

EVALUACIÓN DE MÉTODOS CON REDES NEURONALES  
CONVOLUCIONALES PARA LA CLASIFICACIÓN DE LA ENFERMEDAD DE  
ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA CEREBRAL

JUAN MANUEL ORTIZ SUÁREZ

UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FISICO-MECANICAS  
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA  
BUCARAMANGA

2016

EVALUACIÓN DE MÉTODOS CON REDES NEURONALES  
CONVOLUCIONALES PARA LA CLASIFICACIÓN DE LA ENFERMEDAD DE  
ALZHEIMER EN IMÁGENES DE RESONANCIA MAGNÉTICA CEREBRAL

JUAN MANUEL ORTIZ SUÁREZ

Trabajo de Grado para optar al título de  
Ingeniero de Sistemas

Director

RAÚL RAMOS POLLÁN

PhD, Informatics Engineering

UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FÍSICO-MECÁNICAS  
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA  
BUCARAMANGA

2016

## CONTENIDO

|  | Pág. |
|--|------|
| INTRODUCCIÓN                                       | 11   |
| 1. OBJETIVOS                                       | 12   |
| 1.1 OBJETIVO GENERAL                               | 12   |
| 1.2 OBJETIVOS ESPECIFICOS                          | 12   |
| 2. MARCO DE REFERENCIA                             | 13   |
| 2.1 FUNDAMENTOS TEÓRICOS                           | 13   |
| 2.2 ANTECEDENTES DEL TEMA                          | 15   |
| 3. METODOLOGÍA                                     | 17   |
| 3.1 DATASET  | 17   |
| 3.2 LÍNEA DE BASE                                  | 17   |
| 3.3 PREPROCESAMIENTO DE IMÁGENES                   | 20   |
| 3.4 DISEÑO DE LAS REDES NEURONALES CONVOLUCIONALES | 21   |
| 3.5 DATA SETS Y ENTRENAMIENTO                      | 23   |
| 3.6 EVALUACIÓN DE REDES NEURONALES                 | 24   |
| 3.6.1 Selección de la mejor configuración de red   | 25   |
| 3.6.2 Pruebas experimentales                       | 25   |
| 3.7 EQUIPO Y SOFTWARE                              | 26   |
| 3.7.1 Hardware                                     | 26   |
| 3.7.2 Software                                     | 26   |
| 4. RESULTADOS Y ANALISIS                           | 27   |
| 4.1 BOOTSTRAPPING                                  | 27   |
| 4.2 VALIDACIÓN CRUZADA DEJANDO UNO FUERA           | 29   |
| 4.3 ÁREA BAJO LA CURVA ROC                         | 30   |

|  |    |
|--|----|
| 4.4 COMPARACIÓN EN RENDIMIENTO COMPUTACIONAL | 31 |
| 4.5 RESULTADOS FINALES                       | 31 |
| 5. CONCLUSIONES                              | 33 |
| BIBLIOGRAFÍA                                 | 34 |

## LISTA DE TABLAS

|   | Pág. |
|---|------|
| Tabla 1. Información de la población  | 17   |
| Tabla 2. Arquitectura de redes neuronales   | 22   |
| Tabla 3. Distribución de los datasets   | 24   |
| Tabla 4. Resultados para la selección de la mejor red neuronal                      | 25   |
| Tabla 5. Resultados de la prueba bootstrapping                                      | 27   |
| Tabla 6. Resultados de la prueba de validación cruzada para la línea base           | 29   |
| Tabla 7. Resultados de la prueba de validación cruzada para<br>Las redes neuronales | 29   |
| Tabla 8. Tabla comparativa de rendimiento computacional                             | 31   |
| Tabla 9. Tabla comparativa de métodos de clasificación                              | 32   |

## LISTA DE FIGURAS

|  | Pág. |
|--|------|
| Figura 1. Modelo de una neurona artificial                                 | 13   |
| Figura 2. Modelo de redes neuronales de una y dos capas                    | 14   |
| Figura 3. Modelo de una red neuronal convolucional típica                  | 15   |
| Figura 4. Representación gráfica del problema de factorización de matrices | 18   |
| Figura 5. Curva ROC para la línea de base                                  | 19   |
| Figura 6. Representación gráfica del pre procesamiento de las imágenes     | 21   |
| Figura 7. Representación del proceso experimental                          | 24   |
| Figura 8. Curvas ROC para cada prueba.                                     | 30   |

## RESUMEN

**TÍTULO:** “Evaluación de métodos con redes neuronales convolucionales para la clasificación de la enfermedad de Alzheimer en imágenes de resonancia magnética cerebral”\*.

**AUTORES:** JUAN MANUEL ORTIZ SUÁREZ \*\*

**PALABRAS CLAVE:** Clasificación, redes neuronales, Alzheimer.

**DESCRIPCIÓN:** La enfermedad de Alzheimer es el tipo más común de demencia que afecta a personas mayores de 60 años, aunque el Alzheimer es incurable, el diagnóstico temprano es muy importante para tratar de atenuar los síntomas que empeoran con la edad.

En este trabajo de investigación se propone el uso de redes neuronales convolucionales para la clasificación de la enfermedad de Alzheimer en etapa temprana a partir de imágenes de resonancia magnética cerebral de pacientes entre 60 y 80 años de edad. Las redes neuronales convolucionales han demostrado ser eficaces en tareas de visión computacional como clasificación de imágenes, reconocimiento facial y detección de objetos. Como primer paso se replica un método de clasificación del Alzheimer basado en generación de modelos a partir de la fusión de información neuropsicológica y estructural para tener una línea de base con la cual comparar el rendimiento de las redes neuronales. Seguido de esto se implementan métodos de pre procesamiento de las imágenes de resonancia magnética para posibilitar el entrenamiento de las distintas configuraciones de redes neuronales. Usando datos de validación se selecciona la mejor arquitectura de red y posteriormente con los datos de prueba y diferentes métodos de evaluación se obtienen métricas que reflejen adecuadamente el rendimiento de las redes neuronales convolucionales.

---

\* Trabajo de grado

\*\* Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Raúl Ramos Pollan, PhD. Informatics Engineering.

## ABSTRACT

**TITLE:** “Evaluation of methods with convolutional neural networks for Alzheimer’s disease classification over cerebral magnetic resonance images.”\*

**AUTHORS:** JUAN MANUEL ORTIZ SUAREZ.\*\*

**KEY WORDS:** Classification, neural networks, Alzheimer’s disease.

**DESCRIPTION:** Alzheimer’s disease is the most common type of dementia that affects people over sixty years old, although the disease is incurable, early diagnostic is very important to control symptoms which become worst in the last stages of the illness.

In this research it is proposed the application of convolutional neural networks for Alzheimer’s disease classification in early stages using cerebral magnetic resonance images from patients between sixty and eighty years old. Convolutional neural networks have been shown effectiveness in computational vision tasks like general imaging classification, facial recognition and object detection. In a first phase of this research a classification method of the Alzheimer’s disease has been replicated, this method consists in generate models from neuropsychological and structural information and it is used to obtain results which compare to the ones from the performance evaluation of convolutional neural networks. In the second, phase magnetic resonance images have to be processed to make it possible the training of the different neural network architectures. Finally, with validation images, the best network architecture is selected and then it is tested with other images and different evaluation methods to get the metrics which show the performance of convolutional neural network for this specific problem.

---

\* Bachelor Thesis

\*\* Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Raúl Ramos Pollan, PhD. Informatics Engineering.

## INTRODUCCIÓN

El Alzheimer es una enfermedad neurodegenerativa y es la forma más común de demencia que afecta a las personas de la tercera edad. Este desorden cerebral avanza lentamente y afecta en sus primeras etapas partes del cerebro encargadas del pensamiento, la memoria y el lenguaje.

Actualmente no existen tratamientos que curen esta enfermedad, pero sí medicamentos que ayudan al control de los síntomas previniendo que estos empeoren por un tiempo limitado, por lo tanto, el diagnóstico temprano del Alzheimer es muy importante a la vez que difícil ya que puede confundirse con otros tipos de demencia o incluso con el envejecimiento normal del cerebro. Como métodos de diagnóstico se usan frecuentemente pruebas psicológicas que buscan cuantificar el grado de demencia e imágenes de resonancia magnética que ayudan a detectar partes del cerebro afectadas.

A partir de lo anterior se plantea en este trabajo de investigación el uso de Redes Neuronales Convolucionales para el procesamiento y clasificación de imágenes de resonancia magnética cerebral. Este particular tipo de red neuronal, basado en la corteza visual primaria del cerebro biológico, ha demostrado ser muy eficaz en tareas de reconocimiento de imágenes, clasificación de objetos y análisis de video. Además, gracias al avance tecnológico de los últimos años, el uso de unidades de procesamiento gráfico GPU han mejorado significativamente la eficiencia a la hora de entrenar redes neuronales convolucionales y otros algoritmos de Machine Learning.

Concretamente en este proyecto se buscó evaluar el desempeño de distintas configuraciones de redes neuronales convolucionales con dos de los mejores frameworks actuales usados para el desarrollo de algoritmos de machine learning y teniendo como línea de base el trabajo desarrollado por el grupo de investigación CIM@LAB de la Universidad Nacional de Colombia que consiste en la generación de modelos para detectar el Alzheimer a partir de información estructural y neuropsicológica.

