



# UCA

PONTIFICIA  
UNIVERSIDAD CATÓLICA  
ARGENTINA  
Santa María de los Buenos Aires

## FACULTAD DE INGENIERÍA Y CIENCIAS AGRARIAS

### 1- PROYECTO

**1.1 Título:** Machine Learning aplicado a radiografías de tórax para asistencia al diagnóstico radiológico pediátrico

**1.2 Área Temática**

Disciplina: Ciencias de Datos

Especialidad: Machine Learning. Deep Learning. Procesamiento de Imágenes. Ingeniería de Datos.

**1.3 Área Prioritaria:** Ciencias de Datos

**1.4 Tipo de Proyecto:** Investigación. Implementación.

**1.5 Lugar de Trabajo:** UCA.

### 2- RESPONSABLES

**2.1 Director**

**Apellido y nombre:** Di Pasquale, Ricardo Héctor

**Cargo Docente:** Profesor Pro Titular en la carrera de Ingeniería Informática (Bases de Datos y Bases de Datos Avanzadas y Big Data)

**Dedicación:** 6hs

**Títulos académicos obtenidos:** Ingeniero en Informática (UCA – 2002). Magíster en Ingeniería de Software (ITBA – Universidad Politécnica de Madrid – 2012). Doctor en Ciencia y Tecnología (Universidad Nacional de General Sarmiento - defensa de tesis en trámite).

## 3. PLAN DE INVESTIGACIÓN

### 3.1 Resumen

El objetivo del presente trabajo consiste en implementar un agente de software capaz de asistir al radiólogo pediatra con la clasificación de placas radiográficas (rayos x) frontales de tórax en función de una de las patologías seleccionadas como objeto principal de estudio.

La Ciencia de Datos mediante las técnicas de clasificación de imágenes pone a disposición de los radiólogos una serie de herramientas para su apoyo. Estas herramientas suelen estar asistidas con lo que comúnmente se conoce como Deep Learning, generalmente basadas en el entrenamiento de redes neuronales convolucionales (CNN). Estas redes pueden ser entrenadas para reconocer patrones (sean patologías o lo que busque identificar el radiólogo) en imágenes radiográficas.

### 3.2 Palabras claves

Ciencia de Datos, Machine Learning, Deep Learning. Ingeniería de Datos, procesamiento de imágenes.

### 3.3 Estado actual del conocimiento sobre el tema

La asistencia a los radiólogos por medio de herramientas de vision computacional, mediante el entrenamiento de redes neuronales, es un tema cada vez más presente en los congresos especializados. Cabe destacar que la especialización pediátrica de la radiología tiene aspectos muy similares a la radiología para adultos, es, en sí misma, una especialización. El Hospital Garrahan posee en su ámbito docente un área específica para la formación de radiólogos pediatras (mediante programas de becas y de residencia). Dicho esto ultimo, es importante en la revision bibliográfica, hacer notar la poca cantidad de antecedents de análisis de imágenes pediátricas en la literature asociada al aprendizaje automático en imágenes.

Algunos antecedente en la literature avalados tanto por los profesionales de la medicina, como por ingenieros son:

- “CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning” - Pranav Rajpurkar, Jeremy Irvin et al.
- “Deep Convolutional Neural Networks for Endotracheal Tube Position and X-ray Image Classification: Challenges and Opportunities” - Paras Lakhani1 - J Digit Imaging (2017) 30:460–468 - DOI 10.1007/s10278-017-9980-7

De la misma manera, hacemos notar la difusión de jornadas de Inteligencia Artificial en Radiología, de la Universidad de Stanford:

- <https://www.youtube.com/watch?v=Gigd1rkZTSE>
- <https://www.youtube.com/watch?v=VJRCj-4E2iU>

Y, también, los desafíos y certámenes propuestos por instituciones de prestigio en las plataformas de divulgación de Ciencias de Datos (en particular, en este caso, Kaggle):

- <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge>
- <https://www.kaggle.com/c/rsna-intracranial-hemorrhage-detection>
- <https://www.kaggle.com/c/siim-acr-pneumothorax-segmentation>

### 3.4 Objetivos e hipótesis de la investigación

El objetivo del presente trabajo consiste en implementar un agente de software capaz de asistir al radiólogo pediatra con la clasificación de placas radiográficas (rayos x) frontales de tórax en función de una de las patologías seleccionadas como objeto principal de estudio.

