 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-27

# **Un modelo de aprendizaje profundo para mejorar la calidad de imágenes médicas de resonancia magnética de la mama**

Pablo Campaz Usuga

Ingeniería Electrónica

Carlos Andrés Mera Banguero

Rubén Darío Fonnegra Tarazona

**INSTITUTO TECNOLÓGICO METROPOLITANO**

**2021**

	<p style="text-align: center;">INFORME FINAL DE TRABAJO DE GRADO</p>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

## RESUMEN

---

Las redes de aprendizaje profundo (*Deep Learning*) basadas en redes neuronales artificiales son técnicas de aprendizaje de máquinas que han tomado notable atención por parte de la comunidad académica y la industria en recientes años. Específicamente, esto se debe al buen desempeño que han obtenido los modelos de aprendizaje profundo en diversas aplicaciones que abarcan el procesamiento de señales (como fisiológicas o de audio) o de imágenes (como las resonancias magnéticas). El interés en esta área ha llevado a que el aprendizaje profundo sea uno de los enfoques de estudio en diversos proyectos de investigación, algunos de los cuales se enfocan en el mejoramiento de la calidad de las imágenes médicas. En este trabajo se propone el uso de una estrategia basada en técnicas de aprendizaje profundo para mejorar la calidad de imágenes médicas obtenidas por resonancia magnética. Específicamente, en el trabajo se ha utilizado una base de datos de imágenes denominada *Qin\_breast*. Como resultado, este trabajo contribuye al desarrollo de herramientas que pueden facilitar al personal médico el análisis e interpretación de imágenes de resonancia magnética de la mama que pueden ayudar a hacer diagnósticos más precisos.

*Palabras clave:* Aprendizaje profundo, neuronas artificiales, resonancia magnética de la mama.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

## RECONOCIMIENTOS

---

A mi madre Gleyden Usuga Castaño y padre Victor Campaz Naranjo por el apoyo y la paciencia brindada en todos estos semestres cursados.

A mis asesores Ruben Dario Fonnegra y Carlos Andrés Mera Banguero por el conocimiento, acompañamiento y asesorías brindadas en el desarrollo de este proyecto.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

## ACRÓNIMOS

---

*OpenCV* - Open Computer Vision

GPU - Graphics Processing Units

*GD* – *Algoritmo del descenso de los gradientes*

*MSE* – *Mean Square Error*

*GAN* – Generative Adversarial Networks

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

## TABLA DE CONTENIDO

1.	INTRODUCCIÓN .....	6
2.	MARCO TEÓRICO .....	8
2.1.	APRENDIZAJE PROFUNDO O <i>DEEP LEARNING</i> .....	8
	RED NEURONAL PERCEPTRON .....	9
	PERCEPTRÓN MULTICAPA .....	11
	REDES NEURONALES CONVOLUCIONALES .....	12
2.2.	APRENDIZAJE PROFUNDO O <i>DEEP LEARNING</i> .....	14
	EL USO DE LOS AUTOENCODER .....	15
3.	METODOLOGÍA.....	17
	BASE DE DATOS QIN-BREAST .....	18
3.1.	PREPROCESAMIENTO .....	19
3.2.	MODELO DEEP LEARNING .....	20
3.3.	MÉTRICAS DE DESEMPEÑO.....	22
3.3.1.	Error cuadrático medio (MSE) .....	22
3.3.2.	Relación pico de ruido de la señal (PSNR) .....	23
3.3.3.	Medida de similitud estructural (SSIM).....	23
4	RESULTADOS Y DISCUSIÓN.....	24
5	CONCLUSIONES, RECOMENDACIONES Y TRABAJO FUTURO .....	29
	REFERENCIAS .....	30
	APÉNDICE.....	32

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

# 1. INTRODUCCIÓN

---

En el ámbito de la medicina, el resonador magnético es un equipo especializado de gran utilidad que permite generar imágenes del cuerpo y de sus estructuras internas por medio de la inducción de un campo electromagnético. La interacción del campo magnético con el cuerpo se da por medio de pulsos de radiofrecuencia, los cuales atraviesan los tejidos del cuerpo captando una imagen (resonancia magnética) que se genera a partir de los cambios de intensidades en los píxeles. Este dispositivo es comúnmente utilizado para mostrar distintas estructuras de tejido microscópicas, en las cuales pueden identificarse algunas patologías que se manifiestan a través de cambios en la composición del tejido. El cáncer de mama es una de esas patologías, la cual es además considerada un problema a nivel mundial ya que se sitúa como la primera enfermedad en producir mayor número de incidencias por año en el mundo. Uno de los parámetros críticos que influye directamente en la calidad de las imágenes obtenidas mediante resonancia magnética es la potencia del resonador. Esto debido a que una mayor potencia permite generar un campo electromagnético mayor; el cual a su vez permitiría visualizar con más facilidad algunas estructuras microscópicas en el tejido. Comercialmente, existen resonadores con distintas potencias; sin embargo, aquellos con mayor costo-beneficio para el correcto diagnóstico de cáncer de mama son los de 1.5T (Tesla). Sin embargo, la diferencia principal entre los resonadores que ofrecen mayor calidad de imagen (3T y 7T) respecto al anterior es el elevado costo de adquisición y operación; por lo que las entidades de salud optan por usar los primeros para el diagnóstico.

Algunas de las alternativas en el estado del arte proponen el uso de métodos basados en inteligencia artificial para mejorar la calidad de las imágenes médicas, entre las que se incluye la resonancia magnética [14]. Actualmente, estudios de investigación han demostrado que los usos de técnicas de aprendizaje de máquina han logrado generar imágenes sintéticamente con una calidad similar a aquellas de 7T, a partir de imágenes de

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

3T [14]. Mejorar la calidad de las imágenes de un estudio médico puede ayudar a los profesionales de la salud a establecer un diagnóstico más preciso del paciente sin incurrir en gastos adicionales [14]. Sin embargo, existen algunas limitantes relacionadas con la variabilidad biológica del comportamiento de tejido de los pacientes sospechosos de cáncer de mama. Por este motivo, este aún sigue siendo un tema de investigación vigente en la actualidad.

A partir de lo anterior, en este trabajo se desarrolla un modelo de aprendizaje profundo que permite mejorar la calidad de las imágenes de resonancia magnética de mama. Inicialmente, por medio de una base de datos de imágenes de 3T se degradan las imágenes introduciendo ruido o aplicando transformaciones espaciales. Seguido, se reconstruye la imagen original a partir de las imágenes con menor resolución, minimizando la diferencia entre las imágenes reales y las imágenes generadas. Los resultados de este modelo se reportan utilizando métricas computacionales de similitud.

## OBJETIVOS

### General

Desarrollar un modelo de aprendizaje profundo (*Deep Learning*) para el mejoramiento de la calidad de las imágenes médicas de resonancia magnética de mama.

