y frecuencia cardíaca (FC). El método de detección de picos de presión reportó presenta unos valores mediana (rango intercuartílico) de Se/PPV/F1 de 97.9(4.2)/99.7(11.6) /97.8 (3.1). Los parámetros fisiológicos se midieron con un error absoluto de: 1.12 (3.13) mmHg en PS, 0.97 (0.76) mmHg en PD, 2.11 (4.50) mmHg en PP y 0.36 min⁻¹ en FC.

1. Introducción

La presión arterial invasiva (PAI) es ampliamente utilizada en la medicina moderna para la monitorización, prevención, detección y evaluación de diferentes enfermedades cardiovasculares. [1-3]. Cuando el PAI se adquiere de forma continua, la forma de onda contiene información valiosa sobre la fisiología del paciente y se refleja en valores como la presión sanguínea o la frecuencia del pulso. [4,5].

En el ámbito de la parada cardiorrespiratoria, se recomienda monitorizar la PAI en pacientes hemodinámicamente inestables o en unidades de cuidados durante la atención posterior a la parada [3,6,7]. Para aumentar las probabilidades de supervivencia, las directrices de la Asociación Americana del Corazón y la *Australian Resuscitation Council* recomiendan mantener niveles de presión sistólica (PS) superiores a 90 mmHg y 100mmHg, respectivamente. Además, se ha demostrado

que la señal de PAI puede ser útil a la hora de guiar la terapia de resucitación cardiopulmonar (RCP) [11-14].

Se han propuesto diferentes algoritmos para eliminar el ruido y caracterizar la señal de PAI [4,5,15], que es muy sensible al ruido de alta frecuencia o al movimiento del paciente [1,16]. El filtrado de componentes no deseadas permite calcular variables fisiológicas como la presión sistólica (PS), la presión diastólica (PD), la presión de pulso (PP) y la frecuencia cardíaca (FC) [4,5,15].

Aunque en la literatura se hayan propuesto varios algoritmos automáticos para caracterizar la forma de onda del PAI, ninguno ha sido evaluado durante la parada cardiorrespiratoria extrahospitalaria (PCREH). En dichos escenarios, por un lado, el ritmo cardíaco del paciente es altamente inestable y, por otro lado, el nivel de ruido elevado debido a las compresiones torácicas y ventilaciones de las maniobras de resucitación cardiopulmonar (RCP). Por lo tanto, el objetivo de este trabajo es diseñar un algoritmo automático de detección de picos de la señal de PAI y calcular variables fisiológicas correspondientes durante la PCREH.

2. Materiales

La base de datos utilizada en este estudio ha sido recolectada por los equipos de emergencia del Oslo Emergency Medical System durante la PCREH. Todos los episodios han sido grabados utilizando el desfibrilador LifePak 15, e incluyen la señal de PAI obtenida en la arteria femoral y la señal de ECG, ambas muestreadas a 250 Hz.

En total se extrajeron 119 segmentos de PAI con ritmo cardíaco organizado de 32 pacientes durante las pausas de compresiones torácicas. En el panel superior de la Figura 1 se puede observar un segmento de 3s de señal PAI.

En total la base de datos presenta una duración de 368 minutos, con una duración media de 3 ± 5.7 minutos por segmento. Los instantes de picos sistólicos y diastólicos de cada latido han sido anotados manualmente para ser usadas como valor de referencia.

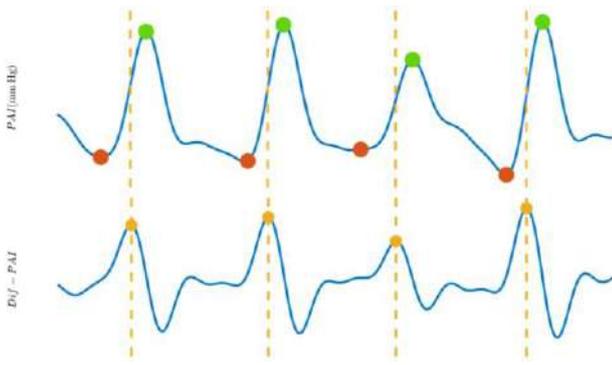


Figura 1: En el panel superior se muestra un segmento de 3s de señal PAI, y en el panel inferior su primera diferencia. Los picos de la señal diferencia y el instante de máxima pendiente ascendente se indican con un punto amarillo y líneas discontinuas amarillas, respectivamente. Los picos de presión sistólica y diastólica se indican en rojo y verde, respectivamente.

3. Métodos

En la Figura 2 se observa el método propuesto en este trabajo para la detección de picos en la señal de PAI y para el cálculo de los valores fisiológicos. El método se divide en tres fases. En la primera, la señal se preprocesa para eliminar las componentes no deseadas. En la segunda fase se estiman los instantes de los picos sistólicos y diastólicos utilizando la señal derivada discreta y un umbral adaptativo. Finalmente, en la última etapa, se calculan los diferentes valores fisiológicos.

3.1. Preprocesado

La señal de PAI es preprocesada para eliminar la componente continua y el ruido de alta frecuencia. La señal se descompone en ocho subbandas utilizando la transformada wavelet estacionaria (SWT) con wavelet madre Daubechies-4 y umbrales suaves. Después, la señal es reconstruida utilizando solo los coeficientes de detalle 6 y 7, correspondientes a la banda entre 1 y 4 Hz.

3.2. Detección de picos

Los potenciales instantes de cada latido se localizan mediante la detección de picos en la primera diferencia de PAI, dPAI. Se discriminan los latidos verdaderos considerando aquellos picos en dPAI que superan un umbral de amplitud adaptativo, descrito por la siguiente formula:

$$Um_i = \frac{\text{mediana}(P_{i-1}:P_{i-5})}{2}$$

donde Um_i es el umbral mínimo para el pico P_i , que se calcula teniendo en cuenta los valores medianos de los cinco picos previos. Además, se establece una distancia mínima de 300ms entre picos consecutivos.

Dichos picos se corresponden con el punto de pendiente máxima de la presión durante la fase sistólica del ciclo cardíaco, que comienza con en el pico diastólico y culmina en el pico sistólico. Por lo tanto, partiendo del pico detectado, se buscan los primeros puntos de inflexión hacia la derecha y hacia la izquierda en la señal de PAI para medir la presión diastólica y sistólica, respectivamente. En la Figura 1 se pueden observar los picos detectados en la señal derivada discreta y los picos sistólicos y diastólicos detectados en la señal de PAI.

3.3. Cálculo de las variables fisiológicas

Los valores de PS y PD se calculan como los valores de presión de la PAI en los instantes de picos sistólicos y diastólicos identificados en la fase anterior. La PP se calcula como la diferencia entre PS y PD. La FC se define como la inversa del tiempo transcurrido entre dos picos diastólicos consecutivos.

3.4. Evaluación estadística

El algoritmo de este estudio ha sido comparado con el algoritmo *wabp* de Physionet propuesto por Zong et al. y que es utilizada en la literatura [4].

Los instantes de pico sistólico y diastólico manualmente anotados se han utilizado como referencia para evaluar los algoritmos de detección de picos. Un pico detectado se considera correcto si está a menos de 300ms de la referencia. Ambos algoritmos se evalúan en términos de sensibilidad (Se): porcentaje de picos detectados correctamente; valor predictivo positivo (PPV): porcentaje de picos detectados que realmente son picos; y valor-F (F1): la media armónica de Se y PPV. Las métricas de desempeño se calculan por paciente y los resultados se dan en términos de mediana (rango intercuartílico, IQR) para todos los pacientes.

