



Síntesis de imagen médica postcontraste en estudios de DCE-MRI de mama usando aprendizaje profundo

Postcontrast Medical Image Synthesis in Breast DCE-MRI Using Deep Learning

  Sara Cañaverall¹;
 Carlos Mera-Banguero²;
 Rubén D. Fonnegra³

¹Instituto Tecnológico Metropolitano, Medellín-Colombia,
saracanaverall207005@correo.itm.edu.co

²Instituto Tecnológico Metropolitano, Universidad de Antioquia,
Medellín-Colombia
carlos.mera@udea.edu.co

³Institución Universitaria Pascual Bravo, Medellín-Colombia,
ruben.fonnegra@pascualbravo.edu.co

Cómo citar / How to cite

S. Cañaverall, C. Mera-Banguero, and R. D. Fonnegra, "Síntesis de imagen médica postcontraste en estudios de dce-mri de mama usando aprendizaje profundo," *Tecnológicas*, vol. 27, no. 60, e3052, 2024.
<https://doi.org/10.22430/22565337.3052>

Resumen

El cáncer de mama es una de las principales causas de muerte en mujeres en el mundo, por lo que su detección de forma temprana se ha convertido en una prioridad para salvar vidas. Para el diagnóstico de este tipo de cáncer existen técnicas como la imagen de resonancia magnética dinámica con realce de contraste (DCE-MRI, por sus siglas en inglés), la cual usa un agente de contraste para realzar las anomalías en el tejido de la mama, lo que mejora la detección y caracterización de posibles tumores. Como limitación, los estudios de DCE-MRI suelen tener un costo alto, hay poca disponibilidad de equipos para realizarlos, y en algunos casos los medios de contraste pueden generar efectos adversos por reacciones alérgicas. Considerando lo anterior, este trabajo tuvo como objetivo el uso de modelos de aprendizaje profundo para la generación de imágenes sintéticas postcontraste en estudios de DCE-MRI. La metodología consistió en el desarrollo de una función de costo denominada pérdida en las regiones con realce de contraste que aprovecha el comportamiento de la captación del agente de contraste. Como resultado se entrenaron dos nuevas arquitecturas de aprendizaje profundo, las cuales hemos denominado G-RiedGAN y D-RiedGAN, para la generación de imágenes postcontraste en estudios de DCE-MRI, a partir de imágenes precontraste. Finalmente, se concluye que las métricas proporción máxima señal ruido, índice de similitud estructural y error absoluto medio muestran que las arquitecturas propuestas mejoran el proceso de síntesis de las imágenes postcontraste preservando mayor similitud entre las imágenes sintéticas y las imágenes reales, esto en comparación con los modelos base en el estado del arte.

Palabras clave

Cáncer de mama, imagen médica, resonancia magnética, generación de imagen postcontraste, aprendizaje profundo.

Abstract

Breast cancer is one of the leading causes of death in women in the world, so its early detection has become a priority to save lives. For the diagnosis of this type of cancer, there are techniques such as dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI), which uses a contrast agent to enhance abnormalities in breast tissue, which improves the detection and characterization of possible tumors. As a limitation, DCE-MRI studies are usually expensive, there is little equipment available to perform them, and in some cases the contrast medium can generate adverse effects due to an allergic reaction. Considering all of the above, the aim of this work was to use deep learning models for the generation of postcontrast synthetic images in DCE-MRI studies. The proposed methodology consisted of the development of a cost function, called CeR-Loss, that takes advantage of the contrast agent uptake behavior. As a result, two new deep learning architectures were trained, which we have named G-RiedGAN and D-RiedGAN, for the generation of postcontrast images in DCE-MRI studies, from precontrast images. Finally, it is concluded that the peak signal-to-noise ratio, structured similarity indexing method, and mean absolute error metrics show that the proposed architectures improve the postcontrast image synthesis process, preserving greater similarity between the synthetic images and the real images, compared to the state-of-the-art base models.

Keywords

Breast cancer, diagnostic imaging, magnetic resonance imaging, postcontrast image generation, deep learning.

1. INTRODUCCIÓN

El cáncer de mama es una enfermedad crónica no transmisible y uno de los principales problemas de investigación en salud pública en el mundo debido a sus altas tasas de incidencia [1]. Esta enfermedad se debe a alteraciones a nivel del ADN, las cuales afectan la división y el crecimiento normal de las células del tejido. Este es uno de los cánceres que más muertes provoca en el mundo [2] y es uno de los de mayor incidencia, especialmente en la población femenina. Según el observatorio global del cáncer (GLOBOCAN), en el 2022 se registraron 11.5 % de casos nuevos de cáncer de mama, respecto del total de casos de cáncer reportados en el mundo [3].

En general, el cáncer de mama se puede clasificar en cinco tipos principales [4]. El primero es el carcinoma lobular in situ (CLIS), el cual es una condición benigna de la mama que se presenta en los lobulillos, pero no se disemina fuera de ellos. El segundo es el carcinoma ductal in situ (CDIS), que es un cáncer no invasivo que se manifiesta en glándulas productoras de leche o en los conductos, sin propagarse fuera de estos. El tercero es el carcinoma ductal infiltrante (CDI), que comienza en los conductos, pero luego se propaga a los tejidos circundantes de la mama. El cuarto es el carcinoma lobulillar infiltrante (CLI), que se origina en las células de los lobulillos mamarios y suele ser el más difícil de detectar porque se propaga de forma dispersa en lugar de formar un bulto definido. El quinto es el cáncer de mama inflamatorio, una forma rara y agresiva de cáncer de mama que, por lo general, no se presenta como un bulto sino como enrojecimiento, hinchazón y calor en la mama. Este último se propaga rápidamente por lo que requiere de tratamiento inmediato.

El tratamiento y el pronóstico dependen del tipo y de las características específicas del cáncer de mama. No obstante, la detección temprana y el tratamiento oportuno son esenciales evitar complicaciones, mejorar el pronóstico del paciente y reducir las tasas de mortalidad [5]. Entre los métodos más utilizados para la detección del cáncer de mama están la mamografía y el ultrasonido de mama. La mamografía consiste en aplicar una dosis baja de rayos X con el fin de visualizar los tejidos internos de la mama, y por su bajo costo y su facilidad de aplicación es considerado como el método estándar para la detección [6]. Por otro lado, el ultrasonido se diferencia del primero en que utiliza ondas de sonido para visualizar el tejido de la mama [7]. Sin embargo, ambos métodos tienen diferentes limitaciones. La mamografía es un examen que, por su dosis de radiación, solo se recomienda en pacientes mayores de 40 años y no se puede usar en mujeres embarazadas ni en seguimiento, mientras que el ultrasonido tiene una alta dependencia de la experticia de quien realiza el examen. Una limitación común es que ambos métodos suelen fallar cuando las mamas tienen una densidad tisular alta [6], [8].

Cuando la mamografía o el ultrasonido no son concluyentes, se usan exámenes más especializados que requieren la aplicación, vía intravenosa, de un compuesto químico denominado agente de contraste [9]. La función del agente de contraste es aprovechar el comportamiento metabólico elevado de las células cancerígenas para resaltarlas en el tejido mamario. Tres exámenes de este tipo son la mamografía digital con realce de contraste (CEDM, por sus siglas en inglés), la imagen de resonancia magnética dinámica con realce de contraste (DCE-MRI) y la ecografía con contraste (CEUS, por sus siglas en inglés). El primero consiste en dos tomas de rayos X de la mama, una antes de aplicar el agente de contraste y con baja radiación (mamografía convencional); y otra posterior a la aplicación de agente de contraste y de alta radiación. A partir de ambas imágenes se genera una nueva imagen en postproceso denominada imagen recombinada, en la cual se resaltan las posibles regiones con absorción del agente de contraste en el tejido [10].

En el caso de DCE-MRI se usan ondas magnéticas para captar la absorción del agente de contraste en el tiempo, el cual reacciona de manera acelerada en los tejidos con posibles lesiones [11]. El examen de DCE-MRI toma unas imágenes de ambos órganos antes de aplicar el medio de contraste y una serie de imágenes posteriores a la aplicación del contraste. Estos métodos son más sensibles a la detección debido a que proporcionan información más precisa sobre el estado de las mamas y las posibles lesiones, lo que permite la caracterización de posibles tumores. Para el caso de la CEUS, emplea un contraste intravascular que permite valorar en tiempo real la microcirculación y la perfusión vascular y tisular del tejido [12].

A pesar de sus ventajas, estos exámenes están limitados por su costo, por el prolongado tiempo de adquisición y por la escasez de equipos para su realización. Adicionalmente, el agente de contraste puede provocar reacciones alérgicas o adversas en los pacientes [13].

Entre la búsqueda de soluciones a estas limitaciones, múltiples trabajos han mostrado el potencial uso de redes neuronales de aprendizaje profundo para la generación de imágenes sintéticas que pueden replicar el efecto del agente de contraste, sin la utilización de este [14]-[18], a su vez que preservan la calidad visual de las imágenes.