# 1. OBJETIVOS

## 1.1 OBJETIVO GENERAL

Diseñar e implementar redes neuronales convolucionales (CNN) y evaluar su desempeño en el procesamiento y clasificación de imágenes de resonancia magnética cerebral como apoyo al diagnóstico de la enfermedad de Alzheimer.

## 1.2 OBJETIVOS ESPECIFICOS

- Diseñar e implementar un método para posibilitar el procesamiento de las imágenes de resonancia magnética cerebral del dataset disponible en [www.oasis-brains.org](http://www.oasis-brains.org) en redes neuronales convolucionales.
- Diseñar un conjunto de redes neuronales convolucionales para la clasificación de la enfermedad de Alzheimer sobre el resultado del objetivo anterior.
- Evaluar el desempeño de las distintas redes convolucionales con las herramientas CAFFE y TORCH.
- Evaluar el desempeño en clasificación y coste computacional de las redes neuronales convolucionales anteriormente diseñadas con el dataset procesado.
- Identificar qué tipos de redes neuronales convolucionales son más adecuadas para este problema.

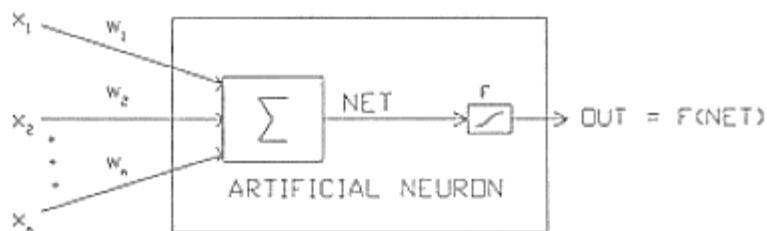
## 2. MARCO DE REFERENCIA

### 2.1 FUNDAMENTOS TEÓRICOS

Las redes neuronales artificiales (RNA) son un paradigma de programación orientado al aprendizaje automático inspirado en la organización del sistema neuronal humano cuya unidad de procesamiento fundamental es un tipo de célula especial conocida como neurona la cual está conectada con otras neuronas mediante estructuras llamadas axones por donde se transmiten señales electroquímicas hacia las dendritas. La conexión entre dendritas es conocida como sinapsis la cual puede excitar o inhibir la neurona según supere o no un umbral de excitación.

La neurona artificial fue diseñada para emular las características funcionales de la neurona biológica, esta recibe un conjunto de entradas multiplicadas cada una por un peso correspondiente análogo al grado de conexión de la sinapsis. Las entradas ponderadas se suman y mediante una función que puede ser lineal, umbral o no lineal se procesa dicha suma y se produce así una salida.

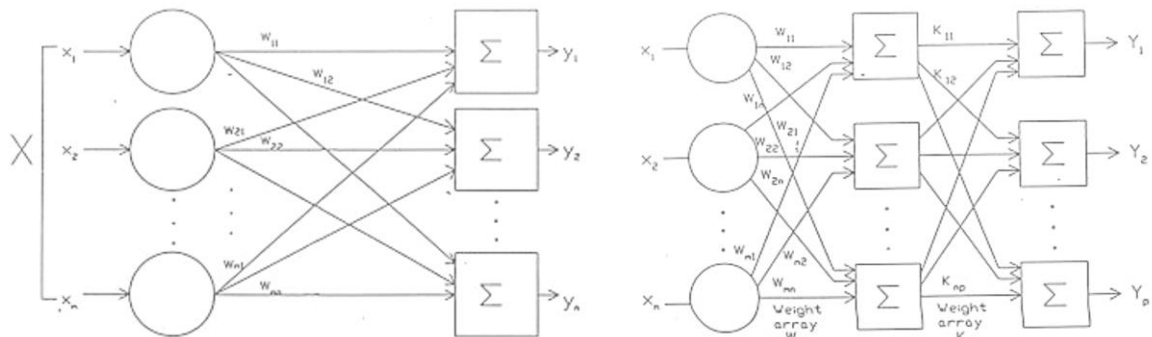
Figura 1. Modelo de una neurona artificial.



Fuente: Basogain X. *Redes Neuronales Artificiales y Sus Aplicaciones*.

Una ventaja de las redes neuronales es que se pueden configurar de muchas formas para adecuarse a la tarea para las que se requieren. Existen dos principales configuraciones de redes neuronales, la más simple que consiste en un grupo de neuronas ordenadas en una capa, estas son redes de una capa, y existen también configuraciones más complejas que ofrecen mejores prestaciones en el cálculo computacional, son redes compuestas por varias capas simples en cascada de denominadas redes multicapa.

Figura 2. Modelos de Redes Neuronales de una y dos capas.



Fuente: Basogain X. *Redes Neuronales Artificiales y Sus Aplicaciones*.

Las redes neuronales convolucionales son un tipo especial de red neuronal artificial multicapa diseñadas para trabajar específicamente con imágenes bidimensionales. Usan una serie de máquinas de Boltzmann, la cuales son redes neuronales de dos capas, una “visible” y otra “oculta”, donde las conexiones se forman entre capas pero no existen conexiones entre la misma capa. Las imágenes van a través de las capas que filtran la información de la imagen entre lo que se considera importante y lo que no. El resultado final es una especie de mapa de la imagen. Procesar un gran número de imágenes entrena al sistema para que “aprenda” en que fijarse al momento de reconocer futuras imágenes de entrada de la misma categoría.

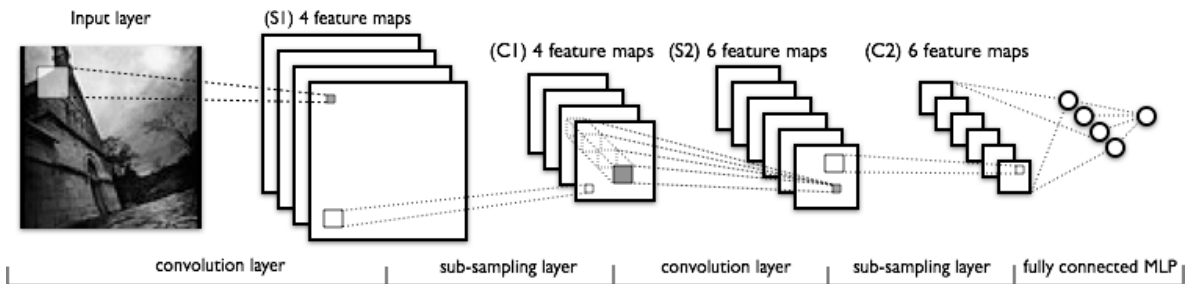
En el diseño de las redes neuronales convolucionales se usan siempre tres tipos de capas: capa convolucional, capa de reducción de muestro o pooling y la capa de clasificación conocida como fully-connected que suele ir al final de la red. A continuación se explica con más detalle el funcionamiento de cada capa en una arquitectura de red sencilla, en su orden:

- ENTRADA: Una imagen de dimensiones  $M \times N \times 3$  si es una imagen de tres canales de color (rojo, verde y azul) o de  $M \times N \times 1$  si es una imagen a blanco y negro.
- CAPA CONVOLUCIONAL: computará la salida de las neuronas que están conectadas a regiones locales de la entrada, esto consiste en el cálculo del producto punto entre los pesos de cada neurona y la región a la cual están conectadas, la salida se puede ver como un volumen de  $M \times N \times K$ , siendo  $K$  el parámetro de profundidad de la capa.
- CAPA ReLU: (abreviación de Rectified Linear Units) es una capa que aplica una función de activación que incrementa las propiedades no-lineales de decisión. Esta capa no afecta el volumen de salida.
- CAPA DE POOLING: esta capa realiza una operación de sub-muestreo que reduce el volumen de salida en sus dimensiones espaciales (ancho y alto) dejando intacta la profundidad.

- **CAPA DE CLASIFICACIÓN:** computa el puntaje para cada clase dando como resultado un volumen de  $1 \times 1 \times C$ , donde  $C$  sería el número de clases definidas en el problema.

Así las redes neuronales convolucionales transforman la imagen original capa por capa de valores por pixel hasta puntajes por clase.

Figura 3. Modelo de una red neuronal convolucional típica.



Fuente: [www.deeplearning.net](http://www.deeplearning.net) *Convolutional Neural Networks (LeNet)*

## 2.2 ANTECEDENTES DEL TEMA

El reconocimiento de objetos resulta ser una tarea trivial para los complejos sistemas nerviosos de humanos y animales en general. Hoy en día, aun con el abrumador avance tecnológico, para los computadores y sistemas, capaces de procesar millones de datos en segundos, a la hora de reconocer objetos o imágenes no resulta ser una tarea fácil que se ve principalmente limitada por la capacidad de procesamiento, capacidad que sí tiene el complejo cerebro de los animales gracias a la inmensa cantidad de neuronas, aproximadamente 100.000 millones en el cerebro humano cada una con cerca de 7000 interconexiones (sinapsis).