La principal hipótesis del Proyecto se sostiene sobre los avances desarrolladores en Deep Learning respecto del procesamiento de imágenes radiológicas y los hallazgos asociados a tal análisis. Estos procesos pueden ser tomados como antecedents, pero sirven también como

fundamento para intentar trasladar resultados de la medicina en general a la práctica pediátrica, que tiene sus particularidades.

### 3.5 Metodología

En base a la experiencia de los especialistas y al estudio de la bibliografía existente en problemas de clasificación de imágenes radiográficas; se considerarán los escenarios posibles para la implementación del modelo aplicado a imágenes específicamente pediátricas. Esto implica la ponderación de escenarios para brindar a las partes la elección de implementación del escenario más adecuado. Una vez seleccionado el objeto de estudio (la patología), se procederá a implementar el modelo específico, lo que implica un trabajo conjunto de recopilación de una cantidad a determinar de imágenes clasificadas por expertos sobre los que se llevará a cabo un primer proceso de entrenamiento masivo del modelo. Se aplicará el ciclo de vida estándar en Data Science<sup>1</sup> con el fin de lograr un umbral de eficiencia adecuado del modelo. Una vez validado el modelo se procederá a implementar los procesos de ingeniería de datos necesarios para la integración a los estándares DICOM<sup>2</sup> (tanto de formato como de query) sobre PACS<sup>3</sup> de forma de permitir la integración tanto de la ejecución del modelo, como del proceso de entrenamiento.

### 3.6 Desarrollo del Trabajo

Luego de una etapa de inyección por parte del equipo de trabajo, se procederá a evaluar las alternativas de trabajo planteadas por la Dirección Académica del Servicio de Diagnóstico por Imágenes del Htal. Garrahan:

- Ubicación del tubo endotraqueal
- Neumotórax, Neumonía u otro hallazgo relacionado

Estas alternativas se evaluarán en función de varios factores (relacionados tanto con problemática médica como informática) como pueden ser:

- Complejidad del modelo
- Cómputo necesario para el proceso de entrenamiento
- Cantidad de placas estimadas
  - Con la patología buscada
  - Sin la patología buscada
    - Con otras patologías
    - Sin otras patologías.
- Posibilidades de extracción de las imágenes.
- Necesidad de trabajo extra radiológico de clasificación de las imágenes ●
- Pre procesamiento de las imágenes.
- Dependencia de la pericia técnica en la adquisición.
- Posibilidad de extensión de enfoques aplicados y validados (ML) en adultos a la problemática pediátrica.

En función de lo anterior, se seleccionará una patología como objetivo central de este trabajo.

<sup>1</sup> Stodden, Victoria, "The Data Science Life Cycle: A Disciplined Approach to Advancing Data Science as a Science", July 2020, Association for Computing Machinery, New York, NY, USA}, Commun. ACM Vol 63 Nr 7, issn 0001-0782, pp 58–66,

<sup>2</sup> The DICOM Standard is managed by the Medical Imaging & Technology Alliance - a division of the National Electrical Manufacturers Association. <https://www.dicomstandard.org/current>

<sup>3</sup> "PACS fundamentals" in "PACS-Based Multimedia Imaging Informatics: Basic Principles and Applications", Third Edition, H.K. Huang D.Sc., FRCR (Hon), FAIMBE, Willey Ed., 2018, ISBN: 9781118795552.

Para la construcción del modelo se trabajará con el lote de imágenes provistos por el hospital (en función de lo determinado en la etapa anterior). El foco en esta etapa estará en la elaboración del modelo de ML más adecuado, así como en el entrenamiento de dicho agente, así como en el pre-procesamiento de las imágenes. Como en todo ciclo de vida de Data Science, se validará iterativamente la eficacia del modelo retroalimentando el proceso.

Una vez logrado un umbral aceptable, se implementarán los procesos necesarios de ingeniería de datos para automatizar el entrenamiento, así como facilitar la integración mediante los estándares DICOM / PACS asociados a la informática médica.

### 3.7 Bibliografía

- “CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning” - Pranav Rajpurkar, Jeremy Irvin et al.
- “Deep Convolutional Neural Networks for Endotracheal Tube Position and X-ray Image Classification: Challenges and Opportunities” - Paras Lakhani1 - J Digit Imaging (2017) 30:460–468 - DOI 10.1007/s10278-017-9980-7

## 4. DESARROLLO DEL PROYECTO

### 4.1 Cronograma de Actividades

Actividad	Años		
<b>Etapas de Desarrollo del Trabajo</b>			
Incepción	2021		
Exploración de Datos	2021		
Elaboración de Prototipos	2021		
Selección de la patología a trabajar	2021		
Extracción set inicial de Datos	2021		
Elaboración modelo v1	2021		
Ing de Datos v1	2021		
Puesta en Producción modelo v1	2021		
Iteración v2	2021	2022	
Iteración v3		2022	
<b>Comentarios y aclaraciones:</b>			

### 4.2 Actividades de Transferencia

- La automatización de las extracciones y entrenamientos requiere una transferencia para con el equipo del Hospital.