En este trabajo se propuso como contribución una función de costo denominada CeR-Loss, la cual aprovecha el comportamiento de la captación del agente de contraste para generar imágenes sintéticas postcontraste a partir de imágenes precontraste en estudios de DCE-MRI. Esta función se usa en dos nuevas arquitecturas de aprendizaje profundo, que hemos denominado G-RiedGAN y D-RiedGAN, las cuales centran su atención en las regiones de realce de contraste para mejorar el proceso de generación de las imágenes sintéticas postcontraste.

2. REVISIÓN DE LITERATURA

El aprendizaje profundo es una rama del aprendizaje de máquinas basado en el uso de redes neuronales artificiales. Estas últimas se crean a partir de la interconexión de capas de neuronas artificiales que tienen la capacidad de auto ajustarse en función de la entrada y la cantidad de datos que procesan [19]. Una de las áreas de aplicación de las redes de aprendizaje profundo es la síntesis de imágenes. Esta consiste en la generación de imágenes artificiales a partir de una descripción visual o textual de su contenido. En el campo de la medicina, la síntesis de imágenes se ha utilizado para múltiples propósitos; por ejemplo, para aumentar los conjuntos de datos que entrenan modelos para ayuda al diagnóstico de enfermedades, como el cáncer [20]; para mejorar la resolución de las imágenes obtenidas en ciertas modalidades de imágenes [21]; para segmentar regiones de interés en la imagen [22]; y para obtener la imagen de una modalidad de examen a partir de los datos de otra modalidad [23], entre otras aplicaciones. La importancia del uso de la síntesis de imágenes en el campo médico radica en su potencial, no solo para ayudar a mejorar la precisión en el diagnóstico de ciertas enfermedades, sino también en reducir el tiempo y costo de captura en ciertas modalidades de imágenes diagnósticas, lo que puede permitir ampliar la cobertura de este tipo de exámenes médicos [24].

Los modelos para la síntesis de imágenes se pueden agrupar, de manera general, en dos: los *autoencoders* y las redes generativas adversarias (GAN, por sus siglas en inglés). Los *autoencoders* son arquitecturas que tienen dos partes, un codificador cuyo objetivo es reducir la dimensionalidad de los datos de entrada para aprender la representación abstracta (o latente) de su distribución; y un decodificador que reconstruye la información desde el espacio latente a un espacio de mayor dimensión [25]. Una variación de los *autoencoders* es una arquitectura denominada U-Net, la cual busca subsanar el problema de pérdida de información de los *autoencoders*, en parte copiando, la

información de las capas del codificador a las capas del decodificador lo que permite que este mejore la reconstrucción de información del espacio de mayor dimensión [25], [26].

Las arquitecturas GAN se componen de un generador y un discriminador. Por un lado, el generador es una red convolucional que intenta aprender la distribución latente de los datos reales para generar información sintética a partir de una muestra de ruido aleatorio. Por otro lado, el discriminador es una red convolucional complementaria que se comporta como un experto que diferencia entre la información real y la sintética. El entrenamiento de ambas redes se realiza de manera adversaria, así, el generador busca mejorar el proceso de generación para engañar al discriminador, mientras que este último busca mejorar su experticia para no dejarse engañar por el generador. Este aprendizaje adversario es lo que le otorga el nombre a esta arquitectura [27].

Ambos tipos de arquitecturas generativas se han utilizado en el ámbito médico y en el contexto de ayuda al diagnóstico de cáncer de mama, los métodos de síntesis de imagen se pueden usar para generar imágenes postcontraste a partir de imágenes precontraste en estudios de DCE-MRI y CEDM. A esta aplicación se le conoce como cambio de dominio y consiste en transformar una imagen x , que es la imagen precontraste, a otra imagen similar y , que es la imagen postcontraste [21], [23]. Entre los trabajos de esta línea de investigación en [15] describen una arquitectura generativa poco profunda, denominada SD-CNN, la cual se entrena para generar parches sintéticos de imágenes recombinadas postcontraste en estudios de CEDM a partir de parches de imágenes de mamografías digitales de campo completo. En este estudio se utilizaron dos bases de datos de imágenes independientes para extraer los parches. La primera es una base de datos pública, denominada INbreast [28], de la cual usaron 89 estudios con clasificación BIRADS 1,2 (como benignos), y 5 y 6 (como malignos). La otra base de datos utilizada es privada y consiste en 49 estudios con clasificación BIRADS 4 o 5, confirmada por biopsia, de los cuales 23 son benignos y 26 malignos. Los autores indican que el uso de los parches sintéticos generados con su arquitectura mejora la precisión de un clasificador de parches de mamografía. Una de las limitaciones de esta arquitectura es que solo permite generar parches sintéticos de 3x3 píxeles a partir de parches de 15x15 píxeles, lo que restringe su uso en la síntesis de imágenes completas, además la poca profundidad de la red conlleva a que haya pocos niveles de abstracción, lo que puede reducir la capacidad para sintetizar estructuras complejas como las que definen los tejidos de la mama.

Dando continuación al trabajo anterior, en [29] propusieron una arquitectura tipo U-Net denominada RiedNet. Entre las modificaciones que propusieron los autores a la arquitectura U-Net de base, se introducen capas convolucionales y deconvolucionales que reemplazan las capas tipo *pooling*, además de la agregación de un tipo de bloque llamado *residual inception*, el cual busca subsanar el problema del desvanecimiento del gradiente por la profundidad de la red. La arquitectura fue diseñada para la generación sintética de imágenes en el contexto médico y se evalúa, en parte en la generación de imágenes recombinadas postcontraste de estudios de CEDM a partir de imágenes de baja energía. Como limitación, la red fue entrenada para sintetizar bloques de 128x128 píxeles por lo que la síntesis de imagen completa considera el promedio de los bloques generados, lo que puede causar un efecto borroso en los tejidos reconstruidos de la mama. En este caso específico, los experimentos se realizaron sobre 139 estudios de mamografía contrastada, de los cuales 112 fueron usados para entrenar la arquitectura y 27 para prueba.

En cuanto al uso de estudios de DCE-MRI para la detección de cáncer de mama, los autores en [30] utilizan una arquitectura tipo GAN condicional, denominada Pix2Pix [31], que tiene como objetivo generar imágenes contrastadas ponderadas en T1, con supresión de grasa, a partir de imágenes sin contraste. En general, la arquitectura Pix2Pix está compuesta por un generador tipo U-Net y un discriminador tipo PatchGAN [31]. Las imágenes usadas

en este trabajo fueron adquiridas con una resolución de 3T, aunque fueron redimensionadas a un tamaño de 512 x 512 píxeles. En total se usaron 2620 pares de imágenes de 48 estudios de DCE-MRI. De estas, 2112 imágenes se destinaron para el entrenamiento de la arquitectura, 418 para validación y 90 para pruebas. De acuerdo con los autores, el modelo suele presentar errores cuando se usa en imágenes de mamas densas; sin embargo, resaltan el potencial de la arquitectura Pix2Pix en la generación sintética de imágenes contrastadas de DCE-MRI. Otro trabajo en esta línea es el presentado por [17] en el que se propone una arquitectura tipo GAN, denominada TSGAN, la cual se compone de cuatro modelos, uno tipo U-Net y que es entrenada para generar imágenes ponderadas en T1 postcontraste, a partir de imágenes precontraste; dos discriminadores: uno global que se centra en diferenciar entre imágenes postcontraste reales y falsas, y otro local que se encarga de diferenciar entre regiones de interés reales y falsas; y finalmente, un modelo tipo U-Net que se entrena para generar máscaras de segmentación sobre las lesiones presentes en la mama.

Igualmente, en [32] los autores proponen una arquitectura denominada síntesis de realce de bordes (EDLS, por sus siglas en inglés) para sintetizar secuencias dinámicas a partir de imágenes T1WI en estudios de MRI. Los autores indican que la arquitectura propuesta mejora la identificación de lesiones sin el uso del agente de contraste. En [33] los autores investigan el uso de arquitecturas GAN para generar imágenes de MRI de mama realistas, esto con el fin de mejorar la detección de lesiones en las mamas. Por otro lado, en [34] se propone el uso de una arquitectura CycleGAN para traducir imágenes entre diferentes dominios sin necesidad de pares de datos emparejados, mejorando la calidad de las imágenes sintetizadas.

En [35] propusieron una arquitectura denominada TDM-StarGAN que se emplea para generar imágenes sintéticas de las fases de estudios DCE-MRI convencionales a partir de imágenes estudios de DCE-MRI ultrarrápidos. En este trabajo los autores modificaron la arquitectura StarGAN [29] para utilizarla con imágenes emparejadas, además de considerar la pérdida entre los mapas de diferencia de las imágenes generadas y la zona de captación, obtenida a partir de la diferencia entre la última imagen postcontraste y la imagen precontraste. Como conclusión, los autores indican que el modelo propuesto mejora los modelos de base usados (Pix2Pix y StarGAN) dado en que este, se enfoca en sintetizar correctamente las regiones asociadas a las lesiones.