Desde su aparición, propuestas por Yann LeCun a principios de la década de los 90's, las redes neuronales convolucionales han demostrado excelente rendimiento en tareas como clasificación de dígitos escritos a mano y reconocimiento facial. A partir de lo anterior y junto a los avances tecnológicos de las últimas dos décadas, se han desarrollado diversas aplicaciones en diferentes campos haciendo uso de estas redes neuronales. Específicamente en el campo de la medicina las redes neuronales, en general se han usado en áreas como son, el diagnóstico médico, análisis bioquímico en muestras de sangre y orina, desarrollo de medicamentos y análisis de imágenes. Para este último tópico es que las redes neuronales convolucionales (CNNs) han venido demostrando su potencial ya que en el campo de la medicina, en muchos estudios se usan imágenes bidimensionales, como radiografías e imágenes de resonancia magnética, que resultan apropiadas para el análisis por parte de las redes neuronales.

Cuando se analizan imágenes médicas lo importante es identificar objetos como lesiones u organismos que no siempre se hacen presentes de la misma forma, por ejemplo, los nódulos pulmonares generalmente se presentan como una esfera sólida, pero en algunos casos los mismo aparecen con varias formas incluso no homogéneas, también los pólipos en el colon se muestran mayormente con forma de bulbo pero en algunas ocasiones aparecen con una forma aplanada. Así, esencialmente la tarea de diagnóstico en imágenes se resume a “aprender de los ejemplos (datos)”.

La función de clasificación de las redes neuronales convolucionales en el ámbito de la medicina ha aportado a la reducción de falsos positivos en el diagnóstico médico de nódulos pulmonares en radiografías de tórax, micro calcificaciones y masas en mamografías.

### 3. METODOLOGÍA

#### 3.1 DATASET

Para la realización de este trabajo de investigación se obtuvieron imágenes y datos de una base de datos o data set público, The Open Acces Series of Imaging Studies (OASIS) creado por el Alzheimer's Disease Research Center de la Universidad de Washington que contiene información de cuatrocientos dieciséis sujetos o pacientes entre hombres y mujeres con edades de dieciocho a noventa y seis años y diferentes grados de Alzheimer diagnosticado.

Teniendo en cuenta la línea de base, descrita en la sección siguiente, se seleccionó un conjunto de pacientes con edades entre sesenta y ochenta años con un puntaje de 0 ó 1 en la clasificación clínica de demencia, CDR por sus siglas en inglés, la cual es una prueba basada en una entrevista estructurada que cuantifica la severidad de los síntomas de demencia en una escala de 0 a 5, siendo 0 un paciente sin síntomas, 1 un paciente con síntomas leves y 5 un paciente con síntomas severos de demencia.

Tabla 1. Información de la población.

|         | <b>Grado de Demencia</b> | <b>Mujeres</b> | <b>Hombres</b> | <b>Total</b> |
|---------|--------------------------|----------------|----------------|--------------|
| CDR = 0 | Ninguno                  | 48             | 18             | 66           |
| CDR = 1 | Leve                     | 13             | 7              | 20           |
| TOTAL   |                          | 61             | 25             | 86           |

Estos pacientes son los mismos que se usaron en el trabajo que se replicó como línea de base, no se toman pacientes con CDR de 0,5 porque son difíciles de diferenciar del envejecimiento normal de cerebro y en el rango de edades seleccionado se acentúan más los síntomas de Alzheimer.

#### 3.2 LINEA DE BASE

Con el fin de tener indicadores con los cuales comparar los resultados obtenidos de la clasificación con la redes neuronales convolucionales, se replicó el trabajo *Finding Models of the Alzheimer Disease by Fusing Information from Neuropsychological Test and Structural MRI Images* desarrollado por el grupo de investigación Computer Image and Medical Applications Laboratory CIM@LAB de la Universidad Nacional de Colombia y presentado en el simposio internacional organizado por la Sociedad Internacional de Procesamiento y Análisis de la Información Médica SIPAIM del año 2015.

Se decidió usar este trabajo como punto de comparación debido a que en la literatura consultada no se encontraron otras investigaciones que usen las redes neuronales convolucionales para clasificar la enfermedad del Alzheimer usando el dataset OASIS. Además, se tiene la posibilidad de contactar directamente con los autores de las Universidad Nacional de Colombia.

Este trabajo consiste en fusionar información estructural y neuropsicológica y mediante un problema de factorización de matrices generar un modelo para cada grado de demencia, nulo o leve. La información estructural son los volúmenes cerebrales que están representados en forma de arreglos tridimensionales y que posteriormente se transforman en columnas en una matriz bidimensional X como se muestra en la Figura 4.

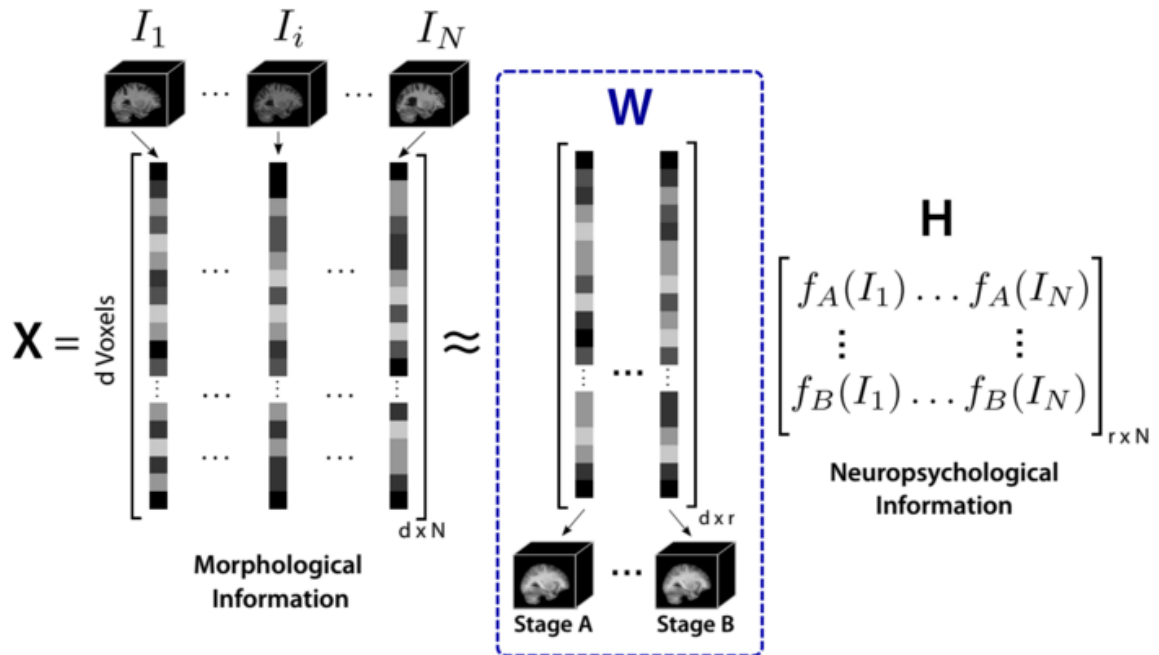
De pruebas psicológicas como el CDR se obtiene la información neuropsicológica la cual se usa para generar valores de pertenencia para ambos grados de demencia para cada paciente, esta información se representa en la matriz H.

Los modelos calculados son entonces las dos columnas de la matriz W. Este problema de factorización de matrices tiene la fórmula:

$$X = WH$$

La cual se ve más claramente representada en la siguiente ilustración.

Figura 4. Representación gráfica del problema de factorización de matrices.



Fuente: *Finding Models of the Alzheimer Disease by Fusing Information from Neuropsychological Test and Structural MRI Images.*

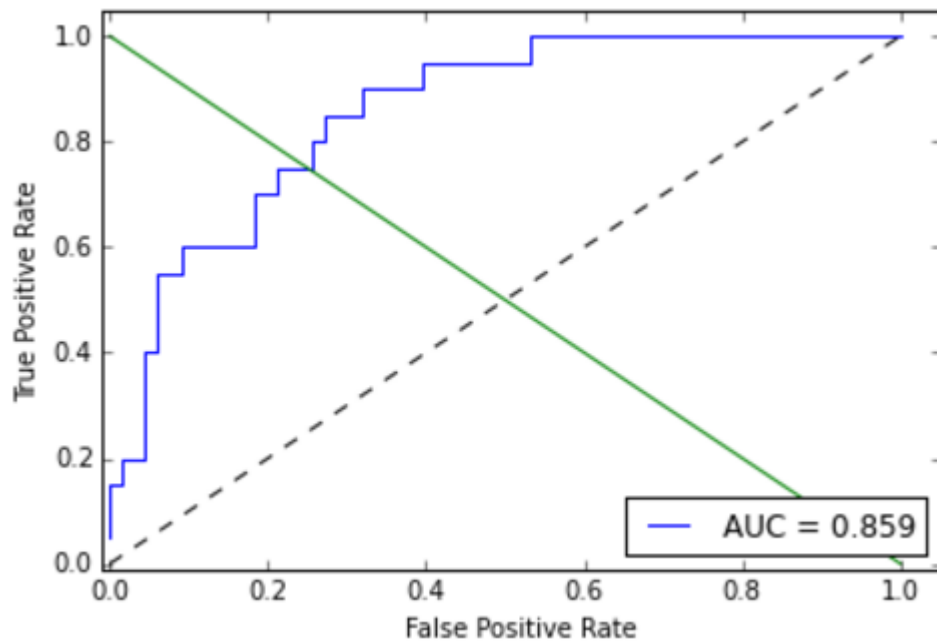
La validación de este método de clasificación se hizo mediante el esquema de validación cruzada dejando uno fuera, esto quiere decir que se deja un único dato, en este caso un paciente, como dato de prueba y todo el resto se usa para entrenar, en este caso calcular los modelos. Este tipo de validación tiene un error bajo pero un costo computacional alto debido a que se tienen que hacer tantas iteraciones como datos se tengan.

