

# CUANTIFICACIÓN DEL VENTRÍCULO CARDÍACO IZQUIERDO MEDIANTE REDES NEURONALES CONVOLUCIONALES ESPACIO-TEMPORALES

**Debus Alejandro**

*Research institute for signals, systems and computational intelligence, sinc(i), FICH-UNL/CONICET  
Director/a: Ferrante Enzo*

**Área:** Ingeniería

**Subárea:** Informática y sistemas de información

**Palabras claves:** redes neuronales convolucionales, imágenes médicas, visión computacional

## INTRODUCCIÓN

Según datos de la OMS, en 2015 alrededor de 17,7 millones de personas murieron en todo el mundo debido a enfermedades cardíacas. La cuantificación de diversos índices que caracterizan el ventrículo izquierdo (VI) es un factor clave para la identificación y el diagnóstico de dichas patologías (*Karamitsos, 2009*). Sin embargo, su estimación es una tarea muy compleja debido a su compleja dinámica temporal y la variabilidad de las estructuras cardíacas inter-sujetos. Índices como el área del miocardio, el espesor de las paredes regionales y las dimensiones direccionales de la cavidad (ver Figura 1) proporcionan información útil para diagnosticar diversos tipos de patologías cardíacas. Las imágenes de resonancia magnética cardiovascular (RMC) son en general utilizadas para los estudios relacionados con el VI, principalmente debido al buen contraste y calidad de la imagen (*Suinesiaputra, 2015*).

El enfoque clásico para la cuantificación del VI consiste en estimar dichos índices mediante técnicas de segmentación automática (*Peng, 2016; Poudel, 2016*). En este caso, la precisión de los índices estimados está condicionada a la calidad de dicha segmentación. Por este motivo, en este trabajo adoptamos una estrategia alternativa que estima directamente los índices de interés de la secuencia de imágenes de entrada. Inspirados por el trabajo de (*Xue, 2018*), nuestro modelo se basa en una red neuronal convolucional (CNN) que opera directamente sobre las secuencias de imágenes utilizando convoluciones espacio-temporales y regresión los índices de interés. En el contexto del análisis de video, estas convoluciones son convoluciones 3D estándar que actúan en volúmenes de video espacio-temporal (*Tran, 2015*). Aquí, serán empleadas para procesar subconjuntos de frames de RMC temporalmente contiguos, aprovechando la información temporal para mejorar la precisión de las predicciones.

## OBJETIVOS

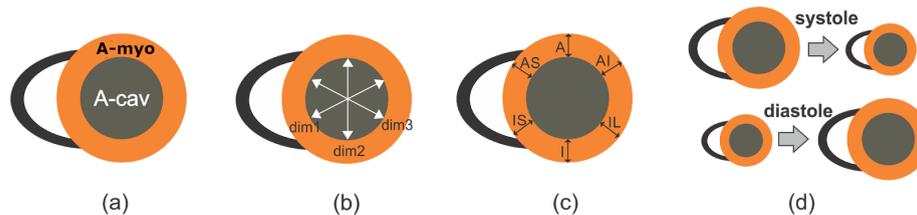
El objetivo general de este trabajo\* es el desarrollo de un algoritmo basado en aprendizaje profundo y redes neuronales convolucionales para la cuantificación del ventrículo izquierdo en imágenes RMC que incorpore la dimensión temporal al análisis. Para ello, se buscará validar experimentalmente la hipótesis de que la calidad de las predicciones puede ser mejorada teniendo en cuenta la dinámica temporal del corazón. Por medio de un completo estudio experimental, se brindará evidencia empírica de que la incorporación de la dinámica temporal del corazón a través de convoluciones espacio-temporales mejora la precisión de la predicción cuando se compara con modelos que operan sólo sobre imágenes individuales.