Finalmente, la capacidad del algoritmo para medir las variables fisiológicas se evalúa utilizando el error absoluto y el error relativo.

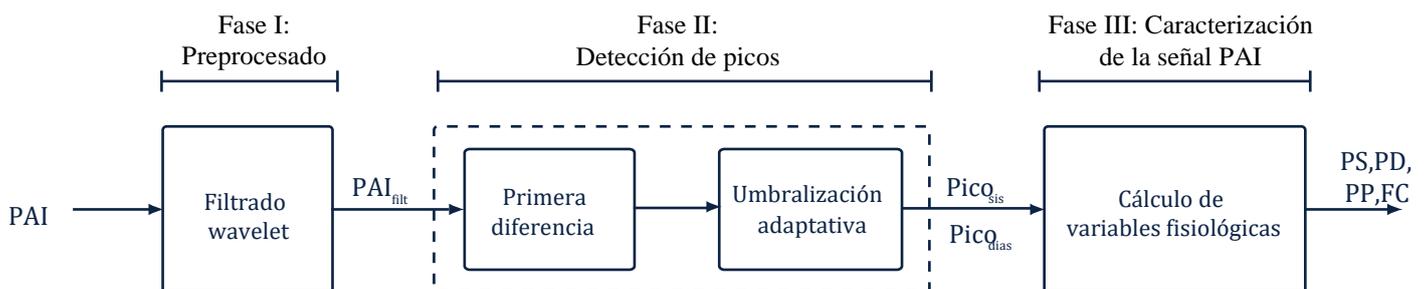


Figura 2: Esquema general del método propuesto para la detección de picos y extracción de variables fisiológicas de la señal de PAI.

4. Resultados

La Tabla 1 muestra las métricas de desempeño del algoritmo propuesto en este trabajo y el algoritmo *wabp* ampliamente usado en la literatura para la detección de picos en la señal PAI. Se observa que la propuesta supera al algoritmo de la literatura en 4.5 puntos de Se, 0.2 puntos de PPV y 1.5 puntos de F1.

	Se (%)	PPV (%)	F1 (%)
Este estudio	97.9 (4.2)	99.7 (11.6)	97.8 (3.1)
<i>wabp</i> [4]	93.4 (40.4)	99.5 (2.1)	96.3 (12.5)

Tabla 1. Métricas de desempeño para los algoritmos evaluados en este estudio, representados como mediana (IQR)

En la Tabla 2 se muestran los errores absolutos y relativos para las variables fisiológicas calculadas por el algoritmo propuesto. En términos de error absoluto la presión sistólica y diastólica están cerca o por debajo de 1mmHg mientras que en términos relativos están por debajo del 2 %. En el caso de la frecuencia cardiaca, el error es de menos de 0.4 latidos por minuto y 0.6 %.

	Error absoluto	Error relativo
Presión sistólica (PS)	1.12 (3.13) mmHg	1.42 (2.30) %
Presión diastólica (PD)	0.97 (0.76) mmHg	1.80 (2.28) %
Presión de pulso (PP)	2.11 (4.50) mmHg	5.09 (6.68) %
Frecuencia cardiaca (FC)	0.36 min ⁻¹	0.58 (0.87) %

Tabla 2. Error absoluto y relativo en términos de mediana (IQR) para las variables fisiológicas calculadas de forma automática.

5. Discusión

La señal PAI continua es ampliamente utilizada para monitorizar el sistema cardiovascular. La mayoría de los métodos de caracterización de dicha señal han sido diseñados para pacientes hemodinámicamente estables.

Esta es la primera vez que se ha caracterizado la señal PAI adquirida en la arteria femoral durante la PCREH. Esta propuesta viene a completar los algoritmos diseñados para la señal PAI adquirida en la arterial radial propuestos por Urteaga et al. [17].

Se consigue así cubrir la parametrización de la señal PAI durante la RCP independientemente de la línea arterial utilizada. La comparativa con el algoritmo *wabp*, muestra que éste presenta un desempeño peor que las propuestas de Urteaga et al. para el caso de la arteria radial, 4.3 puntos de F1 por encima, y para la arteria femoral 1.5 puntos de F1 por encima. La mejora de 4.5 puntos de Se del algoritmo

propuesto respecto al *wabp*, se debe principalmente a que se ha mejorado el filtrado y usado un umbral adaptativo que hacen que el algoritmo se adecue mejor a pacientes hemodinámicamente inestables.

El análisis comparativo de los algoritmos con las bases registradas en la arteria radial [17] y en la arteria femoral, en este estudio, muestra que la propuesta mantiene muy buena precisión en ambos casos (97.2% y 97.8% de F1 respectivamente). Sin embargo, el comportamiento de *wabp* es mejor con la arteria femoral (96.3% de F1) que con radial (92.3% de F1) una media de 94.3%, probablemente debido a una mayor variabilidad de los valores de presión de pulso. Se ha detectado una varianza de 2.1 mmHg en la señal femoral y 3.3 mmHg arterial.

Este trabajo está sujeto a una serie de limitaciones. Por un lado, se trata de una base de datos reducida de tan solo de 32 pacientes. Por otro lado, solo se consideraron casos con ritmos organizados. En futuros estudios se deberían de considerar todo tipo de ritmos presentes durante la parada cardiorrespiratoria.

El desarrollo de algoritmos automáticos de caracterización de la PAI permitiría incluir la información de la presión sanguínea en el guiado de la terapia RCP y en la predicción de evolución del paciente en PCREH.

Referencias

- [1] Marino, Paul L. *The ICU book*. Lippincott Williams & Wilkins, 2007.
- [2] Williams, Jonathan S., Stacey M. Brown, and Paul R. Conlin. "Blood-pressure measurement." *N Engl J Med* 360.5 (2009): e6.
- [3] Muntner, Paul, et al. "Measurement of blood pressure in humans: a scientific statement from the American Heart Association." *Hypertension* 73.5 (2019): e35-e66.
- [4] Zong, W., et al. "An open-source algorithm to detect onset of arterial blood pressure pulses." *Computers in Cardiology, 2003*. IEEE, 2003.
- [5] Li, Bing Nan, Ming Chui Dong, and Mang I. Vai. "On an automatic delineator for arterial blood pressure waveforms." *Biomedical Signal Processing and Control* 5.1 (2010): 76-81.
- [6] Avolio, Alberto P., Mark Butlin, and Andrew Walsh. "Arterial blood pressure measurement and pulse wave analysis—their role in enhancing cardiovascular assessment." *Physiological measurement* 31.1 (2009): R1.
- [7] Li-wei, H. Lehman, et al. "Methods of blood pressure measurement in the ICU." *Critical care medicine* 41.1 (2013): 34.
- [8] Bray, Janet E., et al. "The association between systolic blood pressure on arrival at hospital and outcome in adults surviving from out-of-hospital cardiac arrests of presumed cardiac aetiology." *Resuscitation* 85.4 (2014): 509-515.
- [9] Bray, Janet E., et al. "The association between systolic blood pressure on arrival at hospital and outcome in adults surviving from out-of-hospital cardiac arrests of presumed cardiac aetiology." *Resuscitation* 85.4 (2014): 509-515.