En la réplica de este clasificador al aplicar la validación mencionada anteriormente se obtuvo una precisión de 77.907% es decir, de los ochenta y seis pacientes se clasificaron correctamente sesenta y siete e incorrectamente diecinueve.

Se debe mencionar que con la réplica de este trabajo no se obtuvieron los resultados exactos publicados por sus autores. La pequeña diferencia se debe a la falta de información neuropsicológica, procedente de otro dataset, usada para generar la función de pertenencia que asigna los puntajes para cada clase.

Un mejor indicador para conocer el rendimiento de pruebas diagnósticas, como las que en este trabajo se evalúan, es el área bajo la curva ROC o AUC, por sus siglas en inglés. Más adelante se expondrá detalladamente el funcionamiento de este método de evaluación junto con otros, por el momento basta con saber que cuanto más cercano sea el valor del AUC a 1 mejor será el modelo diagnóstico al discernir entre enfermedad y no enfermedad.

Figura 5. Curva ROC para la línea de base.



### **3.3 PREPROCESAMIENTO DE IMÁGENES**

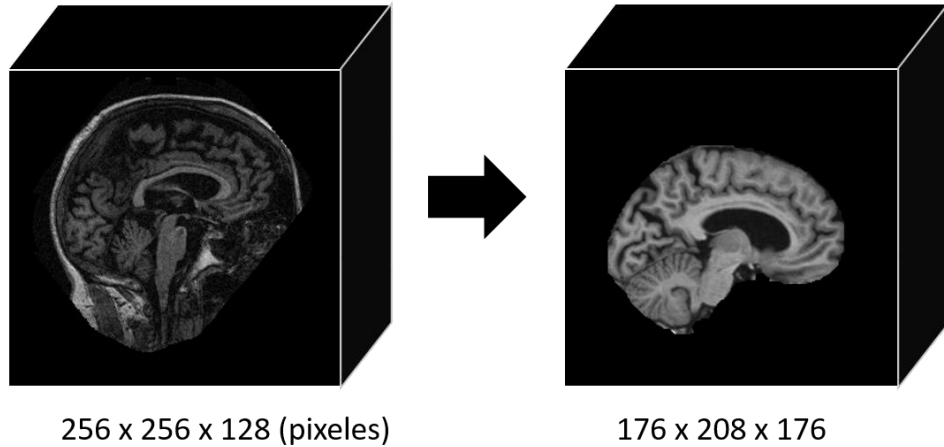
El pre procesamiento de imágenes es un conjunto de técnicas que tienen como objetivo obtener, a partir de una imagen origen, otra final que sea más adecuada para una aplicación específica. Comúnmente estas técnicas se usan para eliminar algunas características o resaltar otras.

Las imágenes usadas en este trabajo son en el resultado de dos etapas de pre procesamiento. La primera, realizada por los creadores del dataset, consiste en promediar las imágenes de varias sesiones de resonancia magnética, eliminar la estructura craneal y ajustarlas espacialmente según un sistema de coordenadas tridimensionales que mapea las distintas partes del cerebro conocido como atlas de Talairach.

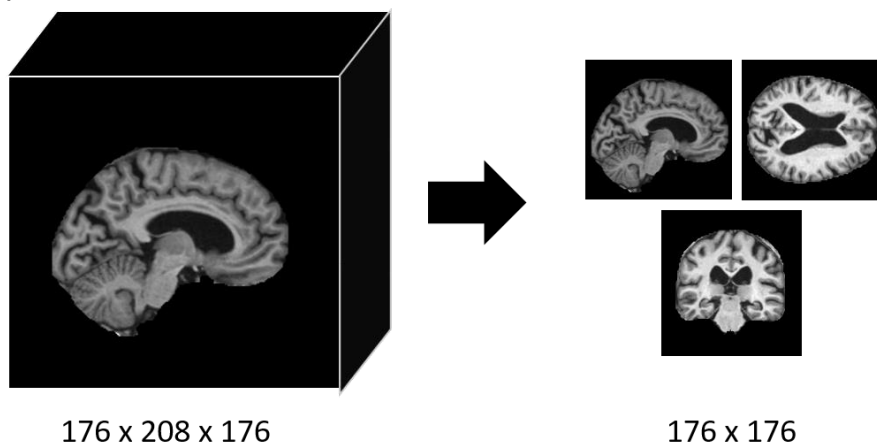
De lo anterior se obtiene una imagen tridimensional corregida, por lo tanto en la segunda etapa de pre procesamiento, realizada en este trabajo, se deben extraer “capas” bidimensionales con las que se puedan entrenar las redes neuronales convolucionales. Esto se consiguió mediante un script escrito en lenguaje Python y con el uso de las librerías NiBabel, especializada en el manejo de imágenes neurológicas y NumPy, fundamental para el manejo de arreglos N-dimensionales. Este proceso consiste en transformar la imagen tridimensional obtenida del primer pre procesamiento en una matriz o arreglo de tres dimensiones del cual se pueden obtener arreglos bidimensionales a lo largo de los ejes sagital, transversal y frontal, anatómicamente hablando, que a su vez se transforman en imágenes bidimensionales las cuales se ajustan finalmente a un tamaño fijo para su uso en el entrenamiento de las redes neuronales convolucionales.

Figura 6. Representación gráfica del pre procesamiento de las imágenes.

Etapa 1



Etapa 2



Como resultado se toman trescientas imágenes por cada volumen cerebral, es decir, cien imágenes por cada eje en su zona central esto debido a que las imágenes de los extremos no son útiles para el entrenamiento de las redes neuronales ya que muestran muy poca área cerebral. Todas las imágenes tienen formato PNG.

### 3.4 DISEÑO DE LAS REDES NEURONALES CONVOLUCIONALES

En este trabajo de investigación se probaron siete distintas configuraciones o arquitecturas de redes. Las principales variaciones entre arquitecturas son: el algoritmo de inicialización de pesos, el número de capas y el tamaño de los filtros. Se usa una estructura básica de dos capas, de convolución y pooling, ya que se vio que el uso de más capas aumenta en gran medida el coste computacional y el tiempo de entrenamiento, pero los resultados finales no mejoran significativamente. En la siguiente tabla se describen las configuraciones de red usadas, los cambios entre cada una se ven resaltados en negrita.

Tabla 2. Arquitecturas de redes neuronales.

| ConvNet 01 (Inicialización de pesos: gaussian)     |            |               |      |            |
|--|------------|---------------|------|------------|
| Capa   | Entrada    | Tamaño filtro | Paso | Salida     |
| Conv1(ReLU)  | 176x176x1  | 3x3           | 1    | 174x174x60 |
| Pool1  | 174x174x60 | 2x2           | 2    | 87x87x60   |
| Conv2(ReLU)  | 87x87x60   | 3x3           | 1    | 85x85x120  |
| Pool2  | 85x85x120  | 2x2           | 2    | 43x43x120  |
| FC   | 221880     |               |      | 2          |
| ConvNet 02 (Inicialización de pesos: <b>MSRA</b> ) |            |               |      |            |
| Capa   | Entrada    | Tamaño filtro | Paso | Salida     |
| Conv1(ReLU)  | 176x176x1  | 3x3           | 1    | 174x174x60 |
| Pool1  | 174x174x60 | 2x2           | 2    | 87x87x60   |
| Conv2(ReLU)  | 87x87x60   | 3x3           | 1    | 85x85x120  |
| Pool2  | 85x85x120  | 2x2           | 2    | 43x43x120  |
| FC   | 221880     |               |      | 2          |
| ConvNet 03 (Inicialización de pesos: MSRA)         |            |               |      |            |
| Capa   | Entrada    | Tamaño filtro | Paso | Salida     |
| Conv1(ReLU + <b>LRN</b> )                          | 176x176x1  | 3x3           | 1    | 174x174x60 |
| Pool1  | 174x174x60 | 2x2           | 2    | 87x87x60   |
| Conv2(ReLU)  | 87x87x60   | 3x3           | 1    | 85x85x120  |
| Pool2  | 85x85x120  | 2x2           | 2    | 43x43x120  |
| FC   | 221880     |               |      | 2          |
| ConvNet 04 (Inicialización de pesos: MSRA)         |            |               |      |            |
| Capa   | Entrada    | Tamaño filtro | Paso | Salida     |
| Conv1(ReLU + <b>LRN</b> )                          | 176x176x1  | 3x3           | 1    | 174x174x60 |
| Pool1  | 174x174x60 | 2x2           | 2    | 87x87x60   |
| Conv2(ReLU)  | 87x87x60   | 3x3           | 1    | 85x85x120  |
| Pool2  | 85x85x120  | 2x2           | 2    | 43x43x120  |
| FC1( <b>ReLU</b> )                                 | 221880     |               |      | 100        |
| <b>FC2</b>   | 100        |               |      | 2          |