### 4.3 Vinculación del proyecto con la actividad docente desarrollada en UCA

- En el ámbito de la carrera de informática:
  - El proyecto servirá como una plataforma práctica para ámbitos como:
    - Ciencias de Datos

- Ingeniería de Datos
- Big Data
- Posee una vinculación directa con las asignaturas de:
  - Estructuras de Datos
  - Bases de Datos
  - Bases de Datos Avanzadas y Big Data
  - Modelos y Simulaciones
  - Inteligencia Artificial (opt)
  - Seminario Metodológico de Trabajo Final
- También tiene una importancia fundamental para los trabajos finales. Muchos alumnos pueden basar su TF en una colaboración en esta línea de investigación.
- La coherencia de trabajos en esta línea fomentará la formación de una comunidad de alumnos y docentes en la orientación a datos (ciencias e ingeniería de datos) que puede dejar una impronta importante en los egresados. Es importante destacar que no hay oferta académica que refuerce uno de los perfiles más necesitados por la industria del software hoy en día: el ingeniero de datos. Este proyecto puede aportar mucho en ingeniería de datos, y en generar la impronta en la que la Facultad puede destacar.

#### 4.4 Vinculación del proyecto con problemas de la Comunidad

La colaboración con una institución de relevancia nacional e internacional como el Hospital Garrahan es fundamental para los proyectos de investigación de nuestra Facultad. El reconocimiento de la comunidad para con los profesionales del Hospital es de público conocimiento, por lo que no es necesario realizar otra ponderación sobre la relevancia del proyecto en la problemática de la comunidad.

### 5. PERSONAL ASIGNADO AL PROYECTO

5.1 Completar la tabla de datos para cada uno de los integrantes en el siguiente orden: Director, Codirector, Investigadores e Investigadores en formación.

#### 5.1.1. Por la UCA

<b>Función:</b>	Director
<b>Apellido y Nombre:</b>	Di Pasquale, Ricardo Héctor
<b>Tipo y No. Documento:</b>	DNI 26.518.829
<b>No. de Legajo en UCA:</b>	42331-3
<b>Lugar y Fecha de Nacimiento:</b>	Buenos Aires, 15/03/1978
<b>Nacionalidad:</b>	Argentino
<b>Domicilio:</b>	Quintino Bocayuva 682
<b>TE Particular/celular:</b>	+5491169757993
<b>E -mail:</b>	<a href="mailto:rdipasquale@uca.edu.ar">rdipasquale@uca.edu.ar</a>
<b>Título de Grado:</b>	Ingeniero en Informática
<b>Máximo Título Obtenido:</b>	Magíster en Ingeniería de Software
<b>Cargo Docente:</b>	Protitular

<b>Si reviste como investigador en otra Institución (Ej.: CONICET, etc.), consignar:</b>	<b>Institución</b>	<b>Cargo</b>	<b>Dedicación</b>
--	--------------------	--------------	-------------------

<b>Función:</b>	Investigador
<b>Apellido y Nombre:</b>	Quintero Rincón, Antonio
<b>Tipo y No. Documento:</b>	DNI 93.964.856
<b>No. de Legajo en UCA:</b>	482163
<b>Lugar y Fecha de Nacimiento:</b>	Manizales, 13/12/1971
<b>Nacionalidad:</b>	Colombiano
<b>Domicilio:</b>	Viamonte 2932 CABA
<b>TE Particular/celular:</b>	+5491135801897
<b>E -mail:</b>	<a href="mailto:tonioquintero@gmail.com">tonioquintero@gmail.com</a>
<b>Título de Grado:</b>	Ingeniero Electrónico
<b>Máximo Título Obtenido:</b>	Doctor en Ingeniería
<b>Cargo Docente:</b>	Protitular

<b>Si reviste como investigador en otra Institución (Ej.: CONICET, etc.), consignar:</b>	<b>Institución</b>	<b>Cargo</b>	<b>Dedicación</b>
--	--------------------	--------------	-------------------

## 6. ALUMNOS COLABORADORES

### 6.1 Por la UCA

- Héctor Buena Maizon
- Julián Maestri
- Martín Pordomingo
- Francisco Ortiz Peñaloza