### Específicos

- Preparar un conjunto de datos de entrenamiento y prueba para el diseño experimental

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

- Implementar un modelo de aprendizaje profundo para el mejoramiento de las imágenes médicas de resonancia
- Diseñar una estrategia de evaluación comparativa de las imágenes reales y las generadas por el modelo

## 2. MARCO TEÓRICO

---

### 2.1. APRENDIZAJE PROFUNDO O *DEEP LEARNING*

El aprendizaje profundo (*Deep Learning*), hace referencia al conjunto de algoritmos de aprendizaje automático encargados de modelar abstracciones de altas cantidades de datos usando arquitecturas que están compuestas por transformaciones no lineales. Esta clase de algoritmos de aprendizaje automático se han aplicado en campos de investigación como reconocimiento de voz, reconocimiento biométrico, visión por computador, entre otros, que han logrado resultados exitosos en varios campos de investigación. El caso más relevante ocurrió en 2012, en el concurso de ImageNet [6] en el cual los autores entrenaron una red neuronal convolucional profunda para clasificar los 1,2 millones de imágenes de alta resolución en el concurso ImageNet LSVRC-2010 [15]. Cada una de las imágenes debía ser clasificada en 1000 categorías diferentes. En los datos de prueba, lograron las tasas de error de top-1 y top-5 de 37.5% y 17.0%, lo cual fue considerablemente mejor que el estado del arte.

Los modelos de aprendizaje en *Deep Learning* están basados en redes neuronales que permiten aprender a través de múltiples capas de procesamiento las diferentes representaciones a partir de datos con múltiples niveles de abstracción. Este tipo de modelos computacionales se habían dejado de utilizar durante muchos años debido al desarrollo de nuevas y eficientes técnicas de aprendizaje de máquina como los *Autoencoders* y por el elevado requerimiento computacional empleado para la realización

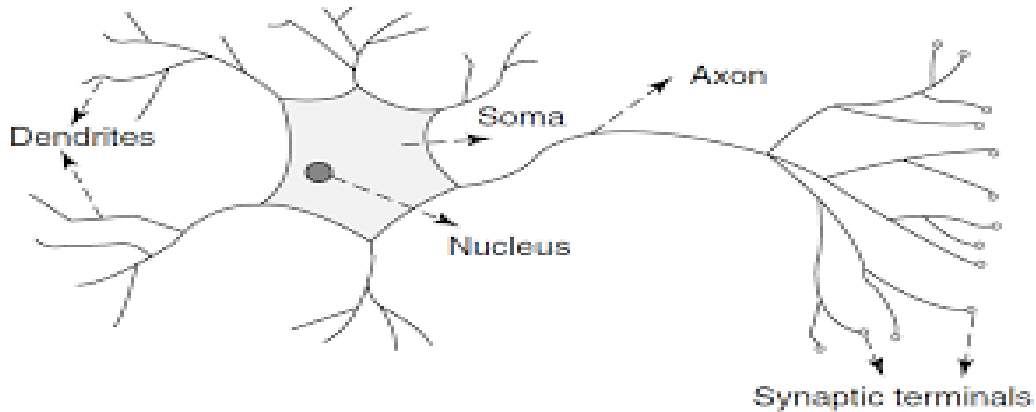


	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

de los cálculos matemáticos para su ajuste de parámetros. No obstante, recientemente, han sido retomados durante los últimos años debido al elevado crecimiento de los procesadores dedicados (como las GPU) las cuales permiten optimizar a través de programación paralela el cálculo de los parámetros para el ajuste de los modelos. Además, algunos de estos han sido implementados para mejorar considerablemente el estado de la técnica en reconocimiento de voz, reconocimiento de objetos visuales, detección de objetos entre otros.

### RED NEURONAL PERCEPTRON

Dado que los modelos de *Deep Learning* están basados, por lo general en arquitecturas con capas de redes neuronales, estas poseen una estructura compleja que permite identificar en grandes conjuntos de datos diversos patrones representativos de la señal de entrada a distintos niveles mediante las capas, permitiendo realizar tareas de clasificación en los cuales, en algunos casos, no se requiere de implementación de técnicas de preprocesamiento o extracción de características. Las arquitecturas de las redes neuronales artificiales trabajan como un modelo de aprendizaje computacional desarrollado con el fin de que las máquinas imiten el comportamiento cerebral producido en la actividad cognitiva de los humanos. En este sentido, los investigadores han propuesto imitar computacionalmente el comportamiento de las células principales encargadas de la función cognitiva humana: las neuronas. En la figura 1 se muestra una neurona biológica en la cual las señales provenientes de otras neuronas son ponderadas a través de las dendritas, antes de llegar al núcleo. El efecto de la neurona es la suma de las entradas ponderadas y si el efecto supera un umbral, la neurona es activada; por el contrario, no se activa.



**Figura 1.** Neurona biológica y sus partes. Fuente: [4,5]

Con el fin de recrear el comportamiento de una neurona biológica en modelos computacionales se propuso la estrategia mostrada en la Figura 2; a la cual se le denominó perceptrón simple. La arquitectura del perceptrón simple, está conformada de la siguiente manera:

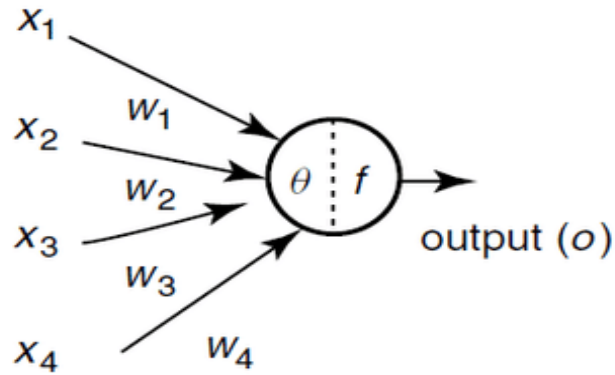
$x$ : es el vector de entrada de la neurona.

$w$ : Pesos relativos para cada entrada (ponderación de las entradas)

$output (o)$ : salida de la neurona o perceptrón simple.

La unión sumadora y la función de activación en la figura corresponden a las reglas de aprendizaje, en las que se propone que todos los valores en la sumatorias menores a cero les correspondan una salida negativa y a todos los valores en la sumatoria mayores o iguales a cero les corresponda una salida positiva.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22



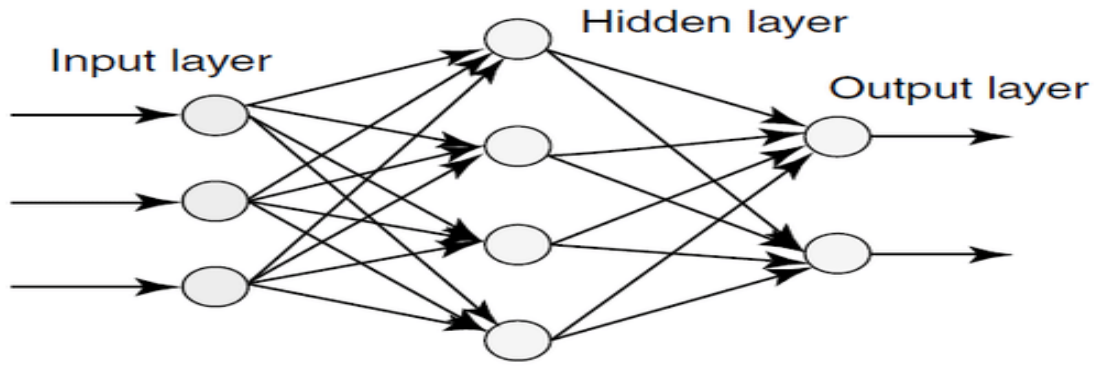
**Figura 2.** Arquitectura del perceptrón simple. Fuente: [4]

## PERCEPTRÓN MULTICAPA

En busca de crear modelos de computación que sean similares al cerebro humano y que imitan la función cognitiva, se introduce la utilización de complejas redes de perceptrones con diversas conexiones y número de entradas. Estos grupos de redes de perceptrones se denominaron perceptrones multicapa. El perceptrón multicapa se caracteriza porque tiene diversos grupos de neuronas agrupadas en varias capas a distintos niveles de profundidad. Las capas en las cuales se conectan las entradas y se obtiene la salida se denominan capas de entrada y de salida; así mismo, a las capas intermedias se les llama capas ocultas. La capa de entrada se encarga de recibir las señales y propagar dichas señales a las capas ocultas que realizan procesamientos no lineales de los patrones recibidos, por último, la capa de salida actúa como salida de la red garantizando al exterior la respuesta de la red a cada uno de los patrones de entrada.