| ConvNet 05 (Inicialización de pesos: MSRA)           |            |               |          |            |
|--|------------|---------------|----------|------------|
| Capa   | Entrada    | Tamaño filtro | Paso     | Salida     |
| Conv1(ReLu + LRN)                                    | 176x176x1  | 3x3           | 1        | 174x174x60 |
| Pool1  | 174x174x60 | 2x2           | 2        | 87x87x60   |
| Conv2(ReLu)  | 87x87x60   | 3x3           | 1        | 85x85x120  |
| Pool2  | 85x85x120  | 2x2           | 2        | 43x43x120  |
| FC1(ReLu + <b>Dropout</b> )                          | 221880     |               |          | 100        |
| FC2  | 100        |               |          | 2          |
| ConvNet 06 (Inicialización de pesos: MSRA)           |            |               |          |            |
| Capa   | Entrada    | Tamaño filtro | Paso     | Salida     |
| Conv1(ReLu + LRN)                                    | 176x176x1  | <b>6x6</b>    | <b>2</b> | 86x86x60   |
| Pool1  | 86x86x60   | 2x2           | 2        | 43x43x60   |
| Conv2(ReLu)  | 43x43x60   | 3x3           | 1        | 41x41x120  |
| Pool2  | 41x41x120  | 2x2           | 2        | 21x21x120  |
| FC1(ReLu + Dropout)                                  | 52920      |               |          | 100        |
| FC2  | 100        |               |          | 2          |
| ConvNet 07 (Inicialización de pesos: <b>Xavier</b> ) |            |               |          |            |
| Capa   | Entrada    | Tamaño filtro | Paso     | Salida     |
| Conv1(ReLu + LRN)                                    | 176x176x1  | 6x6           | 2        | 86x86x60   |
| Pool1  | 86x86x60   | 2x2           | 2        | 43x43x60   |
| Conv2(ReLu)  | 43x43x60   | 3x3           | 1        | 41x41x120  |
| Pool2  | 41x41x120  | 2x2           | 2        | 21x21x120  |
| FC1(ReLu + Dropout)                                  | 52920      |               |          | 100        |
| FC2  | 100        |               |          | 2          |

### 3.5 DATA SETS Y ENTRENAMIENTO

Para el entrenamiento de redes neuronales y modelos de clasificación en general el conjunto de datos disponibles, o data set, se subdivide en tres conjuntos: el conjunto de entrenamiento, el conjunto de validación que se usa para seleccionar el mejor modelo y el conjunto de prueba con el cual se consigue el error real del modelo de clasificación.

En el caso específico de las redes neuronales hay que tener en cuenta la cantidad de datos que se tienen por clase a clasificar ya que el resultado en general es que posteriormente la red clasificará bien datos de la clase con más muestras en el dataset de entrenamiento y erróneamente las clases con menos muestras. Esta situación se presenta con la población seleccionada para la evaluación de las redes neuronales, como se puede ver en la Tabla 1 el número de pacientes considerados como sanos (CDR = 0) es tres veces más que el número de pacientes con la enfermedad (CDR = 1) por esta razón se procedió a hacer un balance del dataset que consistió en seleccionar aleatoriamente veinte pacientes sanos para igualarlo con el número de pacientes con la enfermedad.

Tabla 3. Distribución de los datasets.

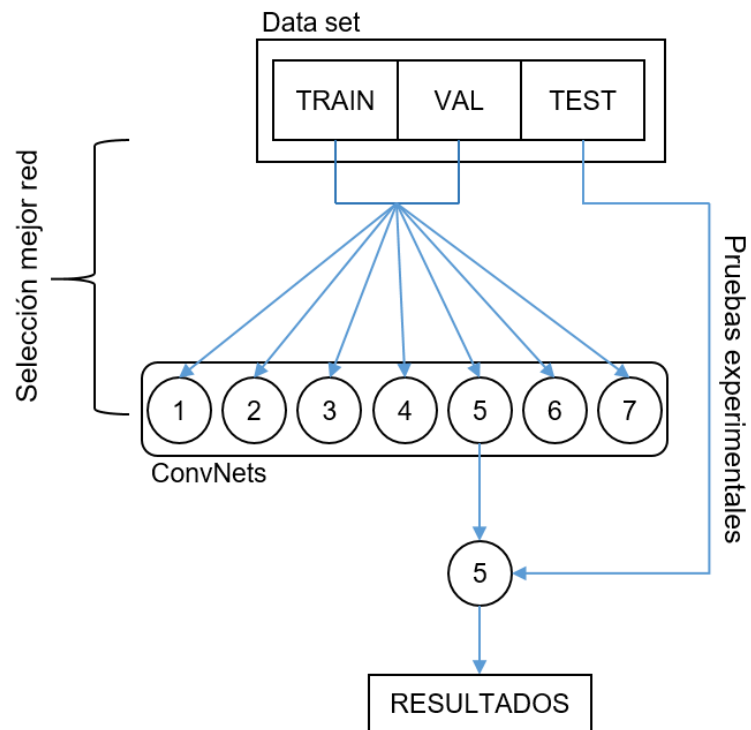
| Data set |            |                          |                       |                   |       |
|----------|------------|--------------------------|-----------------------|-------------------|-------|
| Clase    | #Pacientes | Imágenes (Entrenamiento) | Imágenes (Validación) | Imágenes (Prueba) | Total |
| CDR 0    | 66         | 11880                    | 3960                  | 3960              | 19800 |
| CDR 1    | 20         | 3600                     | 1200                  | 1200              | 6000  |

| Data set (Balanceado) |            |                          |                       |                   |       |
|-----------------------|------------|--------------------------|-----------------------|-------------------|-------|
| Clase                 | #Pacientes | Imágenes (Entrenamiento) | Imágenes (Validación) | Imágenes (Prueba) | Total |
| CDR 0                 | 20         | 3600                     | 1200                  | 1200              | 6000  |
| CDR 1                 | 20         | 3600                     | 1200                  | 1200              | 6000  |

### 3.6 EVALUACIÓN DE LAS REDES NEURONALES

Para obtener indicadores de la eficacia de las redes neuronales convolucionales se establecieron tres métodos de evaluación comúnmente usados para clasificadores o herramientas de diagnóstico.

Figura 7. Representación del proceso experimental.



**3.6.1 Selección de la mejor configuración de red.** La primera fase de evaluación consistió en seleccionar la red convolucional con mejor precisión de las siete configuraciones que se presentaron anteriormente, para esto el dataset se divide en los tres subconjuntos de entrenamiento, validación y prueba en una proporción de 60/20/20 por ciento respectivamente, este proceso de subdivisión del dataset se repite varias veces seleccionando aleatoriamente cada vez los pacientes que se van a ubicar en cada conjunto.

La precisión de las redes para la selección de la mejor configuración se toma a partir del conjunto de validación, en este caso se hizo la prueba con cinco diferentes distribuciones de pacientes en los conjuntos de entrenamiento y validación para cada una de las siete configuraciones de red, los resultados se muestran en la tabla a continuación.

Tabla 4. Resultados para la selección de la mejor red neuronal.

| DATA SET          | RED NEURONAL |            |            |            |                 |            |            |
|-------------------|--------------|------------|------------|------------|-----------------|------------|------------|
|                   | ConvNet 01   | ConvNet 02 | ConvNet 03 | ConvNet 04 | ConvNet 05      | ConvNet 06 | ConvNet 07 |
| 1                 | 61.2917%     | 65.8333%   | 64.9583%   | 66.8750%   | 64.6667%        | 64.8333%   | 64.2916%   |
| 2                 | 58.0833%     | 67.3750%   | 65.6250%   | 67.0833%   | 68.5000%        | 64.3333%   | 66.4167%   |
| 3                 | 68.8333%     | 72.9167%   | 73.7316%   | 72.7083%   | 74.0833%        | 75.8333%   | 74.7917%   |
| 4                 | 64.8333%     | 69.9583%   | 66.0416%   | 69.7500%   | 69.9167%        | 68.5000%   | 66.4583%   |
| 5                 | 71.1667%     | 77.0000%   | 74.5833%   | 77.3333%   | 77.0000%        | 77.2917%   | 74.7083%   |
| <b>Promedio</b>   | 64.8417%     | 70.6167%   | 68.9880%   | 70.7500%   | <b>70.8333%</b> | 70.1583%   | 69.3333%   |
| <b>Desviación</b> | 0.053436     | 0.044682   | 0.047444   | 0.043771   | 0.048193        | 0.060854   | 0.050218   |

A partir de esta prueba se decide utilizar la configuración de la red número cinco para las posteriores pruebas.

**3.6.2 Pruebas Experimentales.** Los indicadores que usan para medir la eficacia y el rendimiento de las redes neuronales convolucionales y para comparar con la línea de base replicada se generan a partir de las pruebas que se describen a continuación.

**3.6.2.1 Bootstrapping** En el campo del aprendizaje de máquina, el bootstrap se utiliza para mejorar la estabilidad y precisión de los clasificadores. Este proceso es similar al que se usó para la selección de la mejor configuración de red, igualmente se busca obtener la precisión en la clasificación de la red neuronal.

En este proceso la configuración de red elegida se prueba varias veces con el conjunto de prueba que a su vez se forma cada vez con una selección aleatoria de paciente. La precisión se calcula a partir de cuantos pacientes se clasifican correctamente, para considerar que un paciente está bien clasificado la red debe clasificar el cincuenta por ciento o más de las imágenes cerebrales con la misma

clase del paciente al que corresponden dichas imágenes, es decir ciento cincuenta o más imágenes.