- [10] Australian Resuscitation Council, New Zealand Resuscitation Council. "Post-resuscitation Therapy in Adult Advanced Life Support. ARC and NZRC Guideline 2010." *Emergency Medicine Australasia* 23.3 (2011): 292-296.
- [11] Martin, Gerard B., et al. "Aortic and right atrial pressures during standard and simultaneous compression and ventilation CPR in human beings." *Annals of emergency medicine* 15.2 (1986): 125-130.
- [12] Niemann, James T., et al. "Coronary perfusion pressure during experimental cardiopulmonary resuscitation." *Annals of emergency medicine* 11.3 (1982): 127-131.
- [13] Sutton, Robert M., et al. "American Heart Association cardiopulmonary resuscitation quality targets are associated with improved arterial blood pressure during pediatric cardiac arrest." *Resuscitation* 84.2 (2013): 168-172.
- [14] Sutton, Robert M., et al. "Patient-centric blood pressure-targeted cardiopulmonary resuscitation improves survival from cardiac arrest." *American journal of respiratory and critical care medicine* 190.11 (2014): 1255-1262.
- [15] Navakatikyan, Michael A., et al. "A real-time algorithm for the quantification of blood pressure waveforms." *IEEE transactions on biomedical engineering* 49.7 (2002): 662-670.
- [16] Li, Qiao, Roger G. Mark, and Gari D. Clifford. "Artificial arterial blood pressure artifact models and an evaluation of a robust blood pressure and heart rate estimator." *Biomedical engineering online* 8.1 (2009): 1-15.
- [17] Urteaga, Jon, et al. "Automated detection of pulse using continuous invasive arterial blood pressure in patients during cardiopulmonary resuscitation." *2021 Computing in Cardiology Conference (CinC)*. IEEE, 2021.

Interfaz para el estudio del “jitter” neuromuscular

César. Valle Morgado¹, Armando. Malanda Triguero¹, Oscar. Garnés Camarena², Daniel W. Stashuk³

¹ Departamento Ingeniería Eléctrica, Electrónica y de Comunicación, Universidad Pública de Navarra, Pamplona, España, valle.108450@e.unavarra.es, malanda@unavarra.es

² Servicio de Neurofisiología Clínica, Hospital de la Fundación Jiménez Díaz, Madrid, España, oscar.garnes@quironsalud.es

³ Department of Systems Design Engineering, University of Waterloo, Ontario, Canada, stashuk@uwaterloo.ca

Resumen

Los potenciales de acción de unidad motora (PAUMs) proceden de la suma de los potenciales de las fibras que forman la unidad motora (UM). Las diferencias temporales entre los disparos de las distintas fibras pueden variar a lo largo de las sucesivas apariciones de un tren de PAUMs. A la dispersión de esas diferencias temporales de los disparos de las fibras de una UM se le llama “jitter” neuromuscular, el cual puede verse aumentado cuando existen patologías que afectan a la unión neuromuscular o a la motoneurona. Su estudio se realiza aplicando algoritmos automáticos o de forma manual por los especialistas neurofisiólogos utilizando las herramientas disponibles en los equipos comerciales de electromiografía (EMG). En el presente trabajo se ha desarrollado una interfaz gráfica basada en el entorno Matlab, que permite realizar mediciones manuales de “jitter”, emulando el método manual. Se ha utilizado un banco de registros de EMG de pacientes con patologías de la unión neuromuscular o neuropatías y se ha podido comprobar que los resultados de medición manual de “jitter” no presentan diferencias significativas (p valor de 0,79) con los obtenidos usando el equipo comercial Keypoint (Natus Medical). El interfaz desarrollado permitirá en el futuro hacer estudios comparativos de medición de “jitter” y comprobar las prestaciones de nuevos métodos automáticos.

1. Introducción

El potencial de acción de una motoneurona se propaga por el axón y por sus ramificaciones terminales. Cuando alcanza la llamada unión neuromuscular, se libera un transmisor químico, la acetilcolina, que al llegar a la fibra muscular origina la despolarización local de la membrana de esta fibra. Esta despolarización se propaga a lo largo de la fibra, hacia los tendones, accionando complejos mecanismos fisicoquímicos, que producen su contracción. De manera sincrónica se produce la activación del resto de fibras musculares que conforman la UM, esto es, aquéllas que también están unidas a la misma motoneurona. Un electrodo posicionado en las inmediaciones de esta UM recogerá la suma de los potenciales disparados en todas sus fibras musculares. Al potencial registrado se le llama potencial de acción de unidad motora (PAUM). En las contracciones musculares los potenciales eléctricos en las motoneuronas se producen de manera repetitiva y los subsiguientes PAUMs también aparecen repetidas veces: son los llamados trenes de PAUMs. En situaciones fisiológicamente normales, las diferencias temporales entre los potenciales del PAUM, procedentes de las distintas fibras de la UM, pueden variar ligeramente a lo largo de las sucesivas descargas del tren. A esta dispersión (Fig. 1) se le da el nombre de “jitter”

neurofisiológico o simplemente “jitter”. Este parámetro está aumentado en situaciones patológicas, generalmente asociadas a enfermedades de la unión neuromuscular, como la miastenia gravis, o a procesos de reinervación en las neuropatías o en la degeneración muscular debida a la vejez [1-3].

Tradicionalmente, la medición del “jitter” se realizaba con electrodos de fibra simple, con una superficie pequeña de la parte activa (del orden de 0.0005 mm²). Mientras el paciente ejercía una pequeña fuerza voluntaria, el electromiografista acercaba la aguja a una UM activa tratando de encontrar trenes de PAUMs con dos potenciales claramente separados (Fig. 1). De forma automática se podía computar estadísticamente la variabilidad en la diferencia temporal entre los dos potenciales de cada PAUM. Este tipo de agujas es caro y se utiliza repetidas veces, después de ser debidamente esterilizadas. Para evitar posibles contagios, en los últimos años ha habido una tendencia cada vez más clara de sustituir estas agujas por las agujas concéntricas, que son más baratas y de un solo uso aunque su área activa mayor. Los estudios de “jitter” en registros de aguja concéntrica son más complicados, porque es más difícil encontrar potenciales claramente separados y hay que marcar manualmente qué potenciales del PAUM deben ser comparados y dónde. La técnica de obtención manual del “jitter” se ha estandarizado [4], aunque los métodos automáticos de extracción de “jitter”, son aún muy escasos [5-6].

En este trabajo se presenta un interfaz gráfico basado en Matlab para realizar mediciones manuales de “jitter” reuniendo y extendiendo los métodos que presentan varios equipos comerciales de EMG. El objetivo es poder realizar estudios comparativos utilizando distintos procedimientos de extracción manual de “jitter”, así como servir de “gold standard” para la comparativa de nuevos métodos automáticos que se puedan proponer.

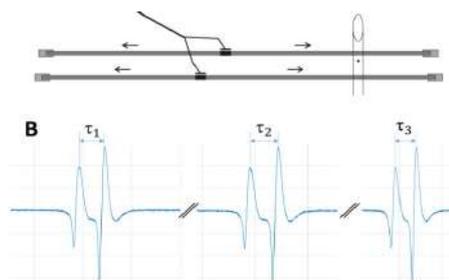


Figura. 1. Registro (A) y dispersión temporal (B) de un PAUM de dos fibras

2. Material y métodos

2.1 Registros EMG

Para la construcción del simulador se usó un banco de señales de EMG tomadas de 7 pacientes con patologías de la unión neuromuscular o neuropatías (periféricas, plexopatía, radiculopatía y ELA), recogidos en el Hospital Universitario de la Fundación Jiménez Díaz de Madrid. Las señales fueron adquiridas siguiendo el protocolo clínico establecido en el Servicio de Neurofisiología Clínica del Hospital y como parte de la exploración neurofisiológica en curso. Para el registro se utilizaron agujas concéntricas faciales, (área de registro, de 0.019 mm²) y un equipo Keypoint v.3.22. Las señales en bruto fueron filtradas con un filtro paso banda con frecuencias de corte de 500 Hz y 20 kHz, muestreadas a 48 kHz, exportadas en ficheros anonimizados y finalmente procesadas con el software DQEMG [7] para su descomposición en trenes de PAUMs, extrayéndose un total de 71 trenes.