Respecto a la síntesis en otras modalidades de imagen, en [36] propusieron el uso de una arquitectura Pix2Pix para mejorar la calidad de imágenes de tomografía por emisión de positrones de recuento bajo (dbPET, por sus siglas en inglés), la cual se ve afectada por la respiración de los pacientes. Los experimentos se realizaron con 49 estudios que incluyen 32 con resultados anormales y 17 con resultados normales. Cada imagen se redimensionó a 958 x 940 píxeles. De acuerdo con los autores, el uso de la arquitectura Pix2Pix puede ser usada para abordar este problema mejorando la calidad de imágenes dbPET adquiridas con tiempos cortos de adquisición. Con miras en reducir las dosis de radiación usadas en los exámenes para el diagnóstico de cáncer de mama en [18], [37] han desarrollado varios trabajos en los que se han entrenado diferentes modelos para generar imágenes de mamografía digital, a partir de imágenes de tomosíntesis. En sus trabajos, los autores presentaron una arquitectura denominada GGGAN que usa un generador tipo U-Net y un discriminador, que es una variante del discriminador Pix2PixHD, que en su función de pérdida usa los mapas de diferencia entre los gradientes de las imágenes generadas en ciertas capas intermedias.

En un contexto diferente al diagnóstico de cáncer de mama, en [38] los autores diseñaron una arquitectura basada en Pix2Pix, denominada Ea-GAN, la cual considera la inclusión de los mapas de diferencias de los bordes entre las imágenes generadas y las imágenes reales, tanto en la función de pérdida del discriminador como del generador, a fin de mejorar la

síntesis de estos elementos y reducir el efecto de suavizado que suele producirse en las redes tipo U-Net [14]. En este trabajo los autores usaron una base de datos denominada BRATS2015 [39] que contienen imágenes de MRI de 74 pacientes con resolución 240x240x155 (voxels), las imágenes fueron preprocesadas y normalizadas con valores de intensidad en el rango [-1, 1].

A pesar de que existen diferentes trabajos que proponen modelos generativos para la síntesis de imágenes diagnósticas para la detección del cáncer de mama, aún se evidencian algunas limitaciones. Esto se debe en cierta medida a que las imágenes de mama poseen una mayor variabilidad en densidad de tejido, lo cual implica que, al trabajar con los agentes de contraste mediante los modelos generativos, existan algunas restricciones debido a que la visualización de dicho agente de contraste disminuye en función de la intensidad de los píxeles.

En la contribución de soluciones al problema de generación sintética de imágenes diagnósticas para la detección del cáncer de mama, en este trabajo se propone una arquitectura denominada D-RiedGAN que sigue el enfoque de una arquitectura Pix2Pix con bloques de inyección residual, pero que centra su atención en las regiones de realce de contraste en estudios de DCE-MRI.

3. METODOLOGÍA

El desarrollo metodológico de este trabajo parte de la implementación de una línea base de tres modelos para sintetizar imágenes de respuesta temprana al medio de contraste, ponderadas en T1 con saturación de grasa, en estudios de DCE-MRI. A partir de la línea base se crean dos arquitecturas ensambladas y se propone dos nuevas arquitecturas, denominadas G-RiedGAN y D-RiedGAN. En general, los modelos generativos utilizados en este trabajo se entrenan para que, a partir de una imagen sin contraste x , se generen imágenes sintéticas postcontraste $\hat{y} = G(x)$. El objetivo es que el modelo generador $G(x)$ aprenda a generar la respuesta temprana al medio de contraste y , por tanto, hacer que las imágenes sintéticas sean similares a las imágenes reales postcontraste y .

3.1 Optimización de regiones con realce de contraste

Los modelos convencionales para la síntesis de imágenes han mostrado importantes avances en el área de procesamiento de imágenes naturales. Sin embargo, también han mostrado múltiples limitaciones, especialmente al tratarse de imágenes especializadas como las imágenes médicas. Para solventar estas limitaciones, este trabajo propone un esquema que integra una función de costo que induce la información de las regiones de realce de contraste durante el entrenamiento. El objetivo de esta función es orientar el proceso de síntesis a la correcta generación del realce de contraste en las imágenes postcontraste.

En términos de intensidad de los píxeles, el realce producido por el agente de contraste se puede identificar a partir de las intensidades más altas en la imagen postcontraste. Específicamente, para detectar los píxeles con mayores niveles de intensidad en la imagen postcontraste se utiliza una estrategia de umbralización global, tal y como se muestra en (1), en donde $y(i, j)$ es el píxel en la posición (i, j) en la imagen postcontraste y T es el valor del umbral.

$$F_y(i, j) = \begin{cases} 1, & \text{si } y(i, j) > T \\ 0, & \text{en otro caso} \end{cases} \quad (1)$$

Por la sensibilidad del parámetro T , respecto a las variaciones de intensidades en las imágenes de diferentes estudios de DCE-MRI, este parámetro se fija para cada imagen utilizando el percentil 90 de su histograma. Es decir, se mantienen como regiones de realce de contraste el 10 % de los píxeles de la imagen con mayor intensidad. Después, se emplean las operaciones morfológicas de cierre y apertura para suavizar los contornos y eliminar pequeñas brechas entre regiones adyacentes. Ambas operaciones se aplican utilizando un elemento estructurante con forma circular de 7×7 píxeles. Este algoritmo se aplica tanto a la imagen sintética postcontraste, como a la imagen real postcontraste, y se les denominan máscaras de realce de contraste real (F_y) y sintetizada ($F_{G(x)}$).

Una vez estimadas las regiones con realce de contraste, se propone una función de costo con el fin de minimizar las diferencias entre esas regiones de las imágenes reales y de las sintetizadas. Dado que las regiones de realce de contraste son de tipo binario, optimizarlas implica emplear una función de costo basada en similitud de conjuntos, como lo es el índice de Jaccard [40]. Específicamente, el índice de Jaccard entre las regiones de contraste real (F_y) y generada ($F_{G(x)}$) se calcula usando (2).

$$Jc(F_y, F_{G(x)}) = \frac{|F_y \cap F_{G(x)}|}{|F_y| + |F_{G(x)}| - |F_y \cap F_{G(x)}|} \quad (2)$$

Como el índice de Jaccard no es una función convexa ni derivable, su optimización usando el algoritmo de gradiente descendente, en el contexto de las redes neuronales, puede llevar a soluciones subóptimas o incluso a errores de convergencia. De acuerdo con [41], es posible aplicar una aproximación convexa con el fin de obtener una función soluble a partir de la función discreta, que a su vez es optimizable por métodos de primer orden como el gradiente descendente. Esta extensión está basada en la aproximación subrogada de Lovasz, y se estima a partir de un conjunto de predicciones erróneas $m(c) \in \mathbb{R}^p$ para la clase $c \in \mathcal{C}$, en vez del vector de predicciones discretas $\{0, 1\}^p$. Considerando que una función $\Delta: \{0, 1\}^p \mapsto \mathbb{R}^p$ es submodular (como la función Jc), su extensión surrogada de Lovasz está definida por (3).

$$\bar{\Delta}: m \in \mathbb{R}^p \mapsto \sum_{i=1}^p m_i g_i(m) \quad (3)$$

Donde, $g_i(m) = \Delta(\{\pi_1 \dots \pi_i\}) - \Delta(\{\pi_1 \dots \pi_{i-1}\})$ siendo π una permutación en orden de los componentes de m en orden descendente. Naturalmente, $\bar{\Delta}$ es la función cerrada estricta de Δ , es lineal por partes e interpola los valores de Δ en \mathbb{R}^p . Finalmente, para calcular la función surrogada de Lovasz del índice de Jaccard en (2) ($\bar{\Delta}_{Jc}$) se utiliza (4).

$$\text{Lovasz}(f(y, y^*)) = \frac{1}{|\mathcal{C}|} \sum_{i=1}^c \bar{\Delta}_{Jc}(m(c)) \quad (4)$$

En donde $f(y, y^*)$ corresponde a la función que permite estimar el vector de errores m a partir de las máscaras de contraste real o generada después de aplicar la función *softmax*. Por último, con el fin de evitar las variaciones por dependencia del tamaño del *batch* y el número de clases, se busca optimizar la función *Lovasz* a través de la combinación de esta con la entropía cruzada binaria (*BCE*, por sus siglas en inglés) descrita en (5), tal como lo sugieren los autores en [41].

$$BCE(F_y, F_{G(x)}) = -\frac{1}{C} \sum_{i=1}^C (F_y \cdot \log(F_{G(x)}) + (1 - F_y) \cdot \log(1 - F_{G(x)})) \quad (5)$$

Finalmente, para optimizar los modelos propuestos se usa una función de costo que combina la BCE y la extensión subrogada de Lovasz sobre las regiones de realce de contraste, cuya tarea es optimizar la síntesis de esas regiones. Esta función se presenta en (6) y se ha denominado CeR-Loss.

