

**DESARROLLO DE UNA RED NEURONAL CONVOLUCIONAL PARA LA
DETECCIÓN DEL CÁNCER DE MAMA MEDIANTE LA CLASIFICACIÓN DE
IMÁGENES MAMOGRÁFICAS****DEVELOPMENT OF A CONVOLUTIONAL NEURONAL NETWORK FOR THE
DETECTION OF BREAST CANCER THROUGH THE CLASSIFICATION OF
MAMMOGRAPHIC IMAGES****Sergio Augusto Celis Esteban*, Jhoan Felipe Sarmiento Ortiz*
PhD. Liliana Calderón-Benavides****** Universidad Autónoma de Bucaramanga**

Facultad de Ingeniería, Ingeniería de Sistemas.

Av. 42 # 48 - 11, Bucaramanga, Santander, Colombia.

Tel.: (57) (7) 643 6111/643 6261

E-mail: {scelis390, jsarmiento99, mcalderon}@unab.edu.co

Orcid: <https://orcid.org/0000-0001-8658-9036>Orcid: <https://orcid.org/0000-0003-3132-0237>Orcid: <https://orcid.org/0000-0001-8658-9036>

Resumen: La inteligencia artificial (IA) ha venido creciendo durante los últimos años en el área de la salud con el desarrollo de sistemas de apoyo a la toma de decisiones clínicas. Con este trabajo se logró desarrollar un algoritmo de aprendizaje profundo capaz de clasificar imágenes mamográficas en cinco categorías (normal, microcalcificación benigna, nódulo benigno, microcalcificación maligna y nódulo maligno) con un enfoque prioritario en la detección temprana del cáncer de mama, aplicando la técnica de transferencia de aprendizaje. Se usaron los conjuntos de datos DDSM y CBIS-DDSM, disponibles en la web, para el entrenamiento y validación de la red neuronal convolucional obteniendo un AUC del 0.9838 y 0.9773 respectivamente. Estos resultados demuestran el gran potencial que la IA trae para el área de la salud, y los beneficios que genera en esta y otras patologías al reducir el porcentaje de falsos positivos y falsos negativos que son elementos importantes en el diagnóstico.

Palabras clave: Inteligencia artificial, aprendizaje profundo, red neuronal convolucional, transferencia de aprendizaje, clasificación de imágenes, cáncer de mama, detección temprana.

Abstract: Artificial intelligence (AI) has been growing in recent years in the health area with the development of support systems for clinical decision making. With this work, it was possible to develop a deep learning algorithm capable of classifying mammographic images into five categories (normal, benign microcalcification, benign nodule, malignant microcalcification and malignant nodule) with a priority focus on the early detection of breast cancer, applying the technique of learning transfer. The DDSM and CBIS-DDSM data sets, available on the web, were used for the training and validation of the convolutional neural network obtaining an AUC of 0.9838 and 0.9773 respectively. These results demonstrate the great potential that AI brings to the health area, and the benefits it generates in this and other pathologies by reducing the percentage of false positives and false negatives that are important elements in the diagnosis.

Keywords: Artificial intelligence, deep learning, convolutional neural network, transfer learning, image classification, breast cancer, early detection.

1. INTRODUCCIÓN

Según el Observatorio Global de Cáncer (GLOBOCAN) en su último informe publicado, el tipo de cáncer más frecuente en las mujeres de todo el mundo es el cáncer de mama con un total de 2'088.849 nuevos casos registrados (Global Cancer Observatory, 2018). Existen múltiples factores de riesgo y la tasa de supervivencia varía de acuerdo al estadio en que se detecta la enfermedad. El diagnóstico tardío disminuye hasta un 7% las probabilidades de supervivencia, mientras que, si el tejido tumoral se logra detectar en estadio temprano, la tasa de supervivencia puede llegar a ser del 99%. En base a esta realidad se propuso el desarrollo de una herramienta soporte con enfoque en la detección temprana de carcinomas mamarios mediante un modelo de red neuronal convolucional para la clasificación de parches de imágenes mamográficas, donde se evaluaron tres arquitecturas distintas para el mismo conjunto de datos determinando que la mejor opción para este caso era la VGG16, tanto por los resultados obtenidos en las métricas de evaluación como por el tiempo reducido de entrenamiento en comparación a las demás (ResNet50 e InceptionV3), (Chacón G, et al, 2018), Rios, Y., et al, 2020).