Por otra parte, las señales en bruto fueron evaluadas por el propio equipo Keypoint, que extrajo diversos trenes de potenciales y mediciones de "jitter".

Las extracciones y todo este estudio fue aprobado por el Comité Ético del Hospital Universitario de la Fundación Jiménez Díaz.

2.2 Alineamiento y análisis del "jitter"

En la Figura 2.A aparecen superpuestos varios potenciales de un tren de PAUMs compuestos por los potenciales de dos fibras. En el ejemplo, para calcular el "jitter" se alinean todos los potenciales por el primer pico positivo y luego se mide la dispersión temporal del segundo pico positivo. En general, dado un punto de alineamiento, se puede medir el "jitter" en cualquier instante temporal de la señal, aunque no se corresponda con un pico (Fig. 2.B).

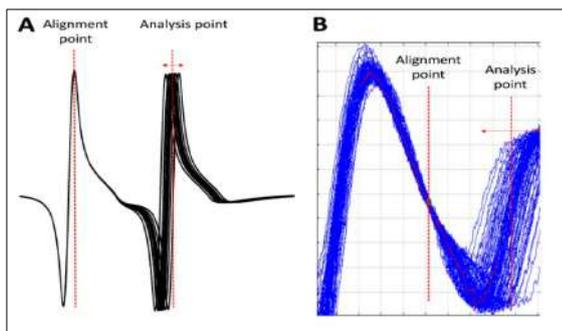


Figura 2. Medición del "jitter"

En síntesis, la medición del "jitter" se realiza en dos pasos sucesivos: primero se alinean los potenciales del tren en un instante temporal concreto y, una vez alineados, se mide la dispersión temporal local de los potenciales en otro instante temporal.

Siguiendo esta metodología, los distintos aparatos comerciales de adquisición de señales EMG plantean la medición manual de jitter utilizando uno de los siguientes métodos:

- Método de máximos: tanto el punto de alineamiento como el punto de medición del "jitter" son picos (máximos) de los PAUMs (Fig. 2.A)

- Método de umbral: tanto el punto de alineamiento como el punto de medición del "jitter" se toman a partir del corte de los potenciales del tren con un valor umbral dado (los umbrales para el alineamiento y para la medición del "jitter" pueden no ser los mismos) (Fig 2.B).

En la interfaz diseñada se incluyen ambos métodos, pero también se permite que el alineamiento se haga por máximos y la medición del "jitter" se haga por umbral y viceversa. Así se tendrán cuatro métodos según la secuencia de alineamiento y medición de "jitter" empleada.

En la Fig. 3.A se muestra un tren con 133 PAUMs, de los cuales se han descartado 16 por considerarse potenciales corruptos (con presencia de potenciales de otra UM). En la Fig. 3.B se tiene este conjunto de potenciales alineados por el método de umbral mientras que en la Fig. 3.C se tiene el mismo conjunto alineado por el método de máximos.

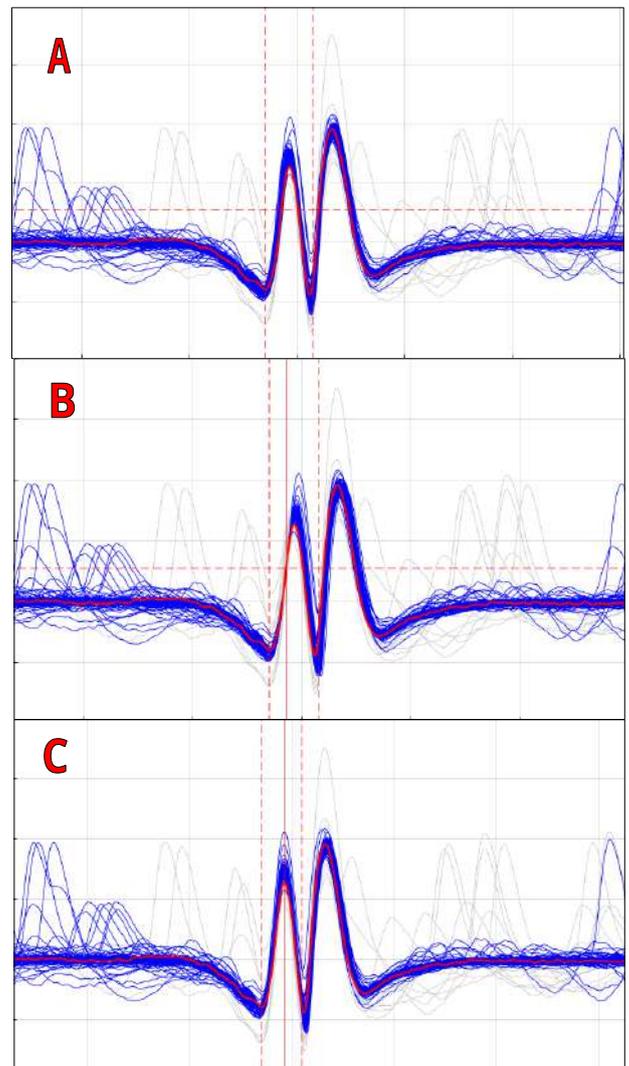


Figura 3. (A) Tren de potenciales sin alinear. (B) Tren de potenciales alineados por el método del umbral (punto de alineamiento marcado con línea horizontal discontinua). (C) Tren de potenciales alineados por el método de máximos (punto de alineamiento marcado con línea vertical continua).

2.3 Medición del “jitter”

Queda por definir cómo se lleva a cabo la medición del "jitter", una vez que el tren de PAUMs ha sido alineado. Tanto el método de máximos, como el de umbral, proporcionan un instante temporal (t_i) concreto para cada potencial del tren. El "jitter" es entonces la medida de la dispersión de estos instantes en el conjunto de potenciales del tren. En la herramienta se han implementado dos mediciones alternativas para el "jitter":

- La media de las diferencias consecutivas de los instantes t_i (“mean consecutive differences” o MCD) [1]:

$$MCD = \frac{|t_1 - t_2| + |t_2 - t_3| + \dots + |t_{i-1} - t_i|}{i - 1}$$

- La desviación estándar de los instantes t_i .

2.4 Otras funcionalidades

Además de los métodos descritos, en la aplicación (Fig.4) se han implementado varias funcionalidades que permiten realizar estudios de “jitter” de forma precisa y eficiente.

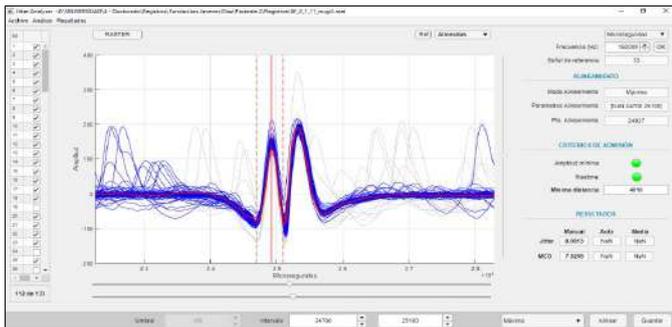


Figura 4. Vista principal de la interfaz desarrollada. JitterAnalyzer.