$$\text{CeR-Loss}(F_y, F_{G(x)}) = (1 - \lambda_{\text{Lovasz}})BCE(F_y, F_{G(x)}) + \lambda_{\text{Lovasz}}\text{Lovasz}(f(F_y, F_{G(x)})) \quad (6)$$

3.2 G-RiedGAN y D-RiedGAN

Con base en los elementos antes descritos, la Figura 1a muestra la arquitectura general del primer modelo propuesto, denominado G-RiedGAN. Esta arquitectura integra, después del generador, el filtro para la detección de las regiones de realce del contraste con el que se busca retroalimentar al generador y guiar su aprendizaje para que replique el realce de contraste. En este caso se mantiene inmutable el discriminador PatchGAN, cuya tarea es identificar si el par de imágenes precontraste y postcontraste son reales o sintetizadas. La (7) presenta la función de pérdida del generador de la G-RiedGAN. Esta función considera la pérdida general causada por la diferencia a nivel de píxeles entre la imagen real y la imagen generada y también la pérdida generada entre las regiones de realce de contraste (CeR-Loss) de esas imágenes.

$$L_{G\text{-RiedGAN}}^G = E_x \left[\log(1 - D(x, G(x))) \right] + \lambda_{l1} E_{x,y} [||y - G(x)||] + \text{CeR-Loss}(F_y, F_{G(x)}) \quad (7)$$

Por otra parte, denominamos D-RiedGAN a la arquitectura que incluye la diferencia entre las regiones de realce de contraste tanto en el generador como en el discriminador. Esto permite que el generador incremente su atención en esas regiones, al considerarlas en la contraparte adversaria, lo que ayuda a mejorar la calidad de las síntesis del modelo. La arquitectura D-RiedGAN se presenta en la Figura 1b.

Para realizar el aprendizaje adversario, el discriminador de la D-RiedGAN se modifica para recibir una tripleta de imágenes: la imagen de entrada, la imagen sintética o real y las regiones de realce de contraste de la imagen real o de la imagen sintética. Así, (8) y (9) definen las funciones de pérdida del generador y del discriminador de la D-RiedGAN, respectivamente, las cuales incluyen la pérdida de las regiones de realce de contraste (CeR-Loss).

$$L_{D\text{-RiedGAN}}^G = E_x \left[\log(1 - D(x, G(x), F(G(x)))) \right] + \lambda_{l1} E_{x,y} [||y - G(x)||] + \text{CeR-Loss}(F_y, F_{G(x)}) \quad (8)$$

$$L_{D\text{-RiedGAN}}^D = -E_{x,y} \left[\log(D(x, y, F(y))) \right] - E_x \left[\log(1 - D(x, G(x), F(G(x)))) \right] \quad (9)$$

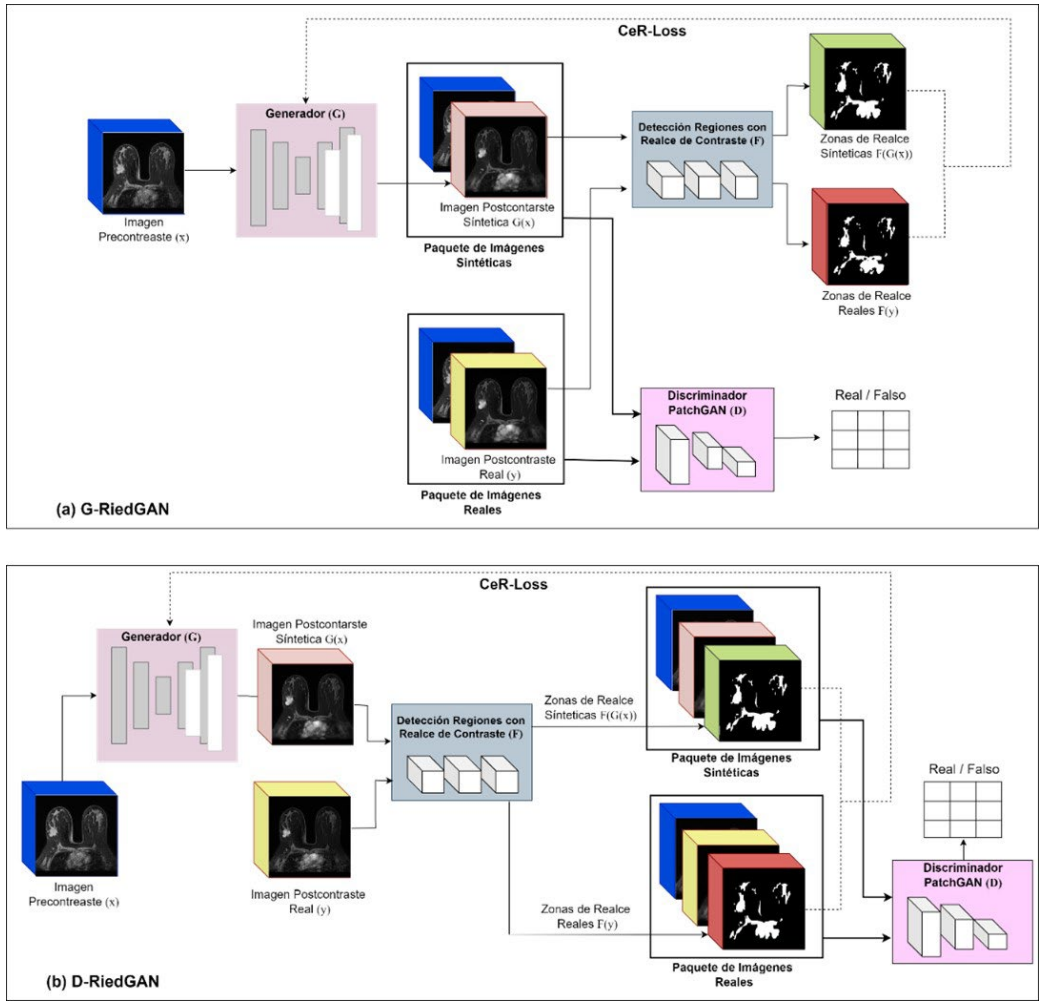


Figura 1. Arquitecturas G-RiedGAN y D-RiedGAN. Fuente: elaboración propia.

3.3 Modelos Base

Como modelos de línea base para la evaluación de nuestro modelo se usaron las arquitecturas Pix2Pix [31], RiedNet [29] y Ea-GAN [38]. La selección de estas arquitecturas obedece a los resultados que se han reportado en su uso en problemas de síntesis de imágenes médicas en distintas modalidades.

3.3.1 Pix2Pix

Pix2Pix [31] es una arquitectura que utiliza información condicional para guiar el proceso de generación, lo que significa que genera una imagen de un dominio a partir de una imagen de entrada que pertenece a otro dominio. La arquitectura Pix2Pix consta de un generador tipo U-Net, un discriminador PatchGAN y una función objetivo que combina la pérdida adversaria que ayuda a que las imágenes generadas traten de ser indistinguibles de las reales, con una pérdida al nivel de los píxeles, que busca que haya una coherencia en el contenido entre la imagen generada y la imagen real. Una de las ventajas discutidas en la literatura de esta arquitectura es que preserva los detalles finos de las imágenes en el proceso de generación, lo que es crucial para la generación de imágenes postcontraste.

3.3.2 RiedNet

RiedNet [29] es una arquitectura tipo U-Net, modificada para usar capas convolucionales y deconvolucionales. Esta arquitectura también incluye un bloque llamado *residual inception* que, en conjunto, buscan limitar el problema del desvanecimiento del gradiente. A diferencia de la propuesta original, en este trabajo la arquitectura RiedNet se entrenó para sintetizar imágenes completas, también se reemplazó la función de activación ReLU de las capas intermedias por la función Leaky ReLU, la cual permite que una pequeña pendiente positiva esté activa, evitando así la completa supresión de la información en esa parte de la red neuronal [42]. La función de activación de la capa de salida se cambió por la función tangente hiperbólica. Estos cambios se implementaron para mantener el rango de valores de las imágenes de entrada a medida que se codifica la información.

3.3.3 Ea-GAN

A diferencia de las dos arquitecturas anteriores, Ea-GAN [38] incluye en el proceso de aprendizaje los bordes de los objetos tanto de la imagen original, como de la imagen sintetizada, calculados a partir de un filtro Sobel. De acuerdo con los autores, incluir la información de los bordes en el proceso de generación ayuda a la arquitectura a centrarse en la síntesis de las texturas y los bordes de los objetos en las imágenes. La arquitectura Ea_GAN tiene dos variaciones; la gEa-GAN, la cual introduce la diferencia entre los bordes de las imágenes sólo en la función de pérdida del generador, y la dEa-GAN la cuál además incluye la diferencia de los bordes de las imágenes en el discriminador.