2. CONJUNTO DE DATOS

Para la construcción del *dataset* se tuvieron en cuenta las bases de datos DDSM con 2.620 estudios de los cuales se tomaron únicamente 695 que hacen referencia a 2.780 imágenes de pantalla completa en formato JPEG pertenecientes a la categoría normal (N) (Rose, 2006); y CIBIS-DDSM con 1.815 mamografías en formato DICOM distribuidas en cuatro categorías: nódulo maligno (NM), nódulo benigno (NB), microcalcificación maligna (MM), microcalcificación benigna (MB), y que, a diferencia de la primera, contaba con las regiones de interés (ROI) (Nolan, 2020). Fue necesario generar las ROI para el primer conjunto de datos, y se formalizó un *dataset* de 4.595 parches de imágenes a los que posteriormente se les hicieron recortes con variaciones en la posición para aumentar el número de datos, obteniendo un total de 23.527. Finalmente, de este total se tomaron únicamente 12.000 imágenes distribuidas equitativamente en las cinco categorías, como se detalla en la Tabla 1, con el fin de tener un conjunto de datos balanceado.

Tabla 1. Distribución de datos por categoría.

	N	NM	NB	MM	MB
N° de datos	2.707	2.661	2.553	1.832	2.247

Nota. Fuente propia

3. METODOLOGÍA

Una vez definido el número total de datos, se inició el procesamiento de los mismos para llevar a cabo el entrenamiento de la red neuronal.

3.1 Metodología de los datos.

Se tuvieron en cuenta cuatro etapas para el tratamiento de los datos (Figura 1).

En primer lugar, se descargaron las bases de datos anteriormente descritas, de los servidores de la Universidad del Sur de Florida y el Archivo de Imágenes de Cáncer.

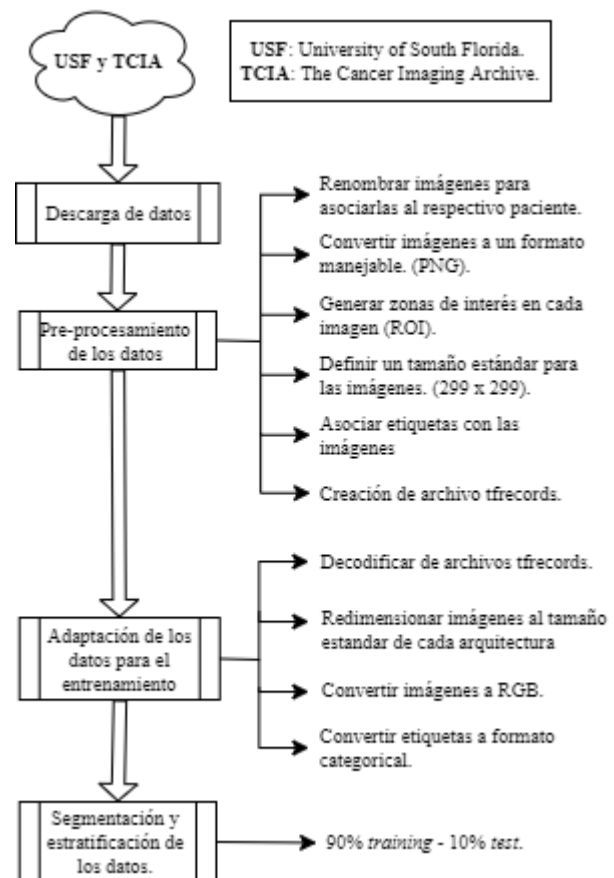


Fig. 1. Metodología empleada para el tratamiento de los datos. Fuente propia.