El indicador final es el promedio de las precisiones resultantes de clasificar ocho pacientes aleatorios en cinco conjuntos de prueba.

**3.6.2.2 Validación cruzada dejando uno afuera.** En esta prueba la red neuronal se entrena utilizando las imágenes de todos los pacientes exceptuando uno del cual se clasifican todas sus imágenes como conjunto de prueba, igual que con la prueba anterior, si cincuenta por ciento o más de sus imágenes se clasifican con la correspondiente clase del paciente, se considera que el paciente se clasificó correctamente.

Con este proceso se reduce el error en la clasificación pero la red se entrena con más datos y tantas veces como datos se tenga lo cual resulta ser un trabajo tedioso y más costoso computacionalmente.

**3.6.2.3 Área bajo la curva ROC.** La curva ROC (Receiver Operating Characteristic) es una representación gráfica que relaciona la tasa de verdaderos positivos o sensibilidad, con la tasa de falsos positivos o falsas alarmas según la varianza de un umbral de clasificación. Estas curvas son ampliamente usadas para probar métodos diagnósticos en las diferentes áreas de la medicina.

El área bajo la curva (AUC) se usa para establecer que tan bueno es un método de diagnóstico al momento de discernir entre paciente enfermos y sanos. Los valores de las AUC están entre 0.5 que representaría un método incapaz de discernir entre clases y 1.0 que representaría un clasificador perfecto.

## **3.7 EQUIPO Y SOFTWARE**

**3.7.1 Hardware.** Este proyecto de investigación se realizó enteramente en único equipo de cómputo del cual se listan aquí las características más relevantes.

- Procesador AMD FX®-8350 de ocho núcleos a 4.00 GHz.
- Tarjeta gráfica Nvidia GeForce® GTX 760 GDDR5 con 2.0 GB de memoria.
- Dos módulos de memoria RAM DDR3 de 8 GB cada uno.
- Sistema operativo de 64 bits.

**3.7.2 Software.** Todo el proceso de diseño, entrenamiento y pruebas de las redes neuronales se llevó a cabo con la herramienta DIGITS (Deep Learning GPU Training System) en su versión 3.0, desarrollada por Nvidia. Este sistema facilita el diseño y entrenamiento de redes neuronales basadas en CAFFE o TORCH, permite también visualizar la topología de las redes y, en tiempo real, el rendimiento de las mismas durante el entrenamiento. Todo con soporte de las unidades de procesamiento gráfico GPU.

Este sistema se ejecuta sobre el sistema operativo Ubuntu 14.04 basado en GNU/Linux.

## 4. RESULTADOS Y ANALISIS

En este capítulo se presentan y analizan los resultados de los tres métodos de evaluación mencionados en el capítulo anterior mediante tablas y gráficas.

### 4.1 BOOTSTRAPPING

La tabla muestra para cada paciente del dataset de prueba su CDR real y el puntaje asignado por el clasificador para las dos clases, en el caso de la línea de base el puntaje para cada clase es resultado de la función de pertenencia generada a partir de la información neuropsicológica, para las redes neuronales el puntaje es el porcentaje de las imágenes de cada clase.

Tabla 5. Resultados de la prueba bootstapping.

| BOOTSTRAPPING (40 PACIENTES) |          |            |        |               |         |               |         |
|------------------------------|----------|------------|--------|---------------|---------|---------------|---------|
| Data set 1                   |          | Línea base |        | ConvNet CAFFE |         | ConvNet TORCH |         |
| ID Paciente                  | CDR Real | CDR 0      | CDR 1  | CDR 0         | CDR 1   | CDR 0         | CDR 1   |
| OAS1_0070_MR1                | 0        | 0.9050     | 0.1000 | 86.667%       | 13.333% | 98.333%       | 1.667%  |
| OAS1_0199_MR1                | 0        | 0.6738     | 0.3353 | 71.667%       | 28.333% | 77.667%       | 22.333% |
| OAS1_0212_MR1                | 0        | 0.2249     | 0.7632 | 45.667%       | 54.333% | 36.000%       | 64.000% |
| OAS1_0259_MR1                | 0        | 0.8541     | 0.1535 | 86.667%       | 13.333% | 84.667%       | 15.333% |
| OAS1_0268_MR1                | 1        | 0.5764     | 0.4211 | 34.667%       | 65.333% | 24.667%       | 75.333% |
| OAS1_0269_MR1                | 1        | 0.5816     | 0.4218 | 35.000%       | 65.000% | 41.000%       | 59.000% |
| OAS1_0425_MR1                | 1        | 0.1675     | 0.8067 | 15.000%       | 85.000% | 13.333%       | 86.667% |
| OAS1_0430_MR1                | 1        | 0.2020     | 0.8045 | 8.000%        | 92.000% | 9.000%        | 91.000% |
| Data set 2                   |          | Línea base |        | ConvNet CAFFE |         | ConvNet TORCH |         |
| ID Paciente                  | CDR Real | CDR 0      | CDR 1  | CDR 0         | CDR 1   | CDR 0         | CDR 1   |
| OAS1_0109_MR1                | 0        | 0.9717     | 0.0286 | 99.667%       | 0.333%  | 99.667%       | 0.333%  |
| OAS1_0199_MR1                | 0        | 0.7304     | 0.2778 | 83.667%       | 16.333% | 74.333%       | 25.667% |
| OAS1_0212_MR1                | 0        | 0.2090     | 0.7790 | 40.667%       | 59.333% | 49.333%       | 50.667% |
| OAS1_0216_MR1                | 0        | 0.9187     | 0.0785 | 86.333%       | 13.667% | 84.333%       | 15.667% |
| OAS1_0184_MR1                | 1        | 0.3103     | 0.6975 | 30.333%       | 69.667% | 34.667%       | 65.333% |
| OAS1_0268_MR1                | 1        | 0.5793     | 0.4175 | 35.333%       | 64.667% | 25.333%       | 74.667% |
| OAS1_0405_MR1                | 1        | 0.0764     | 0.8908 | 36.333%       | 63.667% | 25.000%       | 75.000% |
| OAS1_0430_MR1                | 1        | 0.1885     | 0.8181 | 17.667%       | 82.333% | 19.333%       | 80.667% |

| <b>BOOTSTRAPPING (40 PACIENTES)</b> |          |            |        |               |         |               |         |
|-------------------------------------|----------|------------|--------|---------------|---------|---------------|---------|
| Data set 3                          |          | Línea base |        | ConvNet CAFFE |         | ConvNet TORCH |         |
| ID Paciente                         | CDR Real | CDR 0      | CDR 1  | CDR 0         | CDR 1   | CDR 0         | CDR 1   |
| OAS1_0114_MR1                       | 0        | 0.9258     | 0.0838 | 85.667%       | 14.333% | 92.667%       | 7.333%  |
| OAS1_0259_MR1                       | 0        | 0.8251     | 0.1831 | 81.000%       | 19.000% | 89.000%       | 11.000% |
| OAS1_0280_MR1                       | 0        | 0.1804     | 0.8363 | 24.000%       | 76.000% | 21.000%       | 79.000% |
| OAS1_0358_MR1                       | 0        | 0.8196     | 0.1647 | 76.000%       | 24.000% | 76.000%       | 24.000% |
| OAS1_0073_MR1                       | 1        | 0.1858     | 0.8483 | 13.000%       | 87.000% | 20.000%       | 80.000% |
| OAS1_0134_MR1                       | 1        | -0.1371    | 1.1282 | 30.000%       | 70.000% | 14.667%       | 85.333% |
| OAS1_0269_MR1                       | 1        | 0.5485     | 0.4567 | 46.333%       | 53.667% | 43.333%       | 56.667% |
| OAS1_0291_MR1                       | 1        | 0.7851     | 0.2329 | 85.000%       | 15.000% | 89.333%       | 10.667% |
| Data set 4                          |          | Línea base |        | ConvNet CAFFE |         | ConvNet TORCH |         |
| ID Paciente                         | CDR Real | CDR 0      | CDR 1  | CDR 0         | CDR 1   | CDR 0         | CDR 1   |
| OAS1_0010_MR1                       | 0        | 0.2403     | 0.7671 | 14.667%       | 85.333% | 14.667%       | 85.333% |
| OAS1_0135_MR1                       | 0        | 0.6545     | 0.3366 | 87.667%       | 12.333% | 80.000%       | 20.000% |
| OAS1_0199_MR1                       | 0        | 0.6147     | 0.3931 | 76.000%       | 24.000% | 69.333%       | 30.667% |
| OAS1_0358_MR1                       | 0        | 0.7859     | 0.1970 | 88.333%       | 11.667% | 76.667%       | 23.333% |
| OAS1_0067_MR1                       | 1        | 0.4942     | 0.5210 | 40.333%       | 59.667% | 30.667%       | 69.333% |
| OAS1_0134_MR1                       | 1        | -0.2032    | 1.1868 | 6.667%        | 93.333% | 6.333%        | 93.667% |
| OAS1_0185_MR1                       | 1        | 0.4742     | 0.4985 | 27.333%       | 72.667% | 29.667%       | 70.333% |
| OAS1_0425_MR1                       | 1        | 0.2760     | 0.6963 | 21.333%       | 78.667% | 13.333%       | 86.667% |
| Data set 5                          |          | Línea base |        | ConvNet CAFFE |         | ConvNet TORCH |         |
| ID Paciente                         | CDR Real | CDR 0      | CDR 1  | CDR 0         | CDR 1   | CDR 0         | CDR 1   |
| OAS1_0010_MR1                       | 0        | 0.2273     | 0.7797 | 23.667%       | 76.333% | 19.000%       | 81.000% |
| OAS1_0030_MR1                       | 0        | 0.6987     | 0.2951 | 82.333%       | 17.667% | 76.000%       | 24.000% |
| OAS1_0212_MR1                       | 0        | 0.1122     | 0.8737 | 49.667%       | 50.333% | 39.000%       | 61.000% |
| OAS1_0341_MR1                       | 0        | 0.9343     | 0.0686 | 94.667%       | 5.333%  | 90.333%       | 9.667%  |
| OAS1_0134_MR1                       | 1        | -0.2259    | 1.2097 | 16.333%       | 83.667% | 10.333%       | 89.667% |
| OAS1_0316_MR1                       | 1        | 0.0907     | 0.9121 | 24.333%       | 75.667% | 17.667%       | 82.333% |
| OAS1_0399_MR1                       | 1        | 0.1302     | 0.8419 | 10.000%       | 90.000% | 6.667%        | 93.333% |
| OAS1_0405_MR1                       | 1        | 0.0335     | 0.9314 | 35.333%       | 64.667% | 21.000%       | 79.000% |