Considerando estas arquitecturas de base, y partiendo de la premisa que su fusión puede mejorar el proceso de síntesis, se procedió a realizar un ensamble entre ellas. El primer ensamble, al que hemos denominado RiedGAN, integra a la arquitectura RiedNet un discriminador tipo PatchGAN, esto con el fin de mejorar el proceso de síntesis utilizando un esquema de aprendizaje adversario. Sustancialmente, el cambio principal en esta red, con respecto a la Pix2Pix original, es la estructura del generador, el cual en lugar de usar un generador tipo U-Net tradicional, usa el generador tipo U-Net de la arquitectura RiedNet.

Partiendo de la idea del uso de los mapas de bordes de la arquitectura Ea_GAN, también se integró el uso de los mapas de bordes a la arquitectura RiedGAN. Este ensamble generó dos modelos, el primero denominado gEa-RiedGAN, en el que se integran los mapas de bordes en el generador de la RiedGAN, y el segundo llamado dEa-RiedGAN, el cual integra la información de los mapas de bordes tanto en el generador como en el discriminador.

3.4 Métricas de evaluación

Para validar la calidad de las imágenes sintéticas generadas se utilizaron las métricas cuantitativas más comunes en la literatura: Error absoluto medio (MAE, por sus siglas en inglés), proporción máxima señal ruido (PSNR, por sus siglas en inglés) y el índice de similitud estructural (SSIM, por sus siglas en inglés). El MAE estima la diferencia píxel a píxel entre las intensidades de un par de imágenes calculando el promedio entre esas diferencias. Así, para una imagen real y , y una imagen generada $G(x)$, ambas de tamaño $m \times n$ píxeles, el MAE se calcula como indica (10). En este caso, un MAE bajo indica un error menor entre la imagen sintetizada y la imagen de referencia. Valores cercanos a 0 son ideales, indicando una alta precisión. Por otro lado, un MAE alto indica un mayor error entre la imagen sintetizada y la imagen de referencia. Esto sugiere una baja precisión en la síntesis de la imagen.

$$MAE(y, G(x)) = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n |y_{(i,j)} - G(x)_{(i,j)}| \quad (10)$$

El PSNR es una métrica empleada para definir la relación entre la máxima energía que emite una señal de cualquier tipo, y el ruido que afecta la representación de dicha señal, siendo su unidad de medida los decibeles (dB) [43]. Entre mayor sea el valor del MAE, mayor es la similitud que hay entre las dos imágenes. En (11) se presenta la formulación del PSNR, en donde el término MAX_i es el valor de intensidad máximo posible para las imágenes. Un PSNR alto indica una mayor similitud entre la imagen sintetizada y la imagen de referencia, mientras valor bajo indica una mayor diferencia entre la imagen sintetizada y la imagen de referencia.

$$PSNR(y, G(x)) = 10 * \log_{10} \left(\frac{MAX_i^2}{\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n |y_{(i,j)} - G(x)_{(i,j)}|^2} \right) \quad (11)$$

El SSIM considera que los píxeles tienen fuertes interdependencias, especialmente cuando están cerca. Estas dependencias contienen información sobre luminancia, contraste y estructura de los objetos en la imagen y se pueden estimar de manera conjunta como indica (12) [43], en donde μ , σ y σ^2 son, respectivamente, las medias, desviaciones estándar y covarianzas entre las imágenes, y c_1 y c_2 son dos variables que estabilizan la división para denominadores cercanos a cero. Para este caso, un SSIM cercano a 1 indica una alta similitud estructural entre la imagen sintetizada y la imagen de referencia. Un SSIM bajo indica una menor similitud estructural. Valores por debajo de 0.4 sugieren una baja calidad en términos de estructura y textura de la imagen.

$$SSIM(y, G(x)) = \frac{(2\mu_{G(x)}\mu_y + c_1)(2\sigma_{G(x),y} + c_2)}{(\mu_{G(x)}^2 + \mu_y^2 + c_1)(\sigma_{G(x)}^2 + \sigma_y^2 + c_2)} \quad (12)$$

Por último, los mapas de diferencia se calculan entre una imagen generada y una imagen real, donde se comparan píxeles individuales para evaluar la discrepancia entre ellos. Para calcular estos mapas, se empleó (13). Esta fórmula describe un proceso en el que se analizó cada píxel de las imágenes, y se calculó la diferencia en intensidad entre el mismo píxel en otra imagen. Cada píxel en una imagen tiene un valor que representa su intensidad, la comparación de píxeles implica restar el valor del píxel correspondiente en una imagen del valor del mismo píxel en otra imagen. El propósito de esta comparación es cuantificar y visualizar las diferencias entre las imágenes.

$$DF = ||y_{(i,j)} - G(x)_{(i,j)}|| \quad (13)$$

4. RESULTADOS Y DISCUSIÓN

4.1 Configuración experimental

Los resultados reportados en este trabajo fueron obtenidos bajo la configuración experimental descrita en la Tabla 1. Esta tabla muestra los valores de los hiperparámetros usados en los diferentes modelos y ajustados según las capacidades de cómputo disponibles. Los experimentos se ejecutaron en una estación de trabajo con una CPU Intel Xeon Silver 4108 y una GPU NVIDIA Quadro P2000, con memoria 4GB. Se emplea lenguaje de programación Python versión 3.8 con una versión Pytorch 2.0.

Tabla 1. Hiperparámetros fijados en los experimentos. Fuente: elaboración propia.

Modelo	Tamaño de Batch	Número de épocas	Optimizador	Taza de aprendizaje	Valor λ_{L1}	Valor λ_{edge} o λ_{Iovasz}	Activación Salida
RiedNet	4	100	Adam	0.0002	N/A	-	Linear
Pix2Pix	1	100	Adam	0.0002	100	-	TanH
gEa-GAN	1	100	Adam	0.0002	300	300	Sigmoide
dEa-GAN	1	100	Adam	0.0002	300	300	Sigmoide
RiedGAN	4	100	Adam	0.0002	100	-	TanH
gEa-RiedGAN	1	100	Adam	0.0002	100	300	TanH
dEa-RiedGAN	1	100	Adam	0.0002	100	300	TanH
G-RiedGAN	1	100	Adam	0.0002	100	150	TanH
D-RiedGAN	1	100	Adam	0.0002	100	150	TanH

4.2 Base de datos

Para el entrenamiento de los modelos se utilizó una base de datos privada, retrospectiva, y anonimizada de estudios de DCE-MRI de 197 pacientes. Cada estudio contiene imágenes estructurales potenciadas en T1 y T2, imágenes de difusión (DWI), y de DCE (6 dinámicas). Para este trabajo se tomó la respuesta en la secuencia T1 con saturación de grasa antes de la aplicación de medio de contraste (x) y esta misma imagen con la aplicación posterior del medio de contraste en etapa temprana (y). Debido a la naturaleza retrospectiva de la base de datos, se seleccionaron estudios con distintos tipos de resonadores de 1.5T, con compuestos basados en gadolinio, y con dosis entre 0.014 y 0.016 ml/mol. Estos estudios también cuentan con al menos una anomalía (benigna o maligna) anotada por expertos radiólogos usando el sistema BIRADS. La selección de los estudios se hizo manteniendo un número balanceado entre los estudios benignos y malignos.

Dado que este trabajo se centra en la síntesis de las regiones de contraste, se seleccionaron aquellas imágenes en las cuales se anotan las regiones con el fin de asegurar el comportamiento de la captación. Como resultado, se obtienen 937 imágenes normalizadas en el rango $[-1, 1]$, de las cuales 718 se usaron para el entrenamiento y 219 para la validación. Las imágenes, que originalmente tenían resoluciones desde 480x480 píxeles hasta 512x512 píxeles fueron todas redimensionadas a un tamaño de 240x240 píxeles.

4.3 Evaluación comparativa

La Figura 2 presenta una comparación gráfica del PSNR, SSIM y MAE obtenidos por los modelos sobre el conjunto de imágenes de validación. El gráfico muestra que los modelos G-RiedGAN y D-RiedGAN, propuestos en este trabajo, son los que obtienen el mejor desempeño.

Esto indica que la estrategia de considerar las regiones de realce de contraste en el proceso de síntesis de las imágenes, a través de la función CeR-Loss, mejora la calidad de las imágenes sintéticas en función de las métricas cuantitativas.

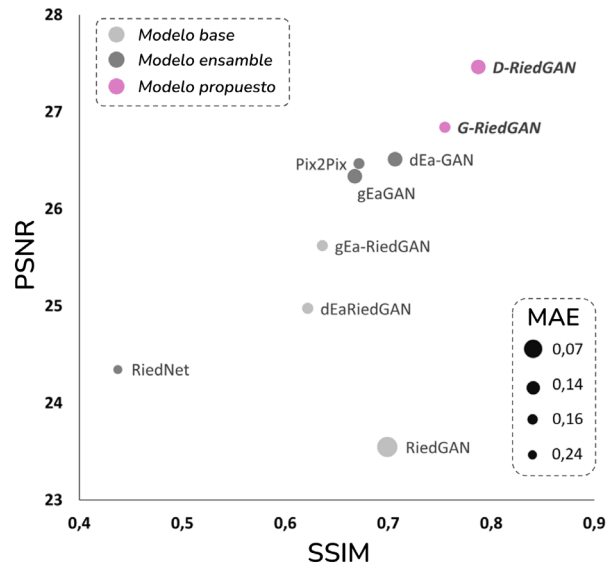


Figura 2. Gráfico de dispersión para el PSNR, SSIM y MAE en los modelos base, ensamblados y propuestos. Fuente: elaboración propia.