El pre-procesamiento de los datos inició con renombrar las imágenes para poder asociarlas a cada paciente ya que la información de cada uno venía separada por directorios y subdirectorios, pero sin un identificador único dentro del conjunto total de datos. Se concatenó el nombre del directorio, del subdirectorio y del archivo como se muestra en la Figura 2.

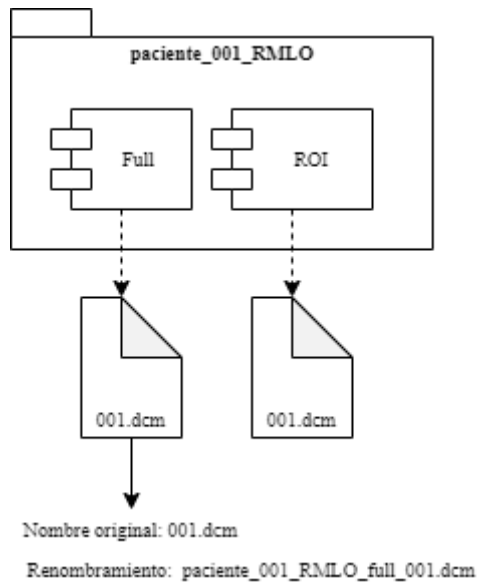


Fig. 2. Ejemplo renombramiento de archivos.
Fuente propia.

Posteriormente, haciendo uso de las herramientas `jpeg.exe`, `ddsmraw2pnm.exe` (Jalandhar y Sharma, 2016) y la librería `Pydicom` se convirtieron las imágenes a formato PNG para poder ser utilizadas en la red neuronal; se eligió este formato ya que es el recomendado para evitar la pérdida de calidad en las imágenes (Nolan, 2020). Además, como se mencionó anteriormente, para el conjunto de datos DDSM fue necesario generar las ROI y se estableció un tamaño de 450×450 px para todas las máscaras. Debido a las dimensiones que exigen las redes neuronales convolucionales, se generaron recortes definiendo un tamaño estándar de 299×299 px. Las etiquetas correspondientes a cada imagen, estaban disponibles en un archivo CSV y fue necesario asociarlas mediante un `dataframe` para poder crear los cinco archivos `trecords` con los que se entrenó cada modelo.

Culminada la etapa de pre-procesamiento fue necesario adaptar los datos para cada red. Para esto, lo primero que se hizo, fue tomar los archivos `trecords` que se habían generado, y se

decodificaron, obteniendo de esta forma, las imágenes (*data*) y etiquetas (*labels*) por aparte. Las dimensiones que anteriormente se habían estandarizado de 299×299 px, tuvieron que modificarse conforme al tamaño de entrada requerido por las arquitecturas seleccionadas (Tabla 2). Este proceso, se llevó a cabo mediante el uso de la librería `OpenCV`, facilitada por `Python`.

Tabla 2. Dimensiones de entrada

	ResNet 50	VGG16	InceptionV3
Tamaño de entrada	224×244	244×244	299×299

Nota. Fuente propia.

Además de los cambios en la dimensión de las imágenes, fue necesario convertir la *data*, a RGB (*red, green and blue*), ya que los modelos requerían que las entradas tuvieran estos tres canales. De igual manera, se hizo mediante la librería `OpenCV`.

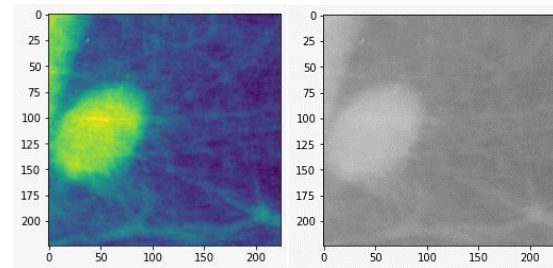


Fig. 3. Imagen diagnóstica antes y después de ser convertida a RGB. Fuente Propia.