*Este resumen está basado en el paper "Left ventricle quantification through spatio-temporal CNNs" realizado por el autor y su director, recientemente aceptado para publicación en el "Statistical Atlases and Computational Modelling of the Heart Workshop 2018" a celebrarse en Septiembre en Granada, España. El paper está a disposición del comité de evaluación si así lo requiere.*

Título del proyecto: Image registration under the big data paradigm: Boosting risk understanding through imaging technologies  
Instrumento: AXA Research Fund Grant  
Año convocatoria: 2017  
Organismo financiador: AXA Research Fund  
Director: Enzo Ferrante

## METODOLOGÍA

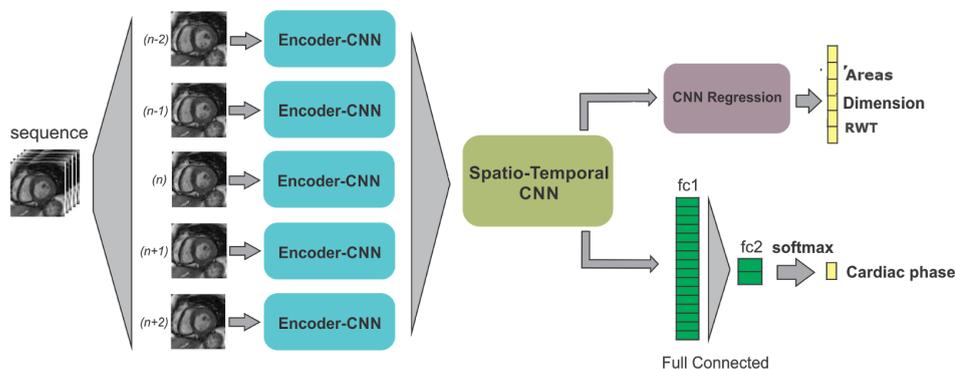
En este trabajo se diseñó un modelo basado en aprendizaje profundo para la predicción de los índices de interés a partir de imágenes RMC. A continuación, se describe la arquitectura del modelo, el corpus de datos utilizado y la metodología de entrenamiento adoptada.



**Figura 1.** Ilustración de índices del ventrículo cardíaco izquierdo. (a) Área de la cavidad (verde) y área del miocardio (naranja). (b) Dimensiones direccionales de la cavidad. (c) Espesores de la pared regional. A: anterior; AS: anteroseptal; IS: inferoseptal; I: inferior; IL: inferolateral; AL: anterolateral. (d) Fase del ciclo cardíaco (sístole o diástole)

### A. Arquitectura

En la Figura 2 se presenta una descripción general de la arquitectura propuesta. La red toma secuencias temporales de  $k$  fotogramas (frames) para estimar los índices correspondientes al frame central. De esta manera, la predicción toma en cuenta la dinámica temporal al observar los frames aledaños. A continuación, describimos de forma general los principales componentes de la arquitectura propuesta.



**Figura 2.** Descripción general de la arquitectura propuesta

**Encoder-CNN.** Usamos una primera CNN (Encoder-CNN en Figura 2) para extraer características informativas de fotogramas individuales. Dicha red está compuesta por capas convolucionales 2D estándar. Utilizamos la función de activación ReLU y la normalización por lotes para aliviar la fase de entrenamiento.

**Spatio-Temporal CNN.** Después de la fase de codificación, los filtros generados por las Encoder-CNN para cada frame individual se utilizan para construir un volumen espacio-temporal con múltiples canales por segmento temporal. Este volumen se procesa utilizando convoluciones 3D que operan en las dimensiones temporal y espacial, produciendo mapas de características compuestas que incorporan información de ambas dimensiones.

**Ramas paralelas finales.** Después de fusionar las características espacio-temporales, se derivan dos ramas paralelas: (i) la primera rama corresponde a una CNN superficial acoplada luego del módulo espacio-temporal, que actúa como un regresor de los índices de la Figura 1 (a, b y c); (ii) en la segunda rama, una tercera capa convolucional se acopla al módulo espacio-temporal, seguida de un perceptrón multicapa totalmente conectado (MLP) para la estimación de la fase cardíaca.

### **B. Corpus de datos y criterios de evaluación**

El método propuesto fue validado experimentalmente utilizando los datos de entrenamiento proporcionados por la competencia LVQuan 2018 (<https://lvquan18.github.io/>), compuesto por imágenes de secuencias de RMC de eje corto pertenecientes a 145 pacientes. Para cada paciente, la base de datos contiene 20 fotogramas correspondientes a un ciclo cardíaco completo. Los mismos fueron divididos en particiones de entrenamiento, validación y prueba siguiendo una metodología de validación por N-folds, con N=5.