Aunque G-RiedGAN y D-RiedGAN presentan un desempeño ligeramente inferior en MAE comparado con RiedGAN, la reducción en el MAE para D-RiedGAN es marginal y se ve compensada por mejoras significativas en PSNR y SSIM. Esto sugiere que, aunque RiedGAN tiene una precisión ligeramente mayor en el promedio de los errores de píxeles individuales, genera imágenes más borrosas en las estructuras internas, lo cual es menos adecuado para la síntesis de imágenes médicas.

En comparación con Pix2Pix [28], un modelo ampliamente utilizado en estudios similares, los modelos G-RiedGAN y D-RiedGAN demuestran un mejor desempeño en PSNR y SSIM, especialmente en la síntesis de regiones de realce de contraste y en la reducción del ruido, superando así las limitaciones de Pix2Pix en mamas densas.

Por otro lado, la Figura 3 muestra algunas imágenes postcontraste reales y generadas a partir de su contraparte sin contraste. En general, es posible evidenciar que los modelos logran generar de manera eficiente las estructuras anatómicas de mayor tamaño, salvo algunas diferencias notorias entre las intensidades de las estructuras de la caja torácica. A pesar de esto, es notable que los modelos G-RiedGAN y D-RiedGAN sintetizan las regiones de realce de contraste de manera superior en comparación con modelos de referencia como RiedNet [12], Pix2Pix [28] y EaGAN [32]. Estos modelos sirvieron como base para el desarrollo de G-RiedGAN y D-RiedGAN y fueron empleados para realizar comparaciones siguiendo la misma metodología. Si bien las redes mencionadas son efectivas en la generación de imágenes en sus respectivas investigaciones, uno de los propósitos de este estudio era evaluar su desempeño específico con este tipo de imágenes médicas. A partir de los resultados obtenidos, se propusieron las redes G-RiedGAN y D-RiedGAN, las cuales integran las mejores características de los modelos preexistentes, logrando una síntesis de imágenes con mayor precisión y calidad en las regiones de realce de contraste. Además, evidencian un buen desempeño en comparación con los modelos del estado del arte, especialmente en términos del ruido y las borrosidades generadas. Esto puede apreciarse de mejor manera en la Figura

4, la cual muestra los mapas de diferencias entre las imágenes sintéticas y las imágenes reales. Como se puede observar, los mapas de diferencia asociados a los modelos propuestos (G-RiedGAN y D-RiedGAN) son los que tienen la menor diferencia entre las imágenes sintéticas y las imágenes reales.

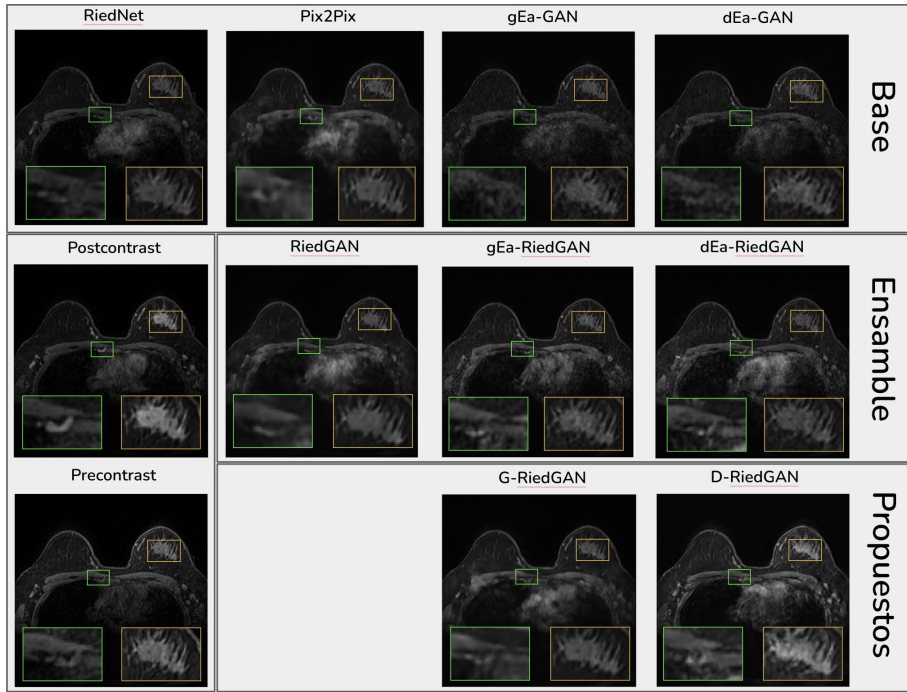


Figura 3. Resultados comparativos entre los modelos evaluados. Las regiones aumentadas corresponden con proyecciones en donde se presenta captación del agente de contraste. Fuente: elaboración propia.

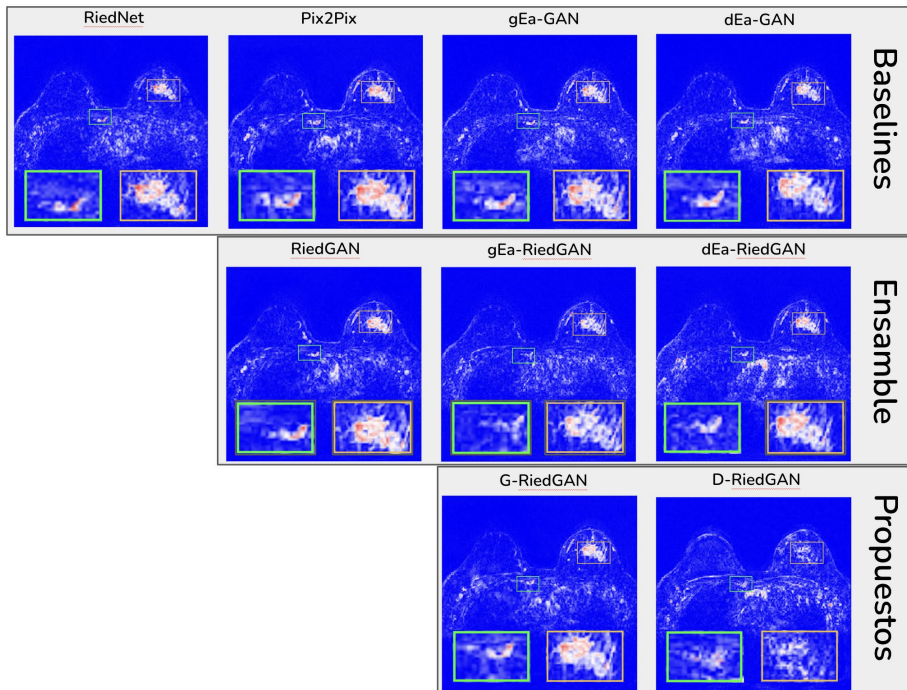


Figura 4. Mapas de diferencia entre las imágenes reales y generadas por los modelos evaluados. Fuente: elaboración propia.

4.4 Impacto de la función de costos de las regiones con realce de contraste (CeR-Loss)

Debido a que la función de costo propuesta en este trabajo, denominada CeR-Loss, es el componente más importante de la arquitectura D-RiedGAN, la cual además es la que permite que el modelo obtenga los mejores resultados en comparación con los demás modelos, se realiza una prueba experimental con el fin de determinar el impacto de esta función en el entrenamiento del modelo. En este sentido, se adopta la misma configuración de la arquitectura D-RiedGAN presentada en la sección anterior, pero se varían los parámetros λ_{l1} y $\lambda_{L_{ovasz}}$ de manera en que se mantiene por encima el $\lambda_{L_{ovasz}}$ en todos los casos. La Figura 5 muestra tres experimentos, para $\lambda_{l1} = 20$ y $\lambda_{L_{ovasz}} = 30$; $\lambda_{l1} = 40$ y $\lambda_{L_{ovasz}} = 60$; y $\lambda_{l1} = 100$ y $\lambda_{L_{ovasz}} = 150$.