Con respecto a las etiquetas, por medio de la librería `Numpy`, se les indicó que adquirieran un valor categorial. De esta manera, cada etiqueta quedaba representada por medio de un arreglo que contenía un 1 en la posición correspondiente a su etiqueta, y lo demás con ceros. Por ejemplo, el arreglo `[1, 0, 0, 0, 0]`, representaba la etiqueta “normal”, el arreglo `[0, 1, 0, 0, 0]`, representaba la etiqueta “microcalcificación benigna”, y así sucesivamente.

Finalmente, se definieron dos escenarios de segmentación de los datos, en primer lugar 70% para *training* y 30% para *test* con el que se realizó una primera prueba con el fin de determinar cuál era el modelo óptimo de acuerdo a los criterios de decisión planteados. Y en segundo lugar el escenario de 90% para *training* y 10% para *test*, aplicado exclusivamente para el modelo elegido.

3.2 Selección y construcción de la red neuronal

Para la selección de la arquitectura se tuvo en cuenta principalmente el tiempo de entrenamiento del modelo debido a las limitaciones computacionales presentados en el desarrollo del proyecto. En tanto, la red seleccionada fue VGG16 y presentó la siguiente configuración de parámetros (Tabla 3):

Tabla 3. Configuración de parámetros VGG16

Parámetro	Valor
Tamaño de entrada	224×224
Tamaño de <i>batch</i>	32
N° de Épocas	70
Optimizador	adam
Tasa de aprendizaje	1×10^{-3} , 1×10^{-4} , 1×10^{-5}
Include_top	false
Weights	imagenet
Pooling	avg
Clases	5
Classifier_activation	Softmax

Nota. Fuente propia.

Se definió como *false* el parámetro *include_top* para tomar únicamente los bloques convolucionales, es decir, las *full connected layers* y la capa densa de predicción, ubicadas en la parte superior de la arquitectura, fueron removidas. El parámetro *weights* carga los pesos predeterminados que el modelo VGG16 trae consigo, esto para que la red no tome pesos aleatorios, sino que inicie con pesos ya ajustados. Se definió, además, las dimensiones de entrada en (224, 224, 3), puesto que esta es la forma recomendada por la arquitectura para su input layer. Para minimizar el número total de parámetros se agregó una *global average pooling layer*, que ayuda a reducir el sobreajuste del modelo. Además, se estableció el parámetro clases como cinco, debido a que este es el número de categorías del modelo y, por último, se utilizó la función de activación *softmax* en la última capa, para obtener una salida en forma de probabilidades.

las últimas capas de la arquitectura fueron reemplazadas por capas personalizadas y adaptadas al objeto de estudio del proyecto. Al momento del entrenamiento, se definió una tasa inicial de aprendizaje de 0.001 (1×10^{-3}) para tres épocas.

Se realizaron entonces, dos entrenamientos posteriores en base al modelo obtenido, con un método progresivo, que consistía en ir descongelando poco a poco capas de la arquitectura y a la vez, reduciendo la tasa de aprendizaje. Para

el segundo entrenamiento la tasa de aprendizaje se definió en 0.0001 (1×10^{-4}), y se descongelaron los tres bloques convolucionales superiores de la arquitectura (doce capas), es decir, se volvieron entrenables. Se manejaron los mismos parámetros que en el caso anterior para diez épocas.

Finalmente, en el tercer entrenamiento se realizó el mismo proceso, esta vez, descongelando todas las capas, se hizo entrenable toda la red, y se utilizó una tasa de aprendizaje de 0.00001 (1×10^{-5}), en setenta épocas.

4. RESULTADOS OBTENIDOS

Para el modelo final de la red seleccionada, se obtuvo la siguiente matriz de confusión (Figura 4).

true label	normal	260 (0.96)	9 (0.03)	0 (0.00)	2 (0.01)	0 (0.00)
	benignCalcification	17 (0.06)	213 (0.80)	4 (0.02)	30 (0.11)	2 (0.01)
	benignMass	21 (0.08)	12 (0.05)	175 (0.69)	9 (0.04)	38 (0.15)
	malignantCalcification	16 (0.09)	38 (0.21)	3 (0.02)	120 (0.66)	6 (0.03)
	malignantMass	14 (0.06)	2 (0.01)	19 (0.08)	12 (0.05)	178 (0.79)
		normal	benignCalcification	benignMass	malignantCalcification	malignantMass
		predicted label				

Fig. 4. Matriz de confusión VGG16

Además, se tuvieron en cuenta, seis métricas de evaluación para la red neuronal, cuyos resultados se observan en la Tabla 4.