El funcionamiento general de la aplicación se describe en el siguiente diagrama de bloques:

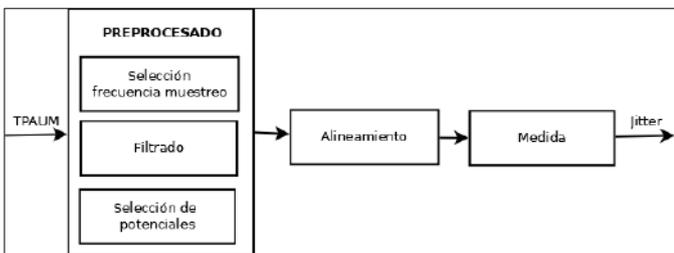


Figura 5. Diagrama de bloques de la aplicación.

Las características que se implementan a tal efecto son:

- Descartar potenciales del tren de PAUMs que puedan estar excesivamente corrompidos por ruido o interferencias con otros potenciales. Se permite también la recuperación de estos potenciales, por si fuera necesario.
- Modificar la frecuencia de muestreo de las señales, con el fin de no estar limitados por la frecuencia de las señales utilizadas y poder afinar los resultados del análisis.

- Realizar filtros paso banda con frecuencias de corte configurables.
- Realizar ampliaciones (zoom) y desplazamientos (“scroll”) de las figuras que contienen los potenciales.
- Permitir visualización en “raster” de los trenes de potenciales (Fig. 6).

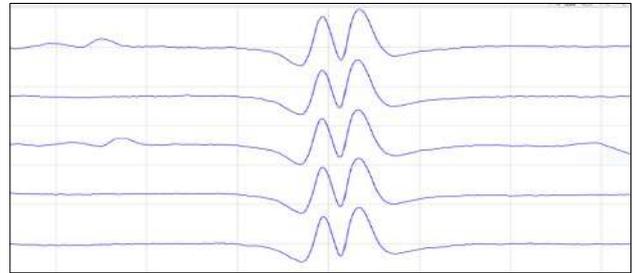


Figura 6. Visualización de potenciales en modo raster. Se muestran de 5 en 5.

2.5 Tecnologías empleadas

Para el desarrollo de la aplicación ha utilizado el software Matlab en su versión R2021b de la compañía MathWorks Inc. Concretamente para la implementación de la interfaz gráfica se ha utilizado el componente AppDesigner, utilizando únicamente el lenguaje propio de Matlab y siguiendo el paradigma de implementación modelo-vista-controlador (MVC).

3. Pruebas y resultados

Para comprobar el funcionamiento de la interfaz se han hecho pruebas comparativas de medición del "jitter" utilizando como referencia las mediciones obtenidas en el propio equipo de registro Keypoint. Este electromiógrafo presenta una facilidad para medir el "jitter" de manera semiautomática que primero lleva a cabo una descomposición automática de la señal en trenes de PAUMs. De cada registro de EMG descompuesto, se elige un solo tren de PAUMs, del cual se descartan aquellos potenciales que se desvían apreciablemente de la tendencia general. Asimismo, el usuario puede eliminar los potenciales que considere no aptos (edición). A continuación, determina sobre qué picos se llevará a cabo la medición del “jitter”, utilizándose el parámetro MCD. El software del electromiógrafo no permite identificar cuál es el tren de PAUMs recogido del conjunto de trenes obtenidos de la descomposición, ni siquiera de cuántos potenciales consta ni cuáles son. Sólo se han podido capturar las pantallas donde se presentan los trenes de PAUMs y los picos seleccionados para el alineamiento y la medición (Fig. 7.A).

La interfaz desarrollada en este trabajo toma como entradas los trenes de PAUMs previamente descompuestos por la herramienta externa DQEMG. (Según se ha expuesto, Keypoint no permite exportar el

resultado de su descomposición, por lo que debe realizarse con una aplicación externa). Sobre cada tren de PAUMs se realiza un descarte manual de potenciales no válidos para el análisis y posteriormente se seleccionan los picos para el alineamiento y para el análisis y se registra el “jitter” medido con los cuatro parámetros antes descritos.

Aunque la señal cruda inicial es la misma, los trenes de PAUMs utilizados para la medición del “jitter” pueden diferir en el electromiógrafo y en la herramienta desarrollada. La inspección de las gráficas de las pantallas capturadas con el equipo Keypoint y de los registros usados en la interfaz desarrollada es la única manera de comprobar que se trata de los mismos trenes (aunque no compuestos exactamente por los mismos potenciales).

Se han descartado algunos de los trenes analizados en las siguientes situaciones: picos del PAUM demasiado próximos o no bien definidos; presencia de bloqueos (algunas fibras de la UM pueden interrumpir esporádicamente sus disparos y no contribuir al PAUM). De los 91 trenes de PAUMs iniciales quedan finalmente 39 que pueden considerarse válidos para la comparación (ver ejemplo de la Fig. 7.A y 7.B), sobre los cuales se ha efectuado la medición del “jitter” con el parámetro MCD. En la Tabla 1 se expresan los valores estadísticos de las mediciones de cada método.

Equipo medida	Media (μ s)	Desv. Estándar (μ s)
Keypoint	32,92	23,65
JitterAnalyzer	36,91	35,05

Tabla 1. Resultados estadísticos de los métodos de medida.

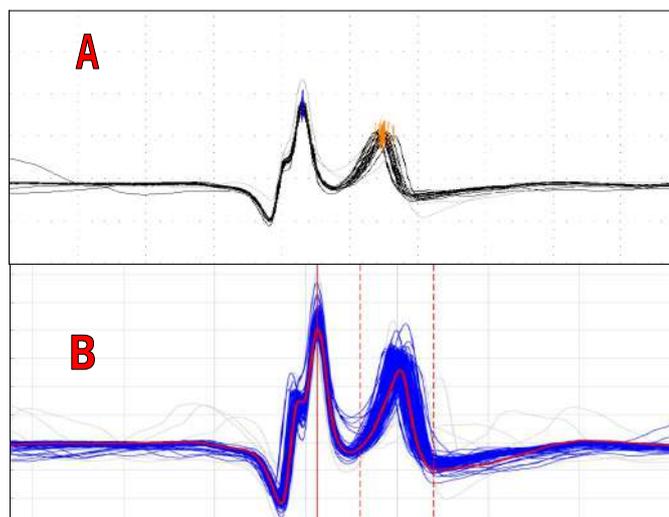


Figura 7. (A). Ejemplo de análisis realizada con el equipo Keypoint. (B). Ejemplo de análisis realizado con la interfaz.

Con los datos obtenidos se ha realizado un análisis de la varianza para determinar si existen diferencias estadísticas significativas entre ambos métodos. En concreto, se ha realizado un análisis de ANOVA pareada con un nivel de significancia del 5% ($\alpha = 0,05$), y se ha obtenido un p-valor de

0,79, que indica que no existen diferencias significativas en la medición del “jitter” al utilizar los dos métodos.