De la tabla anterior se calcula la precisión estableciendo un umbral de 0.5 para considerar un paciente de una clase u otra en el caso de la línea de base o 50% para las redes neuronales. Esto se muestra en la Tabla 9 más adelante en este informe.

## 4.2 VALIDACIÓN CRUZADA DEJANDO UNO FUERA

Esta prueba se repite dos veces para cada uno de los tres métodos de clasificación, una con la población completa de ochenta y seis pacientes y otra con el dataset balanceado de cuarenta pacientes esto sirve para ver como el desbalanceo de muestras en cada clase afecta el desempeño de los clasificadores.

Tabla 6. Resultados de la prueba de validación cruzada para la línea de base.

| VALIDACIÓN CRUZADA LÍNEA DE BASE |    |          |                      |    |          |
|----------------------------------|----|----------|----------------------|----|----------|
| 40 pacientes                     |    |          | 86 pacientes         |    |          |
| Clasificación global             |    |          | Clasificación global |    |          |
| Correctos                        | 28 | 70.0000% | Correctos            | 67 | 77.9070% |
| Incorrectos                      | 12 | 30.0000% | Incorrectos          | 19 | 22.0930% |

Tabla 7. Resultados de la prueba de validación cruzada para las redes neuronales convolucionales.

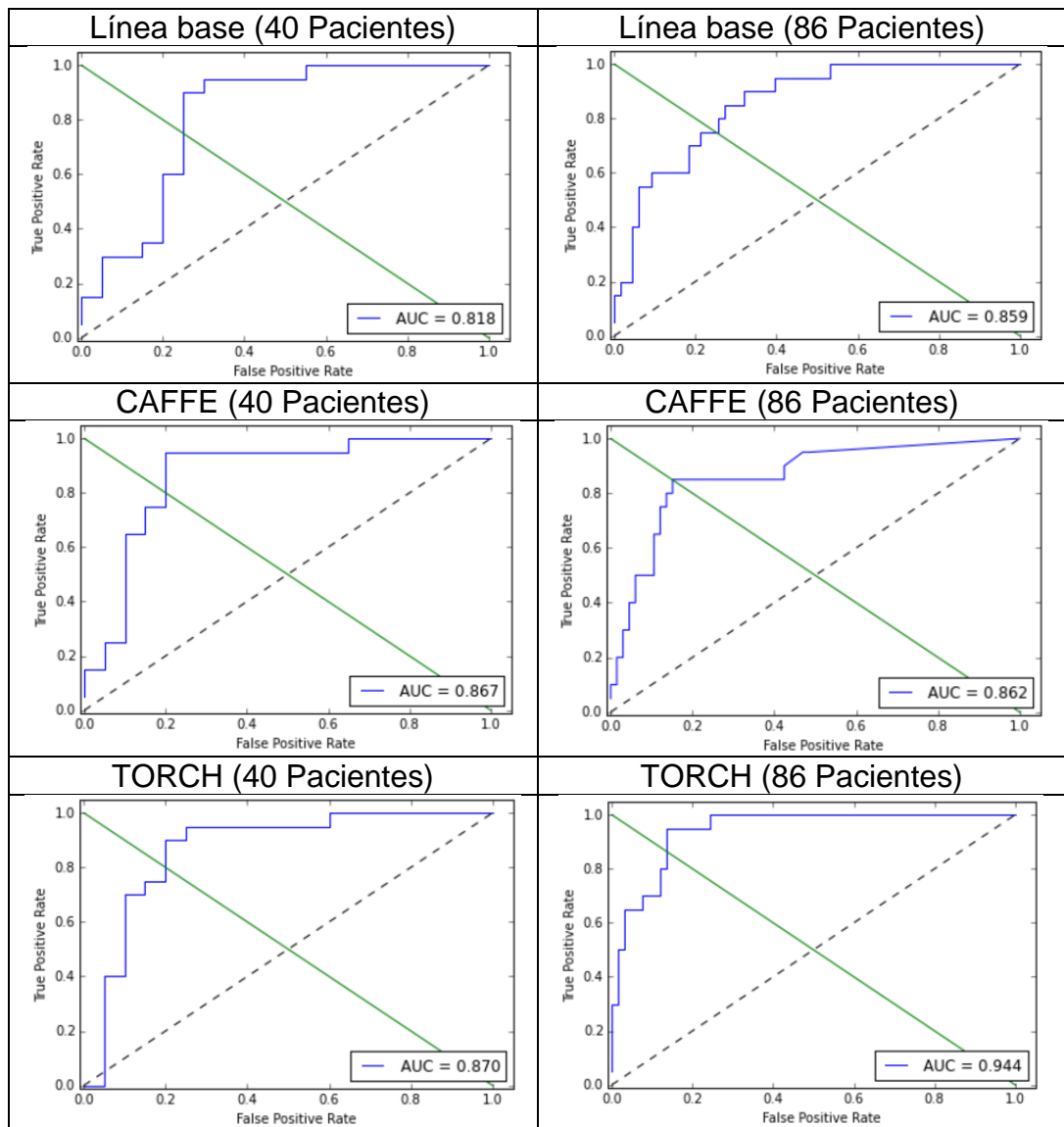
| VALIDACIÓN CRUZADA REDES NEURONALES CONVOLUCIONALES |    |        |                         |    |         |
|---|----|--------|-------------------------|----|---------|
| CAFFE 40 Pacientes                                  |    |        | CAFFE 86 Pacientes      |    |         |
| Clasificación global                                |    |        | Clasificación global    |    |         |
| Correctos   | 32 | 80.00% | Correctos               | 68 | 79.07%  |
| Incorrectos   | 8  | 20.00% | Incorrectos             | 18 | 20.93%  |
| Clasificación por clase                             |    |        | Clasificación por clase |    |         |
| CDR0  |    |        | CDR0                    |    |         |
| Correctos   | 16 | 80.00% | Correctos               | 66 | 100.00% |
| Incorrectos   | 4  | 20.00% | Incorrectos             | 0  | 0.00%   |
| CDR1  |    |        | CDR1                    |    |         |
| Correctos   | 16 | 80.00% | Correctos               | 2  | 10.00%  |
| Incorrectos   | 4  | 20.00% | Incorrectos             | 18 | 90.00%  |
| TORCH 40 Pacientes                                  |    |        | TORCH 86 Pacientes      |    |         |
| Clasificación global                                |    |        | Clasificación global    |    |         |
| Correctos   | 34 | 85.00% | Correctos               | 68 | 79.07%  |
| Incorrectos   | 6  | 15.00% | Incorrectos             | 18 | 20.93%  |
| Clasificación por clase                             |    |        | Clasificación por clase |    |         |
| CDR0  |    |        | CDR0                    |    |         |
| Correctos   | 16 | 80.00% | Correctos               | 66 | 100.00% |
| Incorrectos   | 4  | 20.00% | Incorrectos             | 0  | 0.00%   |
| CDR1  |    |        | CDR1                    |    |         |
| Correctos   | 18 | 90.00% | Correctos               | 2  | 10.00%  |
| Incorrectos   | 2  | 10.00% | Incorrectos             | 18 | 90.00%  |

Se puede ver en la tabla anterior como un dataset desbalanceado afecta el desempeño de las redes neuronales convolucionales disminuyendo el porcentaje en la clasificación global y la capacidad de diferenciar entre clases, esto último se ve claramente en la clasificación por clase, para los ochenta y seis pacientes las redes neuronales clasifican correctamente todos los pacientes con CDR 0 e incorrectamente la mayoría de los pacientes con CDR 1.

### 4.3 ÁREA BAJO LA CURVA ROC

Los resultados de las pruebas anteriores se usan para calcular el área bajo la curva ROC la cual es otro indicador para determinar la capacidad de los clasificadores para discernir entre clases.

Figura 8. Curvas ROC para cada prueba.