Tabla 4. Métricas de evaluación VGG16

	Training	Validation
Accuracy	0.8703	0.8278
AUC	0.9838	0.9773
Loss	0.3462	0.4492
Precision	0.9011	0.8485
Recall	0.8385	0.8037
F1-score	0.8683	0.8262

Nota. Fuente propia.

5. CONCLUSIÓN

El proyecto descrito permite observar que el mejor modelo se obtuvo con la red VGG16 tomando como parámetros discriminatorios la complejidad computacional. Además, la metodología implementada en el entrenamiento de la red resultó eficiente para la optimización de las métricas de evaluación a medida que aumentaban las iteraciones del entrenamiento. Encontrando así, que para el tercer entrenamiento implementando transferencia de aprendizaje y seleccionando 1×10^{-5} como tasa de aprendizaje se obtuvieron las mejores métricas de desempeño: AUC 0.9838 para *training* y 0.9773 en *validation*. Además, se evidenció que el método implementado para el aumento de datos y el balanceo del *dataset* fue eficiente presentando imágenes sintéticas que aportaron correctamente al proyecto.

Otro punto importante de la investigación fue evidenciar la importancia que conlleva tener un pleno conocimiento del *dataset* que se está manejando y tener un real balance en el número de los datos de cada clase para evitar problemas comunes como overfitting y underfitting. Una buena solución para este tipo de problemas es la aplicación de funciones de regularización de los datos (Ej.: *dropout*), así como hacer uso del *transfer learning* que es de gran ayuda en la construcción de los modelos, y aporta a obtener mejores resultados evitando en parte problemas de sobre ajuste. En esta investigación se aplicó tres veces esta técnica, con el objetivo que el modelo aprendiera cada vez mejor de sí mismo. Se recomienda, mantener un número equitativo de los datos, así como realizar una estratificación de los mismos al momento de entrenar la red, para que el modelo pueda aprender sin sesgo o preferencia por alguna categoría.

Cabe resaltar, que se pudo haber construido un sistema de clasificación binaria, es decir clasificar los casos únicamente como benignos o malignos, y esto habría generado resultados incluso más altos en el rendimiento del sistema. Sin embargo, se quiso desarrollar un modelo capaz de identificar el cáncer en su etapa temprana, por lo que fue necesario la implementación de nuevas clases en la arquitectura. Se habla entonces, de las microcalcificaciones, las cuales al igual que los nódulos se dividieron en malignas y benignas. El hallazgo del cáncer en esta fase trae ventajas al paciente pues le permite iniciar el tratamiento de la enfermedad justo a tiempo. Esto genera también

una optimización mayor en la labor de los radiólogos ya que las microcalcificaciones traen en ocasiones complicaciones para identificarse a simple vista. Además, se decidió añadir la clase “normal” puesto que, el tejido mamario común por su densidad puede llegar confundirse con la existencia de nódulos mamarios.