El desarrollo de este tipo de perceptrón llevó al desarrollo de un algoritmo de optimización basado en el descenso del gradiente (GD), el cual consiste en minimizar la función de costo basada en el error obtenido por la red a partir del cálculo de los gradientes de los valores

paramétricos de la red (pesos  $w$  y sesgos  $b$ ). Además, usa un valor de control de descenso en el algoritmo llamado tasa de aprendizaje.

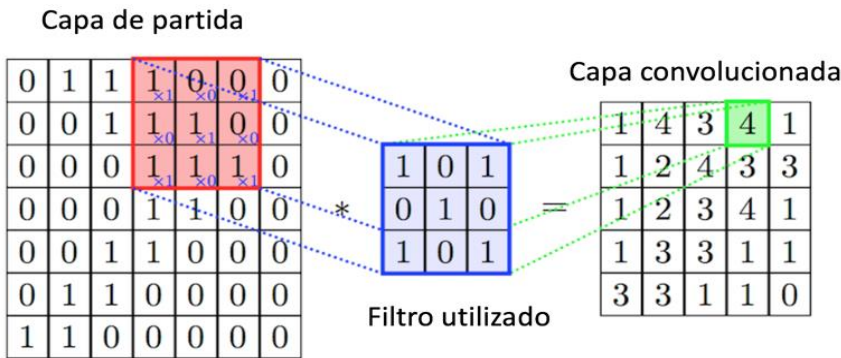


**Figura 3.** Estructura gráfica Perceptrón Multicapa. Fuente: [4]

## REDES NEURONALES CONVOLUCIONALES

Las redes neuronales convolucionales también se componen por neuronas que tienen pesos y sesgos que modifican su valor mediante el proceso de aprendizaje de la red, realizando un producto escalar para luego ejecutar la función de activación. Sin embargo, la diferencia principal en la utilización de estas redes neuronales consiste principalmente en la extracción de datos o características a través de la operación de convolución.

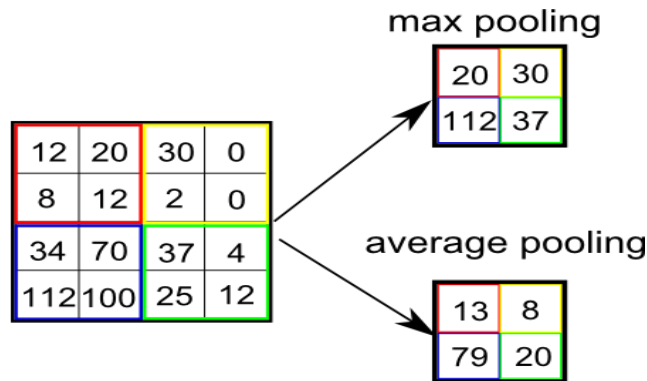
La convolución es una operación matemática que permite representar la magnitud de superposición de diferentes señales de entrada. En las capas convolucionales se realiza un recorrido en la señal superponiendo distintos *kernels* o filtros convolucionales con el fin de extraer ciertos patrones en la señal. Además de esto, las capas convolucionales también poseen un parámetro llamado *stride*, el cual corresponde al avance del superposicionamiento de los *kernels* en el recorrido de la matriz de entrada. En la figura 4 se muestra un ejemplo de una convolución de una matriz de entrada de  $7 \times 7$  con un filtro de  $3 \times 3$ .



**Figura 4.** Extracción de características utilizando la convolución.  
 Fuente: <http://www.diegocalvo.es/red-neuronal-convolucional-cnn/>

La operación de convolución, al ser operaciones sobre matrices, tiene un consumo significativo de memoria y de procesamiento. Con el fin de reducir la cantidad de operaciones y de evitar el sobreajuste de la red se utilizan capas de reducción (*pooling*). Las reglas de *pooling* mayormente utilizadas en la literatura consisten en la selección de las regiones con mayor valor de magnitud o valor con mayor relevancia (*max pooling*) y en obtener un promedio a partir de la sumatoria de todos los valores de la misma (*average pooling*). En la figura 5 se muestra un ejemplo de las reglas de *max pooling* y *average pooling* para una matriz de entrada aleatoria, en la cual se aplica la regla usando ambas operaciones.

**Figura 5.** *Max pooling* y *average pooling* para una matriz de entrada aleatoria. Fuente:



[10,13]

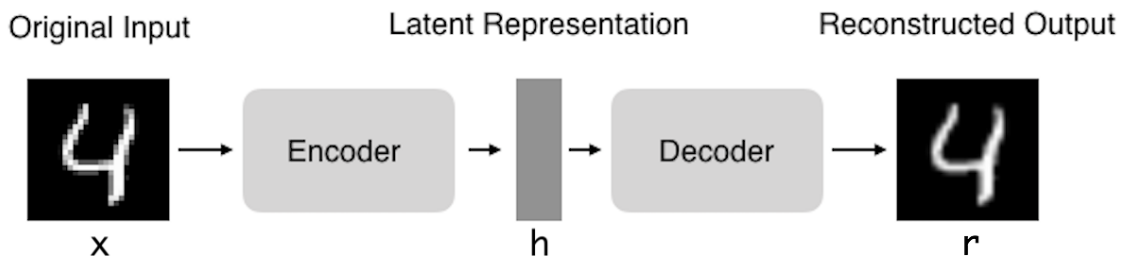
 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

## 2.2. APRENDIZAJE PROFUNDO O *DEEP LEARNING*

Los *autoencoders* hacen parte de las redes neuronales y su objetivo es generar nuevos datos, primero comprimiendo la entrada y luego reconstruyendo la salida en base a la información adquirida. Este tipo de red consta de dos partes:

- **Codificador (*Encoder*):** Se encarga de comprimir la entrada en un espacio de variables latentes y que puede representarse mediante la función de codificación  $h = f(x)$ .
- **Decodificador (*Decoder*):** Se encarga de reconstruir la entrada en base a la información recolectada previamente. Se representa mediante la función de decodificación  $r = g(h)$ .

El objetivo al momento de entrenar un *autoencoder* y copiar la entrada, es que el espacio de variables latentes  $h$  pueda tomar características útiles para nosotros, logrando imponer límites a la acción de codificación, forzando el espacio  $h$  que sea de menores dimensiones que  $x$ . En este caso, el autoencoder se denomina *undercomplete*. Al entrenar el espacio *undercomplete*, se capturan las características más relevantes de los datos de entrenamiento. Si no les damos suficientes restricciones, la red se limita a la tarea de copiar la entrada en la salida, sin replicar ninguna información útil sobre la distribución de datos. Esto puede suceder incluso cuando el tamaño del subespacio tiene el mismo tamaño que el espacio inicial; cuando tenemos un caso de autoencoder *overcomplete*, significa que el tamaño del espacio de variables latentes es mayor que las entradas. En estos casos, incluso con simples *encoder* y *decoder* lineales, podemos copiar la entrada en la salida sin aprender nada sobre los datos.