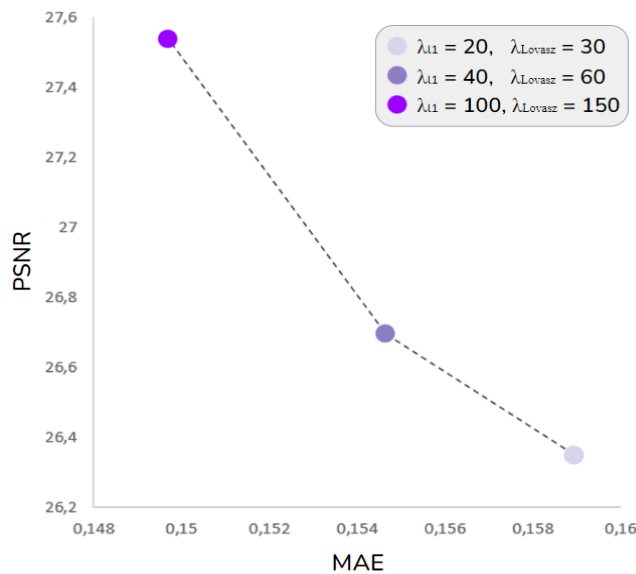


Figura 5. Evaluación de impacto de la función CeR-Loss en el desempeño de la arquitectura D-RiedGAN. Fuente: elaboración propia.

La Figura 5 muestra que para las métricas MAE y PSNR existe una tendencia positiva que mejora los resultados de esas métricas a medida que los valores de λ_{l1} y $\lambda_{L_{ovasz}}$ incrementan, obteniendo su mejor desempeño cuando $\lambda_{l1} = 100$ y $\lambda_{L_{ovasz}} = 150$. Esto confirma el aporte positivo en los resultados por parte de la función de costo CeR-Loss.

5. CONCLUSIONES

En este trabajo se propuso una función de costo denominada CeR-Loss, la cual aprovecha el comportamiento de la captación del agente de contraste para generar imágenes sintéticas postcontraste a partir de imágenes precontraste en estudios de DCE-MRI. Esta función se usa en dos nuevas arquitecturas de aprendizaje profundo, que hemos denominado G-RiedGAN y D-RiedGAN, las cuales centran su atención en las regiones de realce de contraste para mejorar el proceso de generación de las imágenes sintéticas postcontraste. Estas arquitecturas establecen un aporte hacia la eliminación del uso de agentes de contraste y a la reducción del costo de los estudios de DCE-MRI usados como ayuda para la detección del cáncer de mama.

Las arquitecturas G-RiedGAN y D-RiedGAN que se proponen este trabajo integran elementos de las arquitecturas RIED-Net y Pix2Pix bajo la estructura de la arquitectura EaGAN. A diferencia de esta última, la D-RiedGAN integra un filtro para la detección de las regiones de realce de contraste que son las zonas clave de la síntesis en el problema de análisis de imágenes de DCE-MRI para la detección y diagnóstico del cáncer de mama. Las regiones de contraste detectadas en la imagen sintética se usan para guiar el aprendizaje de la red a partir de la función de pérdida Lovász y BCE integradas a la función de pérdida del generador y del discriminador (CeR-Loss).

La evaluación comparativa se realizó teniendo en cuenta dos enfoques. El primero comparó los modelos propuestos con la función CeR-Loss, contra los modelos base del estado del arte y un grupo de modelos ensamblados. Los resultados, considerando las métricas MAE, PSNR y SSIM, muestran que los modelos propuestos obtienen un mejor desempeño en la síntesis de las regiones de realce de contraste en términos del ruido y borrosidades. El segundo enfoque evaluó el impacto de la función de pérdida CeR-Loss en el proceso de aprendizaje. Los resultados, en términos de las mismas métricas, muestran un impacto positivo en la síntesis de las regiones de contraste a medida que se pondera con un valor mayor la función CeR-Loss.

Si bien la validación se realizó utilizando solo métricas cuantitativas basadas en las intensidades de los píxeles de las imágenes sintéticas, se reconoce la necesidad de a futuro hacer una validación cualitativa con radiólogos expertos para determinar la validez diagnóstica de estas imágenes. Como trabajo futuro, se sugiere realizar un análisis sobre cómo los modelos de línea base y los modelos propuestos se comportan al entrenarse con bases de datos de imágenes heterogéneas en términos de la calidad de los estudios (0.5T, 1.5T, 3T y 7T), dosis y medios de contraste diferentes. Igualmente, consideramos que las imágenes sintéticas postcontraste generadas podrían tener un impacto positivo en el entrenamiento de modelos de detección y clasificación de cáncer de mama utilizando imágenes de estudios de MRI convencionales, puesto que estas imágenes pueden brindar información adicional para entrenar esos modelos de clasificación.

6. AGRADECIMIENTO Y FINANCIACIÓN

Este trabajo ha sido financiado parcialmente por el Instituto Tecnológico Metropolitano de Medellín (ITM) a través del proyecto de investigación P20213, la Institución Universitaria Pascual Bravo y Ayudas Diagnósticas SURA S.A.S. a través del convenio específico CE-007-2020. También por la agencia de educación SAPIENCIA de Medellín.

CONFLICTOS DE INTERÉS

Los autores declaran que no existe conflicto de intereses con respecto a la publicación de este trabajo.

CONTRIBUCIÓN DE AUTORÍA

Sara Cañaveral: Diseño y realización de los experimentos, análisis e interpretación de los datos, redacción del borrador del manuscrito y revisión final.

Rubén Fonnegra: contribuyó con el diseño de los experimentos, el análisis e interpretación de los datos, en las revisiones parciales y final del manuscrito.

Carlos Mera-Banguero: contribuyó con el diseño de los experimentos, el análisis e interpretación de los datos, en las revisiones parciales y final del manuscrito. El autor Carlos Mera-Banguero, inició su participación en este trabajo como docente del Instituto Tecnológico Metropolitano de Medellín (ITM), pero sus contribuciones a la versión final del manuscrito se realizaron estando afiliado a la Universidad de Antioquia.

7. REFERENCIAS

- [1] M. P. Jimenez Herrera, “Informe de Evento Cáncer de Mama y Cuello Uterino en Colombia 2018,” Instituto Nacional de Salud, Colombia, Versión 04, May 2018. [Online]. Available: <https://bit.ly/3J1FcnV>
- [2] M. Martín, A. Herrero, and I. Echavarría, “El cáncer de mama,” *Arbor*, vol. 191, no. 773, p. a234, Jun. 2015. <https://doi.org/10.3989/arbor.2015.773n3004>
- [3] IARC. “Data visualization tools for exploring the global cancer burden in 2022.” iarc.who. Accessed: Feb. 20, 2024. [Online.] Available: <https://gco.iarc.who.int/today/en>
- [4] X. Zhou *et al.*, “A Comprehensive Review for Breast Histopathology Image Analysis Using Classical and Deep Neural Networks,” *IEEE Access*, vol. 8, pp. 90931-90956, May. 2020. <https://doi.org/10.1109/ACCESS.2020.2993788>
- [5] H. V. Guleria *et al.*, “Enhancing the breast histopathology image analysis for cancer detection using Variational Autoencoder,” *Int. J. Environ. Res. Public Health.*, vol. 20, no. 5, p. 4244, Feb. 2023. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10002012/>
- [6] Instituto Nacional del Cáncer. “Tratamiento del cáncer de seno.” cancer.gov. Accessed: Feb. 20, 2024. [Online.] Available: <https://www.cancer.gov/espanol/tipos/seno/paciente/tratamiento-seno-pdq>
- [7] S. G. Macias, “Métodos de imagen en el estudio de la mama - Ecografía mamaria,” Editorial Medica Panamericana, Bogotá, Colombia, Módulo 1, 2019. <https://bit.ly/4aFlg4y>
- [8] P. E. Freer, “Mammographic breast density: Impact on breast cancer risk and implications for screening,” *Radiographics*, vol. 35, no. 2, pp. 302–315, Mar. 2015. <https://doi.org/10.1148/rg.352140106>
- [9] P. Campáz-Usuga, R. D. Fonnegra, and C. Mera, “Quality Enhancement of Breast DCE-MRI Images Via Convolutional Autoencoders,” in *2021 IEEE 2nd International Congress of Biomedical Engineering and Bioengineering (CI-IB&BI)*, Bogotá D.C., Colombia, 2021, pp. 1-4. <https://doi.org/10.1109/CI-IBI54220.2021.9626097>
- [10] Y. M. Rodríguez Marcano, I. González, H. Palencia, M. Sandoval, and L. León, “Mamografía espectral con realce de contraste. Nuestra experiencia,” *Revista Venezolana de Oncología*, vol. 26, no. 4, pp. 743–751, Dec. 2014. <https://www.redalyc.org/articulo.oa?id=375633971003>
- [11] I. Pérez-Zúñiga, Y. Villaseñor-Navarro, M. P. Pérez-Badillo, R. Cruz-Morales, C. Pavón-Hernández, and L. Aguilar-Cortázar, “Resonancia magnética de mama y sus aplicaciones,” *Gaceta Mexicana de Oncología*, vol. 11, no. 4, pp. 268–280, 2012. <https://www.elsevier.es/es-revista-gaceta-mexicana-oncologia-305-articulo-resonancia-magnetica-mama-sus-aplicaciones-X1665920112544919>
- [12] C. Balleyguier *et al.*, “New potential and applications of contrast-enhanced ultrasound of the breast: Own investigations and review of the literature,” *Eur. J. Radiol.*, vol. 69, no. 1, pp. 14–23, Jan. 2009. <https://doi.org/10.1016/J.EJRAD.2008.07.037>
- [13] R. Valenzuela, O. Arevalo, A. Tavera, R. Riascos, E. Bonfante, and R. Patel, “Imágenes del depósito de gadolinio en el sistema nervioso central,” *Revista Chilena de Radiología*, vol. 23, no. 2, pp. 59–65, Jul.2017. <https://doi.org/10.4067/S0717-93082017000200005>
- [14] F. Gao, T. Wu, X. Chu, H. Yoon, Y. Xu, and B. Patel, “Deep Residual Inception Encoder-Decoder Network for Medical Imaging Synthesis,” *IEEE J. Biomed. Health Inform.*, vol. 24, no. 1, pp. 39–49, Jan. 2020. <https://doi.org/10.1109/JBHI.2019.2912659>
- [15] F. Gao *et al.*, “SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis,” *Computerized Medical Imaging and Graphics*, vol. 70, pp. 53–62, Dec. 2018. <https://doi.org/10.1016/j.compmedimag.2018.09.004>
- [16] K. Wu *et al.*, “Image synthesis in contrast MRI based on super resolution reconstruction with multi-refinement cycle-consistent generative adversarial networks,” *J. Intell. Manuf.*, vol. 31, no. 5, pp. 1215–1228, Jun. 2020. <https://doi.org/10.1007/s10845-019-01507-7>
- [17] E. Kim, C. Hwan-Ho, J. Kwon, O. Young-Tack, E. S. Ko, and H. Park, “Tumor-Attentive Segmentation-Guided GAN for Synthesizing Breast Contrast-Enhanced MRI Without Contrast Agents,” *IEEE Journal*