Para evaluar el rendimiento de los algoritmos se utilizó el coeficiente de correlación de Pearson (PCC) y el error absoluto medio (MAE). Para evaluar el rendimiento de la clasificación de la fase cardíaca se utilizó la tasa de error (ER).

### **C. Proceso de entrenamiento**

La red propuesta fue entrenada por medio de la minimización de una función de pérdida definida sobre las muestras de entrenamiento, donde cada muestra corresponde a un conjunto de  $\kappa$  frames contiguos donde las estimaciones se realizan sólo para el slice central. Se utilizó el error cuadrático medio como función de pérdida para los índices numéricos (Figura 1 a, b y c) y entropía cruzada para la estimación de la fase cardíaca (Figura 1.d). Para la generación de muestras de entrenamiento y prueba se siguió una hipótesis circular, en el sentido de que el primer y último frame de cada secuencia fueron considerados temporalmente contiguos.

## **RESULTADOS Y CONCLUSIONES**

La efectividad del método propuesto fue validada experimentalmente midiendo la influencia del parámetro  $\kappa$  para valores 1, 3, 5 y 7. Notar que  $\kappa = 1$  implica que no se incorpora ninguna dimensión temporal al análisis. La Tabla 1 presenta estos resultados. Como puede observarse, la incorporación de la dimensión temporal por medio del uso de frames consecutivos ( $\kappa > 1$ ) resulta en un incremento significativo en la precisión del método (menor MAE y mayor PCC), validándose así nuestra hipótesis inicial sobre la importancia de la dimensión temporal en la estimación de índices VI. Tal como puede apreciarse en la tabla, los modelos entrenados considerando  $\kappa=5$  y  $\kappa=7$  no presentan grandes diferencias. Esto indica que el uso de 5 frames provee suficiente contexto temporal al modelo.

En conclusión, se propuso una nueva arquitectura CNN para la cuantificación del VI que incorpora la dinámica del corazón por medio de convoluciones espacio-temporales. A diferencia de otros métodos que se basan en mecanismos más complejos (como las redes neuronales recurrentes (Xue, 2017)) se emplearon convoluciones simples en 3D para fusionar información proveniente de fotogramas de RMC contiguos temporalmente. Los resultados muestran que la incorporación de información temporal al modelo aumenta significativamente el rendimiento de predicción para todos los índices. En el futuro, planeamos explorar el rendimiento del modelo propuesto con uno que emplee convoluciones 2D tradicionales, con el objetivo de minimizar la complejidad del modelo.