**Figura 6.** Arquitectura de un autoencoder. Fuente: [11]

Idealmente, es posible entrenar con éxito cualquier arquitectura basada en el autoencoder eligiendo adecuadamente los parámetros y la capacidad de cada encoder-decoder en función de la complejidad de los datos a modelar.

### EL USO DE LOS AUTOENCODER

La reducción de ruido y la reducción de dimensionalidad para la visualización de datos están entre las que se consideran las aplicaciones más interesantes de los autoencoders. Con la configuración apropiada de la dimensionalidad y las restricciones relacionadas en la dispersión de los datos, a través del autoencoder es posible obtener proyecciones en subespacios de mayor interés en comparación con métodos lineales como PCA. Al igual, los autoencoders se entrenan automáticamente a través de datos de muestra. Esto significa que es fácil entrenar la red para tener un buen rendimiento en tipos similares de entrada, sin la necesidad de generalizar. Es decir, solo la compresión realizada en datos similares a los utilizados en el conjunto de entrenamiento tendrá buenos resultados, pero si se realiza en diferentes datos no será muy efectiva.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Estas redes neuronales están capacitadas para conservar la mayor cantidad de información posible cuando se insertan en el encoder y luego en el decoder, pero también para que las nuevas representaciones adquieran diferentes tipos de propiedades.

**Autoencoder convolucional:** Este modelo generalmente está compuesto por dos capas denominadas *encoder* y *decoder*. Las cuales tienen como objetivo encontrar la mejor arquitectura para cada imagen de entrada, minimizando el *mean square error* (MSE) o el error cuadrático medio entre su entrada y salida, siendo el *MSE* el encargado de medir la diferencia o matemáticamente el promedio de los errores al cuadrado entre la imagen ingresada y la resultante [16].

Como se muestra en La figura [7], la imagen de entrada se muestra para obtener una representación latente, es decir, una reducción dimensional con los datos más relevantes de la imagen y a partir de esta reconstruye la salida en base a los datos adquiridos. De esta manera obliga al autoencoder a aprender de una versión comprimida de la imagen.



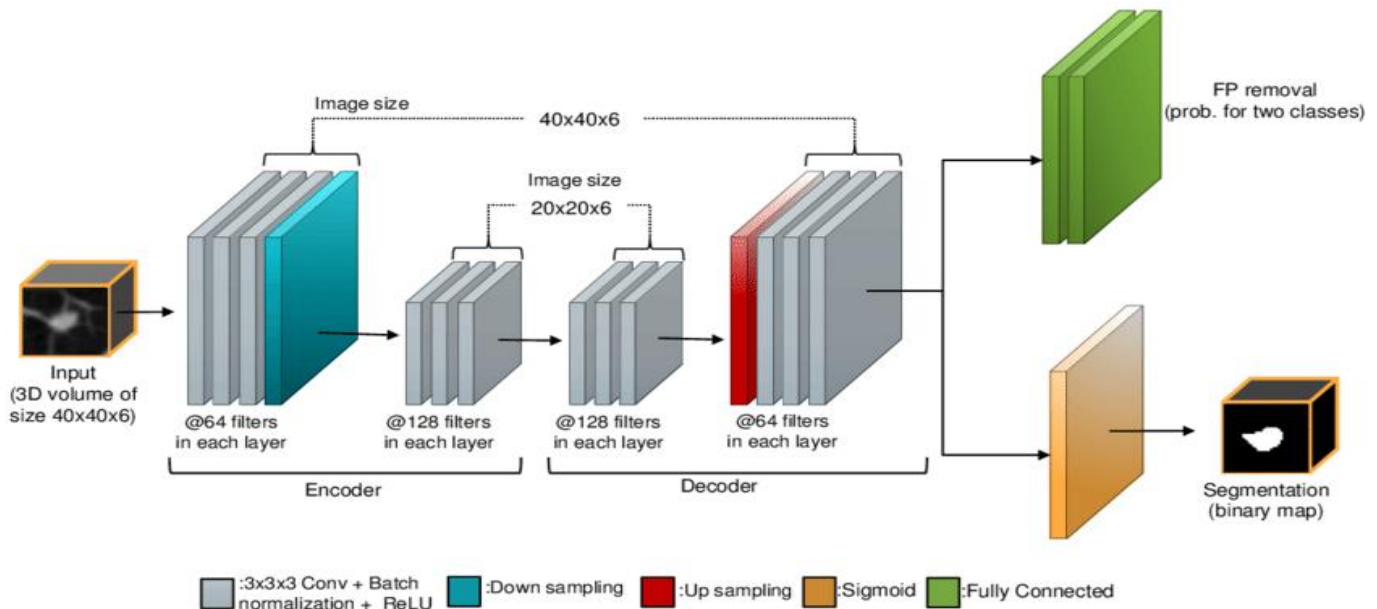


Figura 7. Ejemplo de una arquitectura de un autoencoder convolucional. Fuente: [9]

### 3. METODOLOGÍA

El desarrollo del trabajo se divide en 2 partes; la primera es el preprocesamiento, en esta sesión se realizan procesos de *resize* a las imágenes, también se generan las nuevas imágenes con ruido aleatorio y posterior a este se guardan como *Numpy Array* para facilitar las operaciones matemáticas que se realizan internamente en la red de aprendizaje profundo. La segunda parte incluye el modelo de *Deep Learning*, en el cual se aborda la estructura con la cual se implementó el modelo, realizando una explicación general de cada uno de los componentes que permitieron obtener buenos resultados al momento de mejorar la calidad de las imágenes.

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

## BASE DE DATOS QIN-BREAST

Esta colección contiene imágenes de PET/TC (de las siglas en inglés de *Positron Emission Tomography* o Tomografías por Emisión de Positrones) longitudinales y de resonancia magnética cuantitativas recogidas con el fin de estudiar la evaluación del tratamiento en el cáncer de mama en el entorno neoadyuvante [12]. La base de datos está conformada por un estudio de cáncer de mama que se le realizó a 13 pacientes con tumor cancerígeno confirmado, donde a cada uno se le tomaron entre 2 y 3 estudios, almacenando secuencias de 550 imágenes por estudio para un total de 17,500 imágenes. Las imágenes de esta base de datos se adquirieron en tres momentos: antes del inicio del tratamiento (t1), después del primer ciclo de tratamiento (t2) y después del segundo ciclo de tratamiento o al finalizar todos los tratamientos (antes de la cirugía) (t3). El objetivo de esta colección es proporcionar datos de imágenes clínicas para el desarrollo y evaluación de métodos de imágenes cuantitativos para la evaluación del tratamiento al principio del curso de la terapia para el cáncer de mama. Los datos son proporcionados por Vanderbilt University, PI Dr. Thomas E. Yankeelov [12].