- of Translational Engineering in Health and Medicine*, vol. 11, pp. 32-43, Nov. 2023. <https://doi.org/10.1109/JTEHM.2022.3221918>
- [18] Y. Jiang, Y. Zheng, W. Jia, S. Song, and Y. Ding, "Synthesis of contrast-enhanced spectral mammograms from low-energy mammograms using cGAN-based synthesis network," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, M. de Bruijne, et al., Eds. Cham: Springer International Publishing, 2021, pp. 68–77. https://doi.org/10.1007/978-3-030-87234-2_7
- [19] D. Huangz, and M. Feng, "Understanding Deep Convolutional Networks for Biomedical Imaging: A Practical Tutorial," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany, 2019, pp. 857-863. <https://doi.org/10.1109/EMBC.2019.8857529>
- [20] C. Shorten, and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, Jul. 2019. <https://doi.org/10.1186/s40537-019-0197-0>
- [21] A. Beers et al., "High-resolution medical image synthesis using progressively grown generative adversarial networks," 2018, *ArXiv: 1805.03144*. <https://arxiv.org/abs/1805.03144>
- [22] T. Shen, C. Gou, J. Wang, and F. -Y. Wang, "Collaborative Adversarial Networks for Joint Synthesis and Segmentation of X-ray Breast Mass Images," in *2020 Chinese Automation Congress (CAC)*, Shanghai, China, 2020, pp. 1743-1747. <https://doi.org/10.1109/CAC51589.2020.9326848>
- [23] Y. Pang, J. Lin, T. Qin, and Z. Chen, "Image-to-Image Translation: Methods and Applications," *IEEE Trans. Multimedia*, vol. 24, pp. 3859–3881, Sep. 2021. <https://doi.org/10.1109/TMM.2021.3109419>
- [24] M. Carmen, J. Lizandra, C. Monserrat, A. José, and H. Orallo, "Síntesis de Imágenes en Imagen Médica," *Universidad Politécnica de Valencia*, 2003. <https://josephorallo.webs.upv.es/escrits/ACTA3.pdf>
- [25] A. Anwar "Difference between AutoEncoder (AE) and Variational AutoEncoder (VAE)," towardsdatascience.com Accessed: Feb. 20, 2024. [Online]. Available: <https://towardsdatascience.com/difference-between-autoencoder-ae-and-variational-autoencoder-vae-ed7be1c038f2>
- [26] W. Weng, and X. Zhu, "INet: Convolutional Networks for Biomedical Image Segmentation," *IEEE Access*, vol. 9, pp. 16591-16603, 2021. <https://doi.org/10.1109/ACCESS.2021.3053408>
- [27] I. J. Goodfellow et al., "Generative Adversarial Networks," *Advances in Neural Information Processing Systems*, vol. 14, Jun. 2014. https://papers.nips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afcf3-Paper.pdf
- [28] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso "INbreast: toward a full-field digital mammographic database," *Acad. Radiol.*, vol. 19, no. 2, pp. 236-248, Feb. 2012. <https://doi.org/10.1016/j.acra.2011.09.014>
- [29] F. Gao, T. Wu, X. Chu, H. Yoon, Y. Xu, and B. Patel, "Deep Residual Inception Encoder-Decoder Network for Medical Imaging Synthesis," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 1, pp. 39–49, Apr. 2020. <https://doi.org/10.1109/JBHI.2019.2912659>
- [30] M. Mori et al., "Feasibility of new fat suppression for breast MRI using pix2pix," *Jpn. J. Radiol.*, vol. 38, no. 11, pp. 1075–1081, Nov. 2020. <https://doi.org/10.1007/s11604-020-01012-5>
- [31] P. Isola, Z. Jun-Yan, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 5967-5976. <https://doi.org/10.1109/CVPR.2017.632>
- [32] P. Wang et al., "Synthesizing the First Phase of Dynamic Sequences of Breast MRI for Enhanced Lesion Identification," *Front. Oncol.*, vol. 11, Dec. 2021. <https://doi.org/10.3389/fonc.2021.792516>
- [33] Z. Sani, R. Prasad, and E. K. M. Hashim, "Breast Cancer Detection in Mammography using Faster Region Convolutional Neural Networks and Group Convolution," *ETE J. Res.*, pp. 1–17, May 2024. <https://doi.org/10.1080/03772063.2024.2352643>
- [34] M. Fan et al., "Generative adversarial network-based synthesis of contrast-enhanced MR images from precontrast images for predicting histological characteristics in breast cancer," *Phys. Med. Biol.*, vol. 69, no. 9, p. 095002, Apr. 2024. <https://doi.org/10.1088/1361-6560/ad3889>
- [35] O. Young-Tack, E. Ko, and H. Park, "TDM-Stargan: Stargan Using Time Difference Map to Generate Dynamic Contrast-Enhanced Mri from Ultrafast Dynamic Contrast-Enhanced Mri," in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, Kolkata, India, 2022, pp. 1-5. <https://doi.org/10.1109/ISBI52829.2022.9761463>
- [36] T. Fujioka et al., "Proposal to improve the image quality of short-acquisition time-dedicated breast positron emission tomography using the Pix2pix generative adversarial network," *Diagnostics*, vol. 12, no. 12, p. 3114, Dec. 2022. <https://doi.org/10.3390/diagnostics12123114>
- [37] G. Jiang, Y. Lu, J. Wei, and Y. Xu, "Synthesize Mammogram from Digital Breast Tomosynthesis with Gradient Guided cGANs," *Springer International Publishing*, D. Shen et al., Eds. vol. 11769, Oct. 2019. https://doi.org/10.1007/978-3-030-32226-7_89

- [38] B. Yu, L. Zhou, L. Wang, Y. Shi, J. Fripp, and P. Bourgeat, “Ea-GANs: Edge-Aware Generative Adversarial Networks for Cross-Modality MR Image Synthesis,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 7, pp. 1750–1762, Jan. 2019. <https://doi.org/10.1109/TMI.2019.2895894>
- [39] B. H. Menze *et al.*, “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS),” *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, Dec. 2015. <https://doi.org/10.1109/TMI.2014.2377694>
- [40] D. Duque-Arias *et al.*, “On power jaccard losses for semantic segmentation,” in *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, Setúbal, Portugal, 2021, pp. 561–568. <https://doi.org/10.5220/0010304005610568>
- [41] M. Berman, A. R. Triki, and M. B. Blaschko, “The Lovasz-Softmax Loss: A Tractable Surrogate for the Optimization of the Intersection-Over-Union Measure in Neural Networks,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 4413–4421. <https://doi.org/10.1109/CVPR.2018.00464>
- [42] B. Xu, N. Wang, T. Chen, and M. Li, “Empirical Evaluation of Rectified Activations in Convolutional Network,” 2015, *arXiv:1505.00853*. <http://arxiv.org/abs/1505.00853>
- [43] A. Horé, and D. Ziou, “Image Quality Metrics: PSNR vs. SSIM,” in *2010 20th International Conference on Pattern Recognition*, Istanbul, Turkey, 2010, pp. 2366–2369. <https://doi.org/10.1109/ICPR.2010.579>