4. Conclusiones y líneas futuras

Se ha implementado una interfaz que permite replicar con exactitud las mediciones que realizan los especialistas con los equipos de registros de EMG comerciales para la medición de “jitter” a partir de trenes de PAUMs. La interfaz incluye métodos para el alineamiento del tren, descarte de potenciales y obtención del “jitter”.

Para validar la interfaz, se han llevado a cabo pruebas comparativas utilizando un banco de registros de EMG tomados de sujetos con patología neuropática o de la unión neuromuscular, no apreciándose diferencias estadísticamente significativas entre los resultados medidos con el electromiógrafo comercial Keypoint y la interfaz desarrollada.

Como planteamientos futuros se propone:

- La comparación de los métodos manuales para estudiar si existen diferencias estadísticamente significativas de los valores de “jitter” obtenidos en función del método utilizado tanto para el alineamiento como para la medida.
- La comparación de resultados obtenidos con la interfaz y los obtenidos con nuevos algoritmos automáticos, con el fin de verificar el correcto funcionamiento de estos.

Referencias

- [1] Trontelj JV, Stalberg E. Single-fiber and macro electromyography. In T. E. Bertorini (ed), Clinical evaluation and diagnostic tests for neuromuscular disorders. Butterworth Heinemann, Amsterdam, 2002.
- [2] De Carvalho M, Swash M. Lower motor unit dysfunction in ALS. *Clinical Neurophysiology* 127 (2016) 2670–2681.
- [3] Hourigan ML, McKinnon NB, Johnson M, Rice CL, Stashuk DW, Doherty TJ. Increased motor unit potential shape variability across consecutive motor unit discharges in the tibialis anterior and vastus medialis muscles of healthy older subjects. *Clin Neurophysiol.* 2015 Dec; 126(12):2381-9.
- [4] Stalberg E. Jitter analysis with concentric needle electrodes. *Ann. N.Y. Acad. Sci.* 1274 (2012) 77–85.
- [5] Abdelmaseeh M, Stashuk DW. Motor unit potential jitter: a new measure of neuromuscular transmission instability. *IEEE Trans Neural Syst Rehabil Eng.* 2017 Jul;25(7):1018-1025.
- [6] Malanda A, Stashuk D, Navallas J, Rodríguez-Falces J. Medición automática del “jitter” en registros electromiográficos. Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB). Santander, noviembre de 2019.
- [7] Doherty TJ, Stashuk DW. Decomposition-based quantitative electromyography: methods and initial normative data in five muscles. *Muscle Nerve.* 2003 Aug;28(2):204-11.

An immersive virtual reality training system for paediatric stroke code medical training.

P. Carmona Sánchez¹, M. Rodríguez Matesanz¹, C. Verdú-Sánchez², J. Rubio-Bolivar³, Enrique J. Gómez^{1,4}, M. Quintana^{2,3}, P. Sánchez-González^{1,4}

¹ Biomedical Engineering and Telemedicine Centre, ETSI Telecomunicación, Center for Biomedical Technology, Universidad Politécnica de Madrid, Madrid, Spain

² Hospital Universitario La Paz, Madrid, Spain

³ IdiPaz Instituto de Investigación Sanitaria del Hospital Universitario La Paz, Madrid, Spain

⁴ Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina, Madrid, Spain

Abstract

In Spain, around 120 cases of paediatric stroke are detected each year. At an early age, stroke patients can suffer significant sequelae that they must live with for the rest of their lives. The severity of these consequences is proportional to the time that elapses from the onset of symptoms to their treatment. Although early diagnosis is of vital importance, the lack of experience of primary care centres and their staff in dealing with these cases means that they are not always familiar with the symptoms that precede diagnosis. For this reason, it is important to be able to use tools that allow professionals to train for clinical events by preparing them for these situations in an active and realistic scenario. In this research, a training system based on Virtual Reality scenario has been developed in which outpatient primary care doctors learn and practice a case of paediatric stroke by carrying out the different steps of the NIHSS paediatric stroke scale that allows the diagnosis of the disease and the subsequent activation of the "Código Ictus Pediátrico", a protocol developed for the diagnosis, transport, and treatment of paediatric stroke patients.

1. Introduction

Stroke, also known as cerebrovascular accident (CVA), is a disease that affects the blood vessels of the central nervous system. It consists of a disturbance in circulation that results in poor blood flow to a region of the brain causing it to malfunction [1]. Although it is a disease associated with adults, in the latest hospital morbidity survey published in 2018 by the National Statistics Institute, a total of 242 patients aged up to 14 years diagnosed with cerebrovascular disease were reported in Spain [2], of whom 20 patients died that year. Placing stroke among the 10 leading causes of death in the paediatric population in developed countries [3]. In addition, the degree of long-term sequelae that occur in this population group is of particular value, as this is a population group with a high life expectancy [4]. About 80% of children who survive a stroke will have long-term neurological sequelae [4].

The symptoms of this disease are known in adults but in paediatrics they are often milder, which means that the symptoms can be confused with other more common pathologies. This ends up delaying the diagnosis and the actions of the health services. Therefore, it is of utmost

importance to educate about the symptoms that support the diagnosis of paediatric stroke, as early diagnosis will reduce the sequelae suffered by patients [5 - 6].

For this reason, in 2019 the "grupo ictus pediátrico" was set up, made up of several healthcare professionals from the public hospitals of the Community of Madrid who have proposed the Código Ictus Pediátrico CIP, a protocol that sets out the procedure to be followed by a doctor when diagnosing a stroke in children. The CIP contains the form developed by the National Institute of Health Stroke Scale for Paediatrics (PedNIHSS), a series of questions that evaluate the patient's condition and diagnose him/her as a patient with stroke [4].

Medical simulation has acquired an important position as a training tool in recent years, its educational character is based on the recreation of situations in which students must make decisions within a clinical event. The students will be evaluated by their actions and decisions in an active way to solve a situation. Moreover, the educational character is supported by the evaluation of the students' performance.

Simulation in paediatrics plays an even more important role, as learning to perform procedures on the youngest and most vulnerable children is sometimes a risky practice [7]. Not only because of the procedure itself, but also because of the situations faced by healthcare professionals in paediatrics, as they tend to find themselves in situations of greater stress and delicacy, having to deal not only with the patient, but also with their families [7]. One way to create a realistic and immersive simulation that allows extrapolating what has been learned in it to real life is virtual reality. It improves the acquisition of knowledge in the world of healthcare [8].

The aim of this work is to design and implement a virtual reality-based training system that can provide professionals, such as paediatricians, with a means to learn the paediatric stroke code protocol through the PedNIHSS form that supports the diagnosis.

2. Materials

2.1. Hardware

In order to develop the system proposed in this work, it is necessary to have certain hardware equipment that allows virtual reality scenes to be executed. A Head-Mounted Display device for the visualisation of the scene, and two controls that simulate hands with which the user can interact (pick up, select, rotate) with the different objects. More specifically, the use of the Oculus Quest 2, Figure 1, [9] is recommended. Due to its 6 degrees of freedom that allows us to move freely in a three-dimensional space. They are autonomous, no cable connection is necessary [10] and they weigh 503gr, which facilitates mobility around the stage.



Figure 1. Oculus Quest 2 [11].

2.2. Software

For the development of the simulation the latest versions of Unity 3D has been used [12]. This videogame engine stands out for its portability, as it can be installed on Windows, Linux or MacOS.

In addition, Unity offers the possibility of working with 3D models generated with the modelling programme Blender [13] software with which the modelling, rigging, animation, simulation and rendering of three-dimensional characters has been carried out. In this project, it has been used to create the patient character. To do this, the figure was first created in MakeHuman [14], a humanoid prototyping tool.