Los datos de resonancia magnética consisten en imágenes ponderadas por difusión (DWI), imágenes dinámicas con contraste mejorado (DCE) y datos *multi-flip* para mapeo T1. Las resonancias magnéticas se obtuvieron utilizando una bobina mamaria bilateral de 16 canales dedicada a 3.0T (Philips Achieva con la mesa MammoTrak). Los DWI se adquirieron con una secuencia de imágenes planas de eco (EPI) de eco de espín (SE) de un solo disparo en tres direcciones de codificación de difusión ortogonal (x, y, y z). Para 14 pacientes,  $b = 0$  y  $500 \text{ s} / \text{mm}^2$ ,  $TR / TE = 2500 \text{ ms} / 45 \text{ ms}$   $\Delta = 21,4 \text{ ms}$ ,  $\delta = 10,3 \text{ ms}$  y se obtuvieron 10 adquisiciones de señal. Para 19 pacientes,  $b = 0$  y  $600 \text{ s} / \text{mm}^2$ ,  $TR / TE = \text{"más corto"}$  (rango =  $1800 - 3083 \text{ ms} / 43 - 60 \text{ ms}$ )  $\Delta = 20,7 - 29 \text{ ms}$ ,  $\delta = 11,4 - 21 \text{ ms}$  y 10 adquisiciones de señal fueron adquiridos. Para cuatro pacientes,  $b = 50$  y  $600 \text{ s} / \text{mm}^2$  para dos pacientes),  $TR / TE = \text{"más corto"}$  (rango =  $1840 - 3593 \text{ ms} / 43 - 60 \text{ ms}$ )  $\Delta = 20,6 - 29 \text{ ms}$ ,  $\delta = 11,5 - 21 \text{ ms}$  y se obtuvieron 10 adquisiciones de señales. Antes de la toma de datos de DCE-MRI, los datos

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

para construir un mapa T1 se tomaron con un enfoque de ángulo de giro múltiple de eco gradiente 3D estropeado por RF con diez ángulos de giro de 2 a 20 grados en incrementos de 20 grados. Tanto para el mapa T1 como para las exploraciones DCE, TR = 7,9 ms, TE = 4,6 ms, y la matriz de adquisición fue 192x192x20 (pecho completo) sobre un campo de visión cuadrado sagital (22 cm<sup>2</sup>) con un grosor de corte de 5 mm. Para el estudio DCE, cada conjunto de 20 cortes se recopiló en 16 segundos en 25 puntos de tiempo durante poco menos de siete minutos de escaneo dinámico.

### 3.1. PREPROCESAMIENTO

La preparación de la base de datos consiste en generar imágenes de 1.5T por medio de imágenes de 3T. Esto se logra introduciendo ruido a toda la imagen con la función (*np.random.normal*) de la librería de Numpy, encargada de generar una matriz con valores aleatorios por medio de la densidad de probabilidad de la distribución gaussiana. Posterior a esto se suma el ruido con la imagen original para obtener una resultante con su calidad degradada y de este modo generar una nueva base de datos con imágenes de un aspecto de 1.5T. (Véase Figura 8 y 9)

Principalmente se crea un script que se encarga de buscar las rutas en las cuales se encuentran cada una de las imágenes médicas que por defecto vienen con formato DICOM, para luego guardar estas rutas en una lista. Posterior a esto, por medio de la librería de *NumPy* se procede a generar el ruido, el cual se suma con la imagen original y entrega como resultado una imagen de menor calidad. Por último, se normalizan los datos, se modifica el tamaño de las imágenes dejando un estándar de 512 x 512 y guardando ambas imágenes (originales y con ruido) con un formato “. npy” el cual se utiliza para guardar archivos de matriz NumPy.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22



**Figura 8.** Imagen Original (3T)



**Figura 9.** Imagen con ruido (1.5T)

### 3.2. MODELO DEEP LEARNING

La estructura del autoencoder está compuesta por 2 partes; codificador (*encoder*) y decodificador (*decoder*). El *encoder*, encargado de codificar la información está diseñado con 8 capas convolucionales las cuales se encargan de extraer patrones en la señal. Entre capa y capa se extraen las características más relevantes de la imagen reduciendo cada vez más el tamaño de esta y ampliando su profundidad, siendo esta nueva imagen la entrada de la siguiente capa. Posterior a esto, luego de cada capa convolucional, se aplica una agrupación o *Pooling* con el objetivo de generar otro filtro el cual solo va a reducir el tamaño de la imagen, agrupando varios píxeles en uno solo, esto nos ayuda a evitar el sobreajuste y a minimizar la carga computacional que se requiere entre capa y capa. El número de filtros de las capas convolucionales inicia en 16 y va aumentando ascendentemente hasta 128. Al igual, el tamaño del *kernel* que recorre la matriz para la extracción de características es de 3 x 3, con una función de activación *rectified linear unit* (ReLU) [7].

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Para lograr que la salida reconstruida sea lo más parecida posible a la entrada, se tiene que capturar toda la información relevante extraída en cada una de las capas convolucionales anteriores, guardando esta información en la zona media de la estructura, denominada espacio de variables latente o cuello de botella (*bottleneck*). A partir de este, se empieza a decodificar la información reconstruyendo la imagen en base a los datos codificados, ya que al reducir las dimensiones permite almacenar los datos en un espacio de menores dimensiones al original y de esta manera se conserva la mayor información posible. El decodificador está compuesto por 8 capas convolucionales encargadas de reconstruir la información adquirida, y a diferencia del *encoder* que usa *MaxPooling* para disminuir el tamaño de la imagen, el decoder tiene una capa muestreo ascendente o *UpSampling* antes de cada capa convolucional, con el fin de aumentar la resolución espacial de la imagen [17].

El optimizador utilizado para el ajuste de parámetros es el *Adaptive Moment Estimation*, el cual está basado en el algoritmo del descenso de los gradientes, además de considerar dos momentos matemáticos para el control del descenso durante la optimización. Esto incrementa considerablemente la eficiencia en la optimización. Como función objetivo, se emplea el error cuadrático medio (MSE) y el error medio absoluto (MAE) con el fin de reducir la diferencia entre los valores de intensidad de la imagen real con respecto a la imagen de predicción. La tasa de aprendizaje de  $1e-4$  fue con la que se obtuvo mejores resultados [12]. En la figura 10, se observa la estructura del modelo utilizado y sirve como entendimiento gráfico de lo explicado anteriormente.

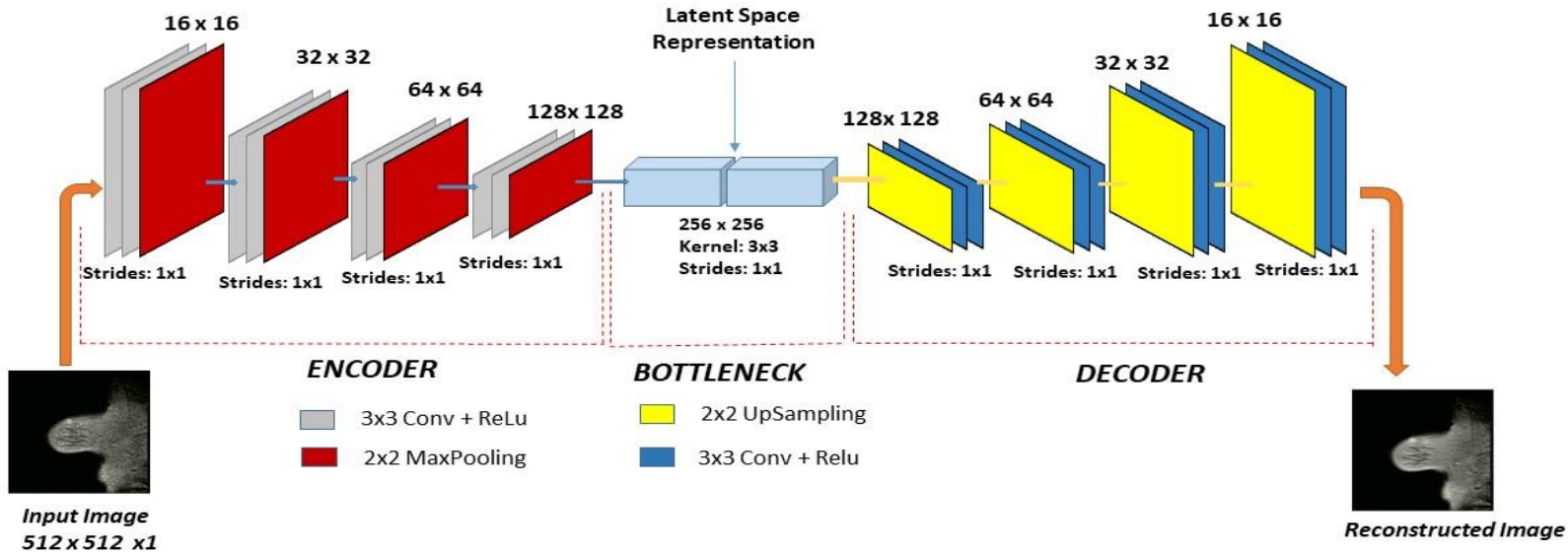


Figura 10. Arquitectura del modelo (Codificador y Decodificador). Fuente: Autores