Para las pruebas de bootstrapping se calcula el área bajo la curva en la clasificación de cada data set diferente y se promedian para cada clasificador.

#### 4.4 COMPARACIÓN EN RENDIMIENTO COMPUTACIONAL

Además de la precisión o el área bajo la curva ROC se tienen en cuenta los recursos computacionales usados en el entrenamiento de las redes neuronales en CAFFE y TORCH. Con DIGITS se puede monitorear el consumo de memoria gráfica en tiempo real en la etapa de entrenamiento, para ambos casos el total de la memoria gráfica que dispone el equipo.

La siguiente tabla comparativa muestra otros indicadores de rendimiento para los dos frameworks con data sets y configuraciones de red neuronal iguales.

Tabla 8. Tabla comparativa de rendimiento computacional.

| CAFFE                   |             |             |             |             |             |          |            |
|-------------------------|-------------|-------------|-------------|-------------|-------------|----------|------------|
| Recurso                 | Data set 01 | Data set 02 | Data set 03 | Data set 04 | Data set 05 | Promedio | Desviación |
| Memoria gráfica (2 GB)  | ~100%       | ~100%       | ~100%       | ~100%       | ~100%       | ~100%    | 0.000000   |
| Tiempo de entrenamiento | 00:19:30    | 00:19:30    | 00:19:28    | 00:19:28    | 00:19:27    | 00:19:29 | 0.000016   |
| Tamaño en disco (GB)    | 1.33        | 1.33        | 1.33        | 1.33        | 1.33        | 1.33     | 0.000000   |
| TORCH                   |             |             |             |             |             |          |            |
| Recurso                 | Data set 01 | Data set 02 | Data set 03 | Data set 04 | Data set 05 | Promedio | Desviación |
| Memoria gráfica (2 GB)  | ~100%       | ~100%       | ~100%       | ~100%       | ~100%       | ~100%    | 0.000000   |
| Tiempo de entrenamiento | 00:15:22    | 00:15:23    | 00:15:25    | 00:15:26    | 00:15:26    | 00:15:24 | 0.000021   |
| Tamaño en disco (GB)    | 1.19        | 1.19        | 1.19        | 1.19        | 1.19        | 1.19     | 0.000000   |

#### 4.5 RESULTADOS FINALES

De toda la experimentación realizada se obtienen indicadores de comparación para cada método de clasificación y de los cuales se pueden sacar las conclusiones, en la siguiente tabla comparativa se muestra el área bajo la curva, el ratio de verdaderos positivos (TPP), el ratio de falsos positivos (FPR) y la precisión de cada clasificador.

Tabla 9. Tabla comparativa de métodos de clasificación.

| PRUEBA                            | INDICADOR  | LÍNEA BASE | CAFFE   | TORCH   |
|-----------------------------------|------------|------------|---------|---------|
| Bootstrap (40 Pacientes)          | Precisión  | 72.500%    | 82.500% | 82.500% |
|                                   | Desviación | 0.0935     | 0.0612  | 0.0612  |
|                                   | AUC        | 0.8750     | 0.8750  | 0.9000  |
|                                   | Desviación | 0.0395     | 0.1186  | 0.0848  |
| Validación cruzada (40 Pacientes) | Precisión  | 70.000%    | 80.000% | 85.000% |
|                                   | AUC        | 0.8180     | 0.8600  | 0.8700  |
|                                   | TPR        | 0.6500     | 0.8000  | 0.9000  |
|                                   | FPR        | 0.2500     | 0.2000  | 0.2000  |
| Validación cruzada (86 Pacientes) | Precisión  | 77.907%    | 79.070% | 79.070% |
|                                   | AUC        | 0.8590     | 0.8620  | 0.9440  |
|                                   | TPR        | 0.6500     | 0.1000  | 0.1000  |
|                                   | FPR        | 0.1818     | 0.0000  | 0.0000  |

A partir de la tabla comparativa ya es posible constatar directamente las diferencias entre cada método de clasificación.

En la prueba de bootstrapping se ve una mejora significativa en la precisión por parte de las redes neuronales convolucionales respecto al clasificador de la línea de base. Ya que la precisión entre las dos redes neuronales es igual con los diferentes frameworks hay que fijarse en el AUC para diferenciar el desempeño entre esta dos, un AUC mayor indica, en este caso, que se clasifica un mayor número de imágenes de una clase respecto a la otra.

Con las pruebas de validación cruzada dejando uno afuera se puede ver la importancia de tener un conjunto de datos balanceado. Para la prueba con cuarenta pacientes, veinte por clase, la precisión es significativamente mejor para la red neuronal en CAFFE y aún más en TORCH con respecto a la línea de base teniendo en cuenta que un paciente es el 2,5% del total. En cambio, con el conjunto de datos desbalanceado de ochenta y seis pacientes la precisión en la línea de base es mejor respecto a la anterior con cuarenta pacientes y la diferencia porcentual corresponde a un solo paciente mal clasificado respecto a la precisión en las redes neuronales convolucionales. El TPR deja en evidencia el inconveniente de clasificadores como las redes neuronales convolucionales, el valor de 0.1 para ambas redes indica que de los veinte pacientes de la clase con la enfermedad la red sólo pudo identificar efectivamente dos como pacientes enfermos. Por lo tanto, no importa si la precisión es mejor para las redes si estas no son capaces de diferenciar alguna de las clases.

De lo anterior se puede afirmar que el método de clasificación replicado en la línea de base tiene la ventaja de no depender del balanceo de las muestras por clase y que esto no tiene gran relevancia al momento de aumentar la población.

## 5. CONCLUSIONES

Con los resultados obtenidos durante todo el trabajo experimental y teniendo en cuenta los objetivos planteados al inicio se listan a continuación las conclusiones a las que se llegaron.

El pre procesamiento de la información es importante primero para posibilitar su procesamiento por parte los diferentes métodos de clasificación y segundo para obtener mejor desempeño de los mismos.

El diseño de redes neuronales convolucionales, a pesar de toda su base teórica resultado de años de estudio, tienen un componente de prueba y error y dependen en gran medida del problema que se esté abordando.

El desempeño de clasificadores basados en redes neuronales convolucionales va más allá de su arquitectura y se cree que las diferencias observadas entre las dos herramientas evaluadas en este trabajo se originan desde su diseño de framework en aspectos como el lenguaje en el que estén programados y los algoritmos de aprendizaje que usen.

Los resultados obtenidos muestran un leve mejoramiento en el trabajo de clasificar o detectar la enfermedad del Alzheimer a partir de imágenes de resonancia magnética cerebral respecto al método replicado como línea de base. Cabe mencionar, sin embargo, que el costo computacional de las redes neuronales es muy grande en comparación ya que el trabajo replicado se basa sólo en cálculos con matrices.

Ya que se comprobó que las redes neuronales convolucionales funcionan mejor con data set balanceados es de esperar que según los datasets crezcan las redes funciones mejor todavía.

Resulta innovadora la aplicación de las redes neuronales convolucionales con data sets especializados como este con imágenes médicas ya que normalmente se usan junto a data sets generales, por ejemplo, en la detección de objetos.

Se espera que con este trabajo se hayan sentado las bases para una posterior investigación que abarque más métodos de pre procesamiento de las imágenes, más formas de entrenamiento de redes neuronales y otros algoritmos de aprendizaje.

## BIBLIOGRAFÍA

BASOGAIN, Xabier. Redes Neuronales Artificiales y sus Aplicaciones. Bilbao: Dpto. Ingeniería de Sistemas y Automática Escuela Superior de Ingeniería de Bilbao. 2014. 79p.

BRETT, Matthew; HANKE, Michael and LARSON, Eric. NiBabel Manual. 2015. <http://nipy.org/nibabel/manual.html> [Consulta: Lunes, 29 de febrero de 2016].

DETTMERS, Tim. Deep Learning in a Nutshell: Core Concepts. 2015. <https://devblogs.nvidia.com/parallelforall/deep-learning-nutshell-core-concepts> [Consulta: Jueves, 11 de febrero de 2016]

GIRALDO, Diana; GARCÍA A., Juan and ROMERO, Eduardo. Finding Models of the Alzheimer Disease by Fusing Information from Neuropsychological Tests and Structural MRI Images. {Online}. {March 24, 2016}. Available in: {<http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=2507239>}

GRAY, Allison. DIGITS: Deep Learning GPU Training System. 2015. <https://devblogs.nvidia.com/parallelforall/digits-deep-learning-gpu-training-system> [Consulta: Martes, 8 de marzo de 2015]

LLOPIS, Jaume. Tema 23: ANÁLISIS ROC. 2013. <https://estadisticaorquestainstrumento.wordpress.com/2013/02/13/tema-23-analisis-roc> [Consulta: Lunes, 11 de enero de 2016]

MARCUS, Daniel S., et al. OASIS Fact Sheet. {Online}. {November 26, 2016} Available in: {[http://www.oasis-brains.org/pdf/oasis\\_cross-sectional\\_facts.pdf](http://www.oasis-brains.org/pdf/oasis_cross-sectional_facts.pdf)}

ZHENG, Alices. How to Evaluate Machine Learning Models: Classification Metrics. 2015. <http://blog.dato.com/how-to-evaluate-machine-learning-models-part-2a-classification-metrics> [Consulta: Lunes, 11 de enero de 2016]