3. Methods

3.1. Virtual environment

The simulation is divided into two scenes (Figure 2), a first one in which the user becomes familiar with the technology and a second part in which the evaluation is carried out.

Throughout the main scene, we find ourselves in an out-of-hospital primary care ward with a patient with a clinical picture compatible with paediatric stroke. In order to train the user, the PedNIHSS form has been implemented through a series of interactive menus, so that the doctor can diagnose the patient.

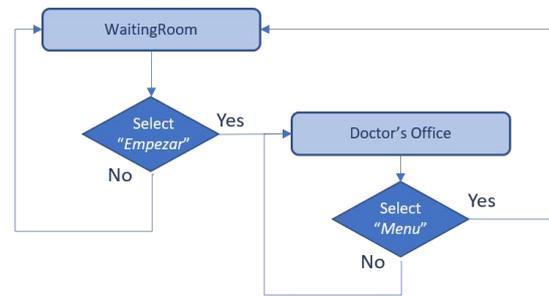


Figure 2. Scene interactions

In addition to the user, there are three other characters in the scene: the patient, her mother and a nurse. With them, a series of interactions can take place throughout the second scenario, the main scene, as shown in Figure 3.

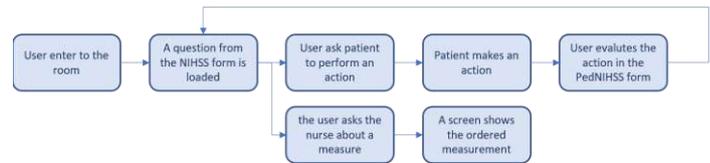


Figure 3. Main scene flow.

3.2. Simulation results

The PedNIHSS form has been implemented as the main form of the CIP. A series of animations have been developed for the user to decide whether the patient is suffering a stroke. So that the user evaluates the patient between a range of 0-3 points in the questions, the metric that is stored is whether the user has evaluated the patient correctly. It is also stored whether you want to activate the protocol, as it is no good if you have got all the questions right if in the end you do not activate the CIP.

In addition, within the protocol, it is important that the PedNIHSS evaluation is carried out in less than 15 minutes. For this reason, the time taken by the physician to perform the assessment is also recorded.

4. Results

A training system has been implemented consisting of a virtual reality scenario in which the user encounters a patient with a stroke-compatible condition. During the execution of the simulation, the doctor is allowed to interact with the patient while performing the PedNIHSS questionnaire, on which the CIP protocol works.

The first part of the simulation (Figure 4) allows the users to familiarise themselves within a VR environment, understanding how both the technology used and the scenario work. This has been done because, being a relatively new technology, which has only recently penetrated the medical world, the users may never have been confronted with a similar situation. Afterwards, they go to a room where they have to perform the different interactions, and it is at this moment when the time starts to count, and the student's activity is evaluated.



Figure 4. WaitingRoom scene

In the main room (Figure 5) the user interacts with the three characters:

- The mother (Figure 6A): she is in charge of explaining the reason for the visit.
- The nurse: the user can order a series of constants to characterise the patient, thanks to the nurse's menu (Figure 6B).
- The patient (Figure 6C): the user makes a series of requests, and the character carries them out, allowing the evaluation of the patient's condition. For this, a series of physical movements and other interactions have been animated, such as lip-synchronisation to simulate the patient's speech.



Figure 5. Doctor'sOffice scene

A panel is displayed to the user in which the different PedNIHSS questions are shown (Figure 6D) together with the answer options. The user selects one option per question and at the end must decide whether the patient is a "CIP" or not.



A

B



C

D

Figure 6. (A) Mother model. (B) Nurse menu (C) Patient model. (D) PedNIHSS form.

In order to make the training more quantifiable, the simulation has been designed in such a way that a document is generated at the end. This document records whether the development has been carried out correctly. The choice of whether a movement was made incorrectly based on the paediatricians' indications. This makes it possible to know whether the trainee has acquired the desired knowledge. The system also allows external visualisation of the user's experience, so that their decision-making can be evaluated by their trainers in real time.

Throughout the development of the simulation, the support and participation of the doctors has been ensured so that a scene adapted to their needs has been created. In this way, the result has been shared with the medical team that will use it and the training system has their approval. Due to the current medical crisis situation, the application has been informally validated by physicians who have been enthusiastic about working with the simulation. Therefore, we are currently working together with them on a comprehensive validation with paediatricians and neurologists to demonstrate the improvement of VR-based simulation systems as training tools.

5. Conclusions

The vulnerability of paediatric patients and the lack of information on cardiovascular accidents in childhood often delay the diagnosis of the disease. For this reason, the hospitals of the Community of Madrid have developed the "Código Ictus Pediátrico", an action protocol based on the

PedNIHSS form. To disseminate the CIP, a virtual reality simulation environment has been developed.

The implemented environment recreates an immersive situation that allows students to place themselves in a real case without endangering patients. This favours the acquisition of knowledge and its subsequent implementation based on previous experience. In addition, this new training allows users to self-assess themselves during the scenario. Performance metrics are stored in order to be able to see whether the trainee has acquired the desired knowledge.

However, it has not been possible to perform a comprehensive clinical validation. Despite this situation, the application has been informally validated by clinicians who have been enthusiastic to try it out. To this end, a validation with out-of-hospital paediatricians is proposed.

Thanks to the proposed simulation, an innovative training tool has been created that offers many training possibilities. Currently, the patient performs the different actions based on a fixed form, but more movements could be performed and there could be more levels of simulation, where the patient erroneously performs other parts of the simulation. In this way, every time a student wants to train the protocol, he or she will encounter different interactions. In addition, an artificial intelligence system could be implemented that would pose the next interaction based on the previous wrong answers selected by the user. Each action would then be adapted to the specific needs of the person immersed in the environment.

References

- [1] Sorribes Capdevila M *et al.* Abordaje de los ictus: colaboración entre Atención Primaria y Especializada. *Semergen*, vol 31, no 7, 2005, pp 314–318, (ISSN: 11383593).
- [2] Instituto Nacional de Estadística. Altas hospitalarias clasificadas por intervalos de estancia, según el diagnóstico principal y el grupo de edad. *Encuesta de morbilidad hospitalaria 2018*, 2018.
- [3] Graziano AP *et al.* Enfermedad cerebrovascular en la infancia: Presentación de casos clínicos. *Archivos Argentinos de Pediatría*, vol 114, no 1, 2016, pp e5–e8, (ISSN: 16683501).
- [4] Federico Ballenilla Marco *et al.* Ictus pediátrico: Grupo de trabajo Código Ictus Pediátrico. 2019.
- [5] Barrena Caballo M *et al.* Atención al ictus pediátrico en Aragón. 2020, (ISBN: ISBN 978-84-09-23721-0).
- [6] Stoodley N. RCPCH clinical guideline, stroke in childhood: an evidence-based guideline for diagnosis, management, and rehabilitation. *Clinical Radiology*, vol 72, 2017, p S25, (ISSN: 00099260).
- [7] Tarrío FJR, De La Oliva Senovilla P. Editorial: Simulation in paediatrics: The new revolution in paediatric training and a guarantee for quality care. *Anales de Pediatría*, vol 73, no 1, 2010, pp 1–4, (ISSN: 16959531).
- [8] Pottle J. Virtual reality and the transformation of medical education. *Future Healthcare Journal*, vol 6, no 3, 2019, pp 181–185, (ISSN: 2514-6645).
- [9] Get Started Developing for the Oculus Quest Platform | Oculus Developers. <https://developer.oculus.com/quest/> (Consultada: Jun 2021).
- [10] Oculus Quest 2: Our Most Advanced New All-in-one VR Headset | Oculus. <https://www.oculus.com/quest-2/> (Consultada: Jun 2021).
- [11] Seymour NE *et al.* Virtual reality training improves operating room performance results of a randomized, double-blinded study. in *Annals of Surgery*, 2002, vol 236, no 4, pp 458–464.
- [12] Plataforma de desarrollo en tiempo real de Unity | Motor de VR y AR en 3D y 2D. <https://unity.com/es> (Consultada: Jun 2021).
- [13] blender.org - Home of the Blender project - Free and Open 3D Creation Software. <https://www.blender.org/> (Consultada: Jun 2021).
- [14] MakeHuman. <http://www.makehumancommunity.org/> (Consultada: Jun 2021).