### 3.3 MÉTRICAS DE DESEMPEÑO

Con el fin de evaluar la calidad de las imágenes, se pretende extraer las siguientes métricas de desempeño evaluadas entre las imágenes con calidad real (3T) y las predicciones entregadas por el modelo. A continuación, se describen algunas de estas métricas.

#### 3.3.1. Error cuadrático medio (MSE)

El error cuadrático medio (MSE) entre dos imágenes pretende medir la fidelidad espectral entre ellas, considerando la proporción de cambio por píxeles. Se basa en las diferencias cuadráticas entre las magnitudes espaciales del mapa de referencia  $\delta_{ref}$ , y la magnitud de en el mapa generado con un modelo convolucional  $\delta_{pred}$ , como se indica en la Ecuación (1). Allí, los valores cercanos a cero indican una alta similitud entre los mapas de diezmos y, por lo tanto, un alto rendimiento del modelo.

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

$$MSE = \sqrt{\frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N [\delta_{ref}(x, y) - \delta_{pred}(x, y)]^2} \quad (1)$$

### 3.3.2. Relación pico de ruido de la señal (PSNR)

La relación pico de ruido a señal (PSNR) tiene como objetivo medir la relación entre la potencia máxima posible de una señal y la potencia del ruido corruptor que afecta la fidelidad de su representación. El valor de PSNR se acerca al infinito cuando el MSE se acerca a cero; esto muestra que un valor PSNR más alto proporciona una calidad de imagen más alta, como se indica en la ecuación (2). En esta métrica, los valores altos indican un mejor desempeño, por el contrario, los valores bajos indican un desempeño bajo.

$$PSNR = 10 \log_{10} \left( \frac{\max(\delta_{ref})}{MSE} \right) \quad (2)$$

### 3.3.3. Medida de similitud estructural (SSIM)

La medida de similitud estructural (SSIM) es una métrica de calidad conocida que se utiliza para medir la similitud entre dos imágenes. Fue desarrollado por Wang et al. [18]. Se considera que está correlacionado con la percepción de la calidad del sistema visual humano (HVS). En lugar de utilizar los métodos tradicionales de suma de errores, el SSIM se diseña modelando cualquier distorsión de la imagen como una combinación de tres factores que son la pérdida de correlación, la distorsión de luminancia y la distorsión de contraste, como se indica en la Ecuación 1. Allí,  $\delta$  es la media,  $\sigma$  es la covarianza y "C1" y "C2" son coeficientes de compensación que dependen del rango dinámico de los píxeles. Los

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

resultados de SSIM cercanos a "1" indican alto rendimiento, los resultados cercanos a "0" indican bajo rendimiento.

$$SSIM = \frac{(2 * \delta_{ref} * \delta_{pred} + C1) * (2 * \sigma + C2)}{(\delta_{ref}^2 + \delta_{pred}^2 + C1) * (\sigma_{ref}^2 + \sigma_{pred}^2 + C2)} \quad (3)$$

## 4 RESULTADOS Y DISCUSIÓN

---

Inicialmente, la base de datos empleada contiene 17.500 imágenes que es dividida en dos subconjuntos: entrenamiento (12.250) y validación (5.250). La división de estos datos se realiza teniendo en cuenta la información de los pacientes con el fin de que *slices* que pertenecen a un mismo paciente no se encuentren en ambos subconjuntos de datos. Esto garantiza que el mejoramiento de la calidad de las imágenes no depende de la distribución de los pacientes que se han usado para el entrenamiento, sino que el mejoramiento depende de la distribución de los datos de la misma imagen. Además, todos los experimentos se realizan sobre una estación de trabajo con un procesador Intel Core i7 con 16 núcleos de procesamiento, 32GB de memoria RAM y una GPU NVIDIA Quadro P2000 (4GB de RAM y 1792 núcleos).

En primera instancia, se presenta la curva de pérdida obtenida por el modelo (figura X). En esta es posible evidenciar una minimización efectiva de la función objetivo. El valor de minimización es presentado por época y posee un valor pequeño debido a la normalización de las imágenes. Debido al alto costo computacional, el modelo es entrenado por 10 épocas (2 días, aprox), sin embargo; un mayor entrenamiento puede concluir en una mayor minimización del error, por consiguiente, en una mejor sintetización y reconstrucción de las imágenes. Esto se planteará como trabajo posterior a este.