REFERENCIAS

- Ahmad, A. M., Muhammad Khan, G., & Mahmud, S. A. (2014). Classification of Mammograms Using Cartesian Genetic Programming Evolved Artificial Neural Networks. *Artificial Intelligence Applications and Innovations. AIAI. IFIP Advances in Information and Communication Technology*. 436, págs. 203-213. Berlin: Springer. doi: https://doi.org/10.1007/978-3-662-44654-6_20
- American Cancer Society. (2019). Cancer.org. Obtenido de La densidad de los senos y el informe de su mamograma: <https://www.cancer.org/es/cancer/cancer-de-seno/pruebas-de-deteccion-y-deteccion-temprana-del-cancer-de-seno/mamogramas/la-densidad-de-los-senos-y-el-informe-de-su-mamograma.html>
- Araque Volk, N. (20 de marzo de 2019). Regularizando nuestra red: Dropout. Obtenido de MC.AI: <https://mc.ai/regularizando-nuestra-red-dropout/>
- Cruz-Roa, A., Gilmore, H., Basavanhally, A., Feldman, M., Ganesan, S., Shih, N., . . . Madabhushi, A. (18 de abril de 2017). Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent. *Scientific Reports*, 7(46450). Recuperado el 30 de octubre de 2019, de <https://doi.org/10.1038/srep46450>
- Chacón G, Rodríguez JE, Bermúdez V, Vera M, Hernández JD, Vargas S, Pardo A, Lamedo C, Madriz D, Bravo AJ. Computational assessment of stomach tumor volume from multi-slice computerized tomography images in presence of type 2 cancer. *F1000Res*. 2018 Jul 17;7:1098. doi: 10.12688/f1000research.14491.2. PMID: 30473775; PMCID: PMC6234734.
- Global Cancer Observatory. (2018). Colombia. World Health Organization, Switzerland. Geneva: Internacional Agency for Research

- on Cancer. Obtenido de <https://gco.iarc.fr/today/data/factsheets/populations/170-colombia-fact-sheets.pdf>
- Jalandhar, D., & Sharma, A. (2016). DDSM Utility v3.1. Obtenido de <https://github.com/trane293/DDSMTUtility/blob/mamast/Tutorial.pdf>
- Jean-Baptiste Lamy, Boomadevi Sekar, Gilles Guezennec, Jacques Bouaud, & Brigitte Séroussi. (2019). Explainable artificial intelligence for breast cancer: A visual case-based reasoning approach. *Artificial Intelligence in Medicine*, 94, 42-53. doi: <https://doi.org/10.1016/j.artmed.2019.01.001>
- Keras. (2020). Layer activation functions. Obtenido de Keras: <https://keras.io/api/layers/activations/>
- Kontos, K., & Maragoudakis, M. (2013). Breast Cancer Detection in Mammogram Medical Images with Data Mining Techniques. 9th Artificial Intelligence Applications and Innovations (AIAI). 412, págs. 336-347. Paphos: Springer. doi: https://dx.doi.org/10.1007/978-3-642-41142-7_35
- Pita Fernández, S., & Pértegas Díaz, S. (07 de diciembre de 2010). *Fisterra*. Recuperado el 03 de septiembre de 2019, de Pruebas diagnósticas: Sensibilidad y especificidad.: https://www.fisterra.com/mbe/investiga/pruebas_diagnosticas/pruebas_diagnosticas. Asp
- PLOS. (13 de marzo de 2019). Convolutional neural network for cell classification using microscope images of intracellular actin networks. Obtenido de PLOS ONE: <https://journals.plos.org/plosone/article/figure?id=10.1371/journal.pone.0213626.g002>
- Prabhu. (04 de marzo de 2018). Obtenido de <https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148>
- Renuka Joshi. (09 de septiembre de 2016). EXSILIO SOLUTIONS. Obtenido de Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures: <https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/>
- Rios, Y., García-Rodríguez, J., Sanchez, E., Alanis, A., Ruiz-Velázquez, E., & Pardo, A. (2020). Neuro-fuzzy control for artificial pancreas: In silico development and validation. [Control neuro-fuzzy para páncreas artificial: Desarrollo y validación in-silico] RIAI - Revista Iberoamericana De Automatica e Informatica Industrial, 17(4), 390-400. doi:10.7326/0003-4819-157-5-201209040-00508
- Rose, C. (2006). DDSM: Digital Database for Screening Mammography. Obtenido de University of South Florida: <http://www.eng.usf.edu/cvprg/Mammography/Database.html>
- Nolan, T. (14 de mayo de 2020). CBIS-DDSM. Obtenido de Cancer Imaging Archive: <https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM>
- Wang, J., Yang, X., Cai, H., Tan, W., Jin, C., & Li, L. (07 de junio de 2016). Discrimination of Breast Cancer with Microcalcifications on Mammography by Deep Learning. *Scientific Reports*, 6(27327). doi:10.1038/srep27327