Validación cruzada con N-folds								
	MAE				PCC			
	$\kappa = 1$	$\kappa = 3$	$\kappa = 5$	$\kappa = 7$	$\kappa = 1$	$\kappa = 3$	$\kappa = 5$	$\kappa = 7$
<i>Áreas (mm<sup>2</sup>)</i>								
a-cav	239 ± 198	194 ± 188	181 ± 130	<b>180 ± 145</b>	.861 ± .05	.922 ± .03	<b>.940 ± .01</b>	.923 ± .01
a-myo	301 ± 243	223 ± 179	<b>199 ± 138</b>	207 ± 141	.852 ± .04	.892 ± .02	.923 ± .01	<b>.931 ± 0.01</b>
<i>promedio</i>	270 ± 154	208 ± 141	<b>190 ± 122</b>	193 ± 115	.857 ± .04	.907 ± .03	<b>.932 ± .01</b>	.927 ± .01
<i>Dimensiones (mm)</i>								
dim1	3.05 ± 2.84	2.63 ± 2.01	<b>2.27 ± 1.79</b>	2.31 ± 1.81	.861 ± .03	.925 ± .01	<b>.961 ± .01</b>	.952 ± .01
dim2	3.23 ± 3.02	2.80 ± 1.89	<b>2.38 ± 1.90</b>	2.41 ± 2.03	.879 ± .06	.932 ± .02	.957 ± .01	<b>.961 ± .01</b>
dim3	3.27 ± 3.12	2.56 ± 1.75	<b>2.22 ± 1.78</b>	2.23 ± 1.67	.912 ± .04	.939 ± .02	<b>.963 ± .01</b>	.959 ± 0.01
<i>promedio</i>	3.18 ± 2.54	2.66 ± 1.75	<b>2.29 ± 1.59</b>	2.31 ± 1.62	.884 ± .04	.932 ± .02	<b>.960 ± .01</b>	.957 ± .01
<i>Espesores de la pared regional (mm)</i>								
wt1 (IS)	2.02 ± 1.32	1.89 ± 1.04	<b>1.23 ± 1.14</b>	1.24 ± 1.17	.625 ± .06	.793 ± .05	<b>.854 ± .01</b>	.846 ± .01
wt2 (I)	2.67 ± 1.69	2.45 ± 1.48	1.44 ± 1.22	<b>1.43 ± 1.87</b>	.618 ± .05	.751 ± .03	.797 ± .01	<b>.801 ± .01</b>
wt3 (IL)	2.95 ± 2.01	1.74 ± 1.56	<b>1.57 ± 1.41</b>	1.60 ± 1.59	.595 ± .04	.735 ± .03	<b>.765 ± .01</b>	.740 ± .01
wt4 (AL)	2.77 ± 1.65	1.66 ± 1.17	1.48 ± 1.13	<b>1.46 ± 1.45</b>	.603 ± .05	.763 ± .02	<b>.785 ± .02</b>	.782 ± .01
wt5 (A)	3.06 ± 2.12	1.49 ± 1.35	<b>1.35 ± 1.19</b>	1.39 ± 1.21	.642 ± .06	.808 ± .02	.842 ± .01	<b>.851 ± .01</b>
wt6 (AS)	2.25 ± 1.72	1.65 ± 1.11	<b>1.46 ± 1.32</b>	1.49 ± 1.37	.651 ± .04	.825 ± .03	<b>.870 ± .01</b>	.866 ± 0.01
<i>promedio</i>	2.62 ± 2.10	1.81 ± 1.05	<b>1.42 ± 0.65</b>	1.43 ± 0.71	.622 ± .05	.779 ± .03	<b>.819 ± .01</b>	.814 ± 0.01
<i>Fase cardíaca (ER %)</i>								
	$\kappa = 1$	$\kappa = 3$	$\kappa = 5$	$\kappa = 7$				
<i>Fase</i>	28.45 ± 5.50	14.67 ± 3.65	<b>3.85 ± 2.82</b>	3.91 ± 2.76				

**Tabla 1.** Resultados obtenidos utilizando el método propuesto realizando validación cruzada.

## BIBLIOGRAFÍA BÁSICA

- Karamitsos, T.D., Francis, J.M., Myerson, S., Selvanayagam, J.B., Neubauer, S.** 2009. The role of cardiovascular magnetic resonance imaging in heart failure. *Journal of the American College of Cardiology* 54(15), 1407–1424
- Peng, P. et al.** 2016: A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. *Magnetic Resonance Materials in Physics, Biology and Medicine* 29(2), 155–195
- Poudel, R.P., Lamata, P., Montana, G.** 2016: Recurrent fully convolutional neural networks for multi-slice mri cardiac segmentation. In: *Reconstruction, Segmentation, and Analysis of Medical Images*, pp. 83–94. Springer
- Suinesiaputra, A. et al.** 2015: Quantification of lv function and mass by cardiovascular magnetic resonance: multi-center variability and consensus contours. *Journal of Cardiovascular Magnetic Resonance* 17(1), 63
- Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M.** 2015: Learning spatiotemporal features with 3d convolutional networks. In: *ICCV 2015* pp. 4489–4497. IEEE
- Xue, W., Lum, A., Mercado, A., Landis, M., Warrington, J., Li, S.** 2017. Full quantification of left ventricle via deep multitask learning network respecting intra-and inter-task relatedness. In: *MICCAI 2017*. pp. 276–284. Springer.
- Xue, W., Brahm, G., Pandey, S., Leung, S., Li, S.** 2018: Full left ventricle quantification via deep multitask relationships learning. *Medical image analysis* 43, 54–65