	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

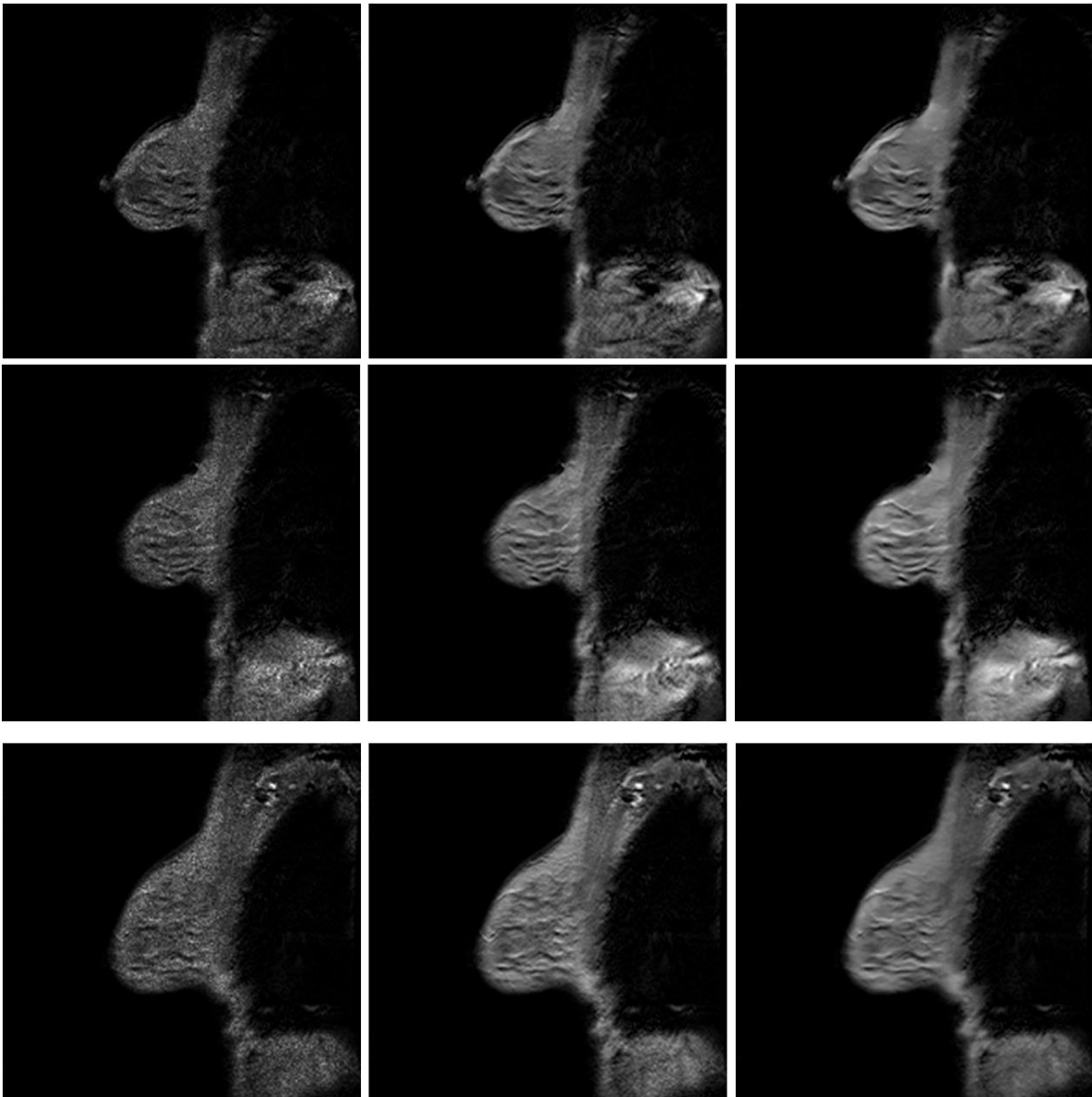
En la figura 12 es posible visualizar algunas de las imágenes generadas por el modelo. En estas es posible encontrar que la reconstrucción corresponde visualmente con las imágenes en la calidad original. Además, los valores de MSE, PSNR y SSIM estimados muestran mejoría en cuanto a similitud entre ambas imágenes. Esto se puede interpretar como que la diferencia entre la imagen de entrada y la de salida es muy mínima. Por otra parte, en la figura 13 en donde se ven algunas de las regiones de tejido aumentadas es posible visualizar que a pesar de que la correspondencia visual es suficiente, existe una pérdida de algunos detalles mínimos en cuanto a los detalles internos en el tejido. Sin embargo, estos detalles no comprometen aspectos importantes como son forma o aspecto de los tejidos, sino la apariencia en cuanto a valores de intensidad, por lo que en principio la correspondencia estructural no es afectada. A pesar de esto, teniendo en cuenta la relevancia de las estructuras microscópicas que pueden determinar el diagnóstico de los pacientes, es importante realizar un estudio más exhaustivo con el fin de determinar la calidad de estos apoyado en expertos radiólogos. Esto se planteará como trabajo posterior a este. La tabla 1 muestra los resultados promedios obtenidos para cada una de las métricas de desempeño en el subconjunto de validación.

MSE	PSNR	SSIM
0.009342	29.181462	0.930107

**Tabla 1:** Valores promedio para las métricas de desempeño en el conjunto de datos de validación

Por otra parte, si bien los resultados muestran una correspondencia visual eficiente; la calidad de las imágenes reales continúa siendo superior. Esto quiere decir que el modelo presentado dista de reemplazar los equipos de adquisición de resonancia magnética en 3T; sin embargo, muestran una herramienta inicial con la cual se podría mejorar significativamente la calidad de las imágenes en baja/mediana resolución. Esto motiva a

continuar con la investigación de los modelos de aprendizaje profundo para el mejoramiento de la calidad de las imágenes médicas con modelos más robustos, no solo bajo la modalidad de resonancia magnética sino también frente a otras modalidades como la malograda completa de rango abierto (FFDM), tomografía computarizada (CT) o tomografía por emisión de positrones (PET).



**Figura 11, 12, 13.** Ejemplos de distintos resultados gráficos obtenidos por el modelo. Las imágenes de la columna izquierda corresponden con las imágenes de menor calidad; las de la columna de en medio corresponden con las imágenes de salida reales y la última columna con las imágenes reconstruidas

MSE	PSNR	SSIM
0.000371	34.302165	0.972619

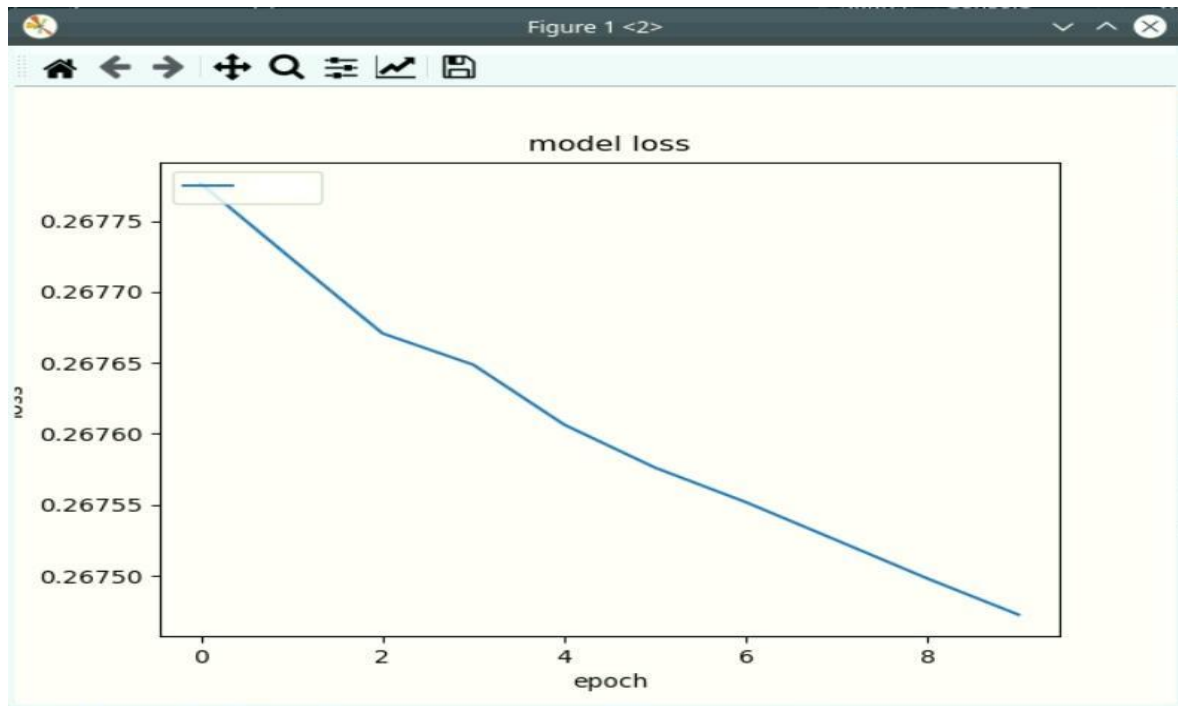
**Tabla 2:** Valores correspondientes a la Fig 11. para las métricas de desempeño en el conjunto de datos de validación

MSE	PSNR	SSIM
0.000766	31.160374	0.965625

**Tabla 3:** Valores correspondientes a la Fig 12. para las métricas de desempeño en el conjunto de datos de validación

MSE	PSNR	SSIM
0.000880	30.554554	0.959446

**Tabla 4:** Valores correspondientes a la Fig 13. para las métricas de desempeño en el conjunto de datos de validación



**Figura 14.** Gráfica de minimización de la función objetivo durante el aprendizaje. El eje vertical corresponde con el valor de MSE ponderado por época y el eje horizontal el número de épocas.