ÍNDICE DE AUTORES

A

Abella M	159, 167, 179, 193
Acosta O	127
Aja-Fernández S	51
Albajar Sigalés A	2
Alberdi Aramendi A	10
Alberola-López C	51, 175
Alcaraz Martínez R	83, 139, 147, 151, 155
Almagro-Pérez C	18
Álvarez González D	63
Álvarez I	42
Añorbe de Frutos A	30
Aramendi Ecenarro E	197
Arroyo Domingo CA	63
Atienza Vicente CM	143

B

Barea R	107
Barrenechea M	10
Barroso FO	6
Bautista-Salineró M	22
Beedle AEM	2
Berdón PM	167
Berruezo Sánchez A	171
Bertomeu-González V	83, 147, 151, 155
Berve PO	197
Blanco-Velasco M	91
Blas JG	167
Boquete L	107
Braxton A	18
Bueno H	123

C

Cabezas E	127
Calle D	103
Calvillo Arbizu J	119
Camara Rey O	171
Cano del Pozo M	22
Cano J	151
Cano M	55
Cano-Manuel JL	179
Canós MA	87
Cañadas M	189
Carmona Sánchez P	205
Carro J	115
Castelli J	127
Castillo-Rabazo J	187
Castro de la Nuez P	119
Cavaliere C	107
Cepeda Chafía S	51

Chen R	10
Chénier F	111
Chil R	75
Cuba Gato MA	183
Cubero L	127
Cussó L	103

D

De Andrés J	87
de Crevoisier R	127
de Luis García R	51
del Campo F	63
Del Cerro CF	159, 179, 193
Desco M	103, 159, 167, 179, 193
Durá JL	87
Durán-Rey D	187

E

Elola Artano A	197
Esparza-Iaizzo M	42

F

Fácil L	151
Ferreira D	95
Ferrer-Bonsoms JA	30

G

Galán A	167, 179
Galán C	34
Gallego-Ferrer G	26
García A	46
García Carmona R	14
García-Blas J	159, 179, 193
García-Martín E	107
García-Sáez G	123, 131
Garnés Camarena O	201
Giménez RC	159
Gómez C	55, 59, 71
Gómez EJ	67, 79, 205
Gómez M	115
Gómez-Cruz C	189
Gómez-Grande A	79
Gómez-Pilar J	55, 59
Gomis P	95
González-Portillo MR	187
Goya Esteban R	22
Gutiérrez R	107
Gutiérrez -de Pablo V	59, 71
Gutiérrez Martín A	6, 99
Gutiérrez Tobal GC	63
Guzmán-García C	67

H			
Hernando ME	123, 131	Moreno-Castán A	103
Hidalgo Torres LA	75	Muñoz-Barrutia A	18, 189
Hornero R	55, 59, 63, 71, 163	Muñoz-Romero S	91
Hornero Sos F	83, 147, 155	N	
Hruban R	18	Navarro González R	51
Huerta Herraiz A	139	Navarro Ruiz E	38
I		Nicolás-Sáenz L	18
Iniesta JM	131	Núñez P	22
Iriarte A	46	O	
Isasa Reinoso I	10	Oropesa I	67
J		Ortiz M	107
Jauregui Garrido B	171	P	
Jiménez García J	63	Palmieri F	95
K		Pardo Picazo MA	38
Kiemen A	18	Pascau J	127
Klinzing JG	42	Pascual-Valdunciel A	6
L		Pérez Valdivia MA	119
Laguna P	95	Pérez-Benito D	75
Laguna S	189	Pérez-Gandía C	131
Larraga García B	99	Pérez-Velasco S	163
López Dorado A	107	Piol A	193
López-Agudelo I	187	Pizarro-Chong A	111
López-Larraz E	42	Plaza de Miguel C	135
Lopo-Martínez V	6	Plaza-Seco C	91
Losada H	143	Poza J	22, 55, 59, 71
Lovas A	10	Pueyo E	95, 115
M		Q	
Malanda Triguero A	201	Quesada Dorador A	83, 147, 155
Manuel-Cano JL	167	Quintana Díaz M	99, 205
Mares-Bou S	26	R	
Margelí C	46	Ràfols P	34
Martí-Bonmatí L	87	Ramírez J	95
Martín González E	175	Reina Tosina J	119
Martínez JP	95	Revilla-Orodea A	175
Martínez Rodrigo A	139	Revilla-Vallejo M	55
Martínez Seijas V	123	Reyes Leiva KM	183
Martínez-Cagigal V	163	Rico-Gómez I	189
Maturana-Candelas A	71	Rieta Ibáñez JJ	83, 139, 147, 151, 155
Mayans B	46	Roa Romero LM	119
Medina López R	119	Roca-Cusachs Soulere P	2
Méndez de Vigo Iriarte I	14	Rodríguez-Fernández J	26
Minguez J	42	Rodríguez-Galván JR	175
Molas C	46	Rodríguez Rodríguez M	99
Möller K	10	Rodríguez Matesanz M	205
Monasterio V	115	Rojo-Álvarez JL	91
Montesano L	42	Romera M	34
Moreno Muñoz G	123	Rubio Bolívar J	99, 205
		Ruiz JE	95

S

Saiz J	87
Sales Bellés C	171
San Martín L	143
San Martín Retegi A	30
Sánchez Anzules R	135
Sánchez-González P	67, 79, 205
Sánchez-Margallo FM	135, 187
Sánchez-Margallo JA	135, 187
Santamaría-Vázquez E	163
Santos Paz JA	14
Sebastián Aguilar R	171
Seiffert AP	79
Serrancolí G	46
Serrano J	127
Serrano Olmedo JJ	183
Sevilla-Ruiz MT	175
Simon A	127
Solanes C	87
Sosenko K	159
Soto Iglesias D	171
Stashuk DW	201

T

Talaminos Barroso A	119
Tapia-Galisteo J	131
Tejo-Otero A	143
Tola-Arribas MA	22, 55
Torner J	46
Trujillo-Muñoz S	26

U

Urgelés D	131
Urrea JF	179
Urteaga Urizarbarrena J	197

V

Valle Morgado C	203
Vaquerizo Villar F	63
Vaquero JJ	75
Verdú C	205
Villarejo-Galende A	79
Vraka A	83, 147, 155

W

Wik L	197
Wirtz D	18
Wood L	18
Wu P	18

Z

Zangróniz R	151
-------------	-----