	<p style="text-align: center;">INFORME FINAL DE TRABAJO DE GRADO</p>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

## 5 CONCLUSIONES, RECOMENDACIONES Y TRABAJO FUTURO

---

Por medio de un modelo de aprendizaje profundo, específicamente un autoencoder, se logra mejorar la calidad de las imágenes médicas de resonancia magnéticas de la mama pasando de una calidad de 1.5T a una de 3T. Para el ámbito médico esto es de mucha importancia ya que permite dar un diagnóstico más detallado al paciente y por medio de este modelo de aprendizaje profundo contribuimos computacionalmente a la problemática de los costos de los resonadores médicos.

La preparación de la base de datos es de gran ayuda para los resultados positivos ya que la cantidad de imágenes que se seleccionan para el entrenamiento y la validación determina que no haya sobreajustes o que el modelo no aprenda de una manera correcta. El entrenamiento se implementó con el 70% de los datos, dejando el 30% de estos para validar. También, se tomaron las medidas necesarias con la selección de los datos para evitar el sobreajuste, generando que este se ajuste bien a los datos de entrenamiento y como resultado realizar una predicción precisa sobre los datos de prueba desconocidos.

Como futuros trabajos se pretende establecer una estrategia basada en una GAN (*Generative Adversarial Networks*) con el fin de mejorar el desempeño de la arquitectura propuesta. Adicionalmente, se realizará una evaluación cualitativa considerando la percepción de expertos en radiología que permita determinar la visualización de las estructuras internas del tejido mamario. Por último, se pretende buscar la implementación de modelos para el mejoramiento de las imágenes de calidad 1.5T a 7T, la cual es la mayor calidad en términos de adquisición actualmente.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

## REFERENCIAS

---

- [1] Y. & G. H. Yann LeCun, “Deep learning,” *Deep Learn.*
- [2] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” *J. Mach. Learn. Res.*, vol. 12, no. Oct, pp. 2825–2830, 2011.
- [3] F. Porikli, S. Shan, C. Snoek, R. Sukthankar, and X. Wang, “Deep Learning for Visual Understanding: Part 2 [From the Guest Editors],” *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 17–19, 2018.
- [4] F. N. Sibai, H. I. Hosani, R. M. Naqbi, S. Dhanhani, and S. Shehhi, “Iris recognition using artificial neural networks,” 2011.
- [5] Mario Campos Soberanis, “Inspiración biológica de las redes neuronales artificiales,” 2017. [Online]. Available: <https://medium.com/soldai/inspiración-biológica-de-las-redes-neuronales-artificiales-9af7d7b906a>. [Accessed: 31-May-2018].
- [6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks.”
- [7] V. Nair and G. E. Hinton, “Rectified Linear Units Improve Restricted Boltzmann Machines,” *Proc. 27th Int. Conf. Mach. Learn.*, no. 3, pp. 807–814, 2010.
- [8] P. Flores and J. Braun, “Algoritmo SIFT: fundamento teórico,” 2011.
- [9] Khosravan, N., Celik, H., Turkbey, B., Jones, E. C., Wood, B., & Bagci, U. (2019). A collaborative computer aided diagnosis (C-CAD) system with eye-tracking, sparse attentional model, and deep learning. *Medical image analysis*, 51, 101-115
- [10] Singhal, H. (2020, August 11). *Convolutional Neural Network with TensorFlow implementation*. Medium. <https://medium.com/data-science-group-iitr/building-a-convolutional-neural-network-in-python-with-tensorflow-d251c3ca8117>.
- [11] Alberti, M. (2018, May 14). *Introducción al autoencoder*. DeepLearningItalia. <https://www.deeplearningitalia.com/introduzione-agli-autoencoder-2/>.
- [12] Li X, Abramson RG, Arlinghaus LR, Kang H, Chakravarthy AB, Abramson VG, Farley J, Mayer IA, Kelley MC, Meszoely IM, Means-Powell J, Grau AM, Sanders M, Yankeelov TE. *Multiparametric magnetic resonance imaging for predicting pathological response after the first cycle of neoadjuvant chemotherapy in breast cancer*. *Investigative Radiology*, 2015 Apr;50(4):195-204. PMID: PMC4471951.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

- [13] Nagi, J., Ducatelle, F., Di Caro, G. A., Cireşan, D., Meier, U., Giusti, A., ... & Gambardella, L. M. (2011, November). Max-pooling convolutional neural networks for vision-based hand gesture recognition. In *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)* (pp. 342-347). IEEE
- [14] Bahrami, K., Shi, F., Zong, X., Shin, H. W., An, H., & Shen, D. (2016). Reconstruction of 7T-like images from 3T MRI. *IEEE transactions on medical imaging*, 35(9), 2085-2097.
- [15] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 1097-1105.
- [16] Guo X., Liu X., Zhu E., Yin J. (2017) Deep Clustering with Convolutional Autoencoders. In: Liu D., Xie S., Li Y., Zhao D., El-Alfy ES. (eds) Neural Information Processing. ICONIP 2017. Lecture Notes in Computer Science, vol 10635. Springer, Cham.  
[https://doi.org/10.1007/978-3-319-70096-0\\_39](https://doi.org/10.1007/978-3-319-70096-0_39).
- [17] Shan, Q., Li, Z., Jia, J., & Tang, C. K. (2008). Fast image/video upsampling. *ACM Transactions on Graphics (TOG)*, 27(5), 1-7.

	<b>INFORME FINAL DE TRABAJO DE GRADO</b>	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

# APÉNDICE

---

## Apéndice A

Archivo de Código en Python llamado “Data\_Factory.py”: Este archivo contiene el generador de datos que se encarga de la carga dinámica de las imágenes para el entrenamiento del modelo.

## Apéndice B

Archivo de Código en Python llamado “Model.py”: Este archivo contiene el modelo de aprendizaje profundo. En él, se definen todos los parámetros dinámicos iniciales del mismo, previo a su entrenamiento.

## Apéndice C

Archivo de Código en Python llamado “Preprocessing.py”: en este archivo se realiza toda la estructuración y preprocesamiento de las imágenes previo al generador de datos. Este script genera los archivos necesarios para el óptimo funcionamiento del generador.

## Apéndice D

Archivo de Código en Python llamado “main.py”: Este archivo crea instancias del generador de datos y de la arquitectura del modelo, ensamblándolos para el aprendizaje. El él, el modelo entrena; es guardado en un archivo de datos en disco duro y demás, genera predicciones aleatorias del subconjunto de prueba y con sus respectivas métricas.



FIRMA ESTUDIANTES Pablo Campaiz

FIRMA ASESOR [Signature]

FIRMA ASESOR [Signature]

FECHA ENTREGA: 16/07/2021

FIRMA COMITÉ TRABAJO DE GRADO DE LA FACULTAD \_\_\_\_\_

RECHAZADO\_\_\_      ACEPTADO\_\_\_      ACEPTADO CON MODIFICACIONES\_\_\_

ACTA NO. \_\_\_\_\_

FECHA ENTREGA: \_\_\_\_\_

FIRMA CONSEJO DE FACULTAD \_\_\_\_\_

ACTA NO. \_\_\_\_\_

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

FECHA ENTREGA: \_\_\_\_\_