

GENERACIÓN DE SECUENCIAS CINE-MRI CARDIACOS UTILIZANDO
MODELOS GENERATIVOS DE APRENDIZAJE PROFUNDO

SANTIAGO GÓMEZ HERNÁNDEZ

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA

2019

GENERACIÓN DE SECUENCIAS CINE-MRI CARDIACOS UTILIZANDO
MODELOS GENERATIVOS DE APRENDIZAJE PROFUNDO

SANTIAGO GÓMEZ HERNÁNDEZ

Una tesis presentada en cumplimiento de los requisitos para el grado de Ingeniero
de Sistemas e Informática

Director:

Ph. D Fabio Martínez Carrillo.

Profesor escuela de Ingeniería de Sistemas e informática

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA

2019

AGRADECIMIENTOS

El autor expresa su agradecimiento:

Al grupo de investigación Biomedical Imaging, Vision and Learning Laboratory (BivL²ab) y al semillero de investigación en análisis de movimiento y visión por computador (MACV), principalmente al profesor Fabio Martínez Carrillo por ser un gran guía, por su paciencia, dedicación, esfuerzo y orientación. También quisiera agradecer por la preocupación de la formación de habilidades íntegras y profesionales que siempre ha inculcado, sin él no hubiera sido posible la realización de este trabajo.

A todos mis amigos de infancia y universidad y a aquellas personas que fueron parte de mi formación, que de una u otra manera me han permitido construir el ser humano que soy, por su fiel compañía y sincera amistad.

A la escuela de Ingeniería de Sistemas e Informática (EISI) y a la Universidad Industrial de Santander (UIS) por la gran formación que me han brindado, lo cual ha sido fundamental para ser un buen profesional.

Finalmente quiero realizar un agradecimiento especial a mis padres, abuelos, hermanos y demás familiares por sus enseñanzas, confianza y apoyo, elementos que han tenido una verdadera importancia para seguir cada día.

TABLA DE CONTENIDO

INTRODUCCIÓN.....	12
1. PLANTEAMIENTO Y JUSTIFICACIÓN DEL PROBLEMA	15
2. OBJETIVOS	16
2.1 OBJETIVO GENERAL.....	16
2.2 OBJETIVOS ESPECÍFICOS.....	16
3. MARCO TEÓRICO Y ESTADO DEL ARTE	17
3.1 AUMENTO DE UN CONJUNTO DE DATOS.....	17
3.2 MODELOS GENERATIVOS VS DISCRIMINATIVOS.....	19
3.3 REDES GENERATIVAS ADVERSARIAS (GANS).....	21
3.4 GANS EN SECUENCIAS CINE-MRI CARDIACAS.....	24
4. MÉTODO PROPUESTO.....	28
4.1 CONDICIONANDO LA GENERACIÓN DE IMÁGENES CON CONDICIONES CARDIACAS.....	28
4.2 ESTRATEGIA PROGRESIVA ADVERSARIA: PGAN.....	30
4.2.1 Añadiendo capas al generador.....	32
4.2.2 Añadiendo capas al discriminador.....	33
4.2.3 Incrementando la variación de los datos producidos.....	33
4.2.4 Aprendizaje en el generador y discriminador.....	35
4.3 GENERANDO ESPACIO LATENTE.....	35
4.4 CONFIGURACIÓN EXPERIMENTAL	36
4.4.1 Dataset.....	37
4.4.2 Configuración de entrenamiento PGAN.....	38
4.4.3 Evaluación con métricas de similitud aparente y similitud entre distribuciones de entrenamiento y sintética.....	39

4.4.3.1 MS-SSIM.....	39
4.4.3.2 Sliced Wasserstein Distance.....	40
4.4.4 Evaluación con tareas de clasificación.....	41
5. EVALUACIÓN Y RESULTADOS.....	42
6. CONCLUSIONES Y PERSPECTIVAS.....	50
BIBLIOGRAFÍA.....	51

LISTA DE FIGURAS

Figura 1. Algunas transformaciones geométricas aplicadas sobre imágenes de resonancias magnéticas cardiacas.....	18
Figura 2. Algunas transformaciones fotométricas aplicadas sobre imágenes de resonancias magnéticas cardiacas.....	19
Figura 3. Diferentes modelos discriminativos para clasificar un conjunto de datos.	20
Figura 4. Aproximación de una distribución a otra utilizando un modelo generativo.	20
Figura 5. Marco de referencia GANs modelan la distribución de datos de resonancias magnéticas cardiacas.....	22
<i>Figura 6. Algunas de las arquitecturas más relevantes de las GANs. Imagen adaptada de Generative Adversarial Network in Medical Imaging: A Review.....</i>	<i>23</i>
Figura 7. Estructura del enfoque propuesto para generación de resonancias magnéticas cardiacas. La arquitectura convolucional de ambas redes neuronales va creciendo progresivamente a medida que cada nivel de detalle llega a un punto óptimo.....	31
Figura 8. Introducción gradual de nuevas capas al generador. Antes de la transición (1), durante la transición (2) y después de la transición (3). El bloque 2x significa aumentar espacialmente al doble. El bloque toRGB representa una capa que proyecta vectores d. 33	
Figura 9. Introducción gradual de nuevas capas al discriminador. Antes de la transición (1), durante la transición (2) y después de la transición (3). El bloque 0.5x significa reducir espacialmente a la mitad. El bloque fromRGB representa una capa que proyecta.....	34
Figura 10. Estructura general del enfoque propuesto para la recuperación del vector de espacio latente embebido z a través de la retropropagación utilizando una función de costo residual.	36
Figura 11. Ejemplos de imágenes de resonancias magnéticas en diferentes dimensiones espaciales para el entrenamiento de la GAN progresiva.	38
Figura 12. Ejemplos de resonancias magnéticas generadas.....	43
Figura 13. Ejemplos de resonancias magnéticas generadas utilizando una estrategia GAN sin condición.	44

Figura 14. Similaridad estructural entre las imágenes generadas en diferentes etapas del entrenamiento y las imágenes del conjunto de entrenamiento.....	45
Figura 15. Distancia entre la distribución de entrenamiento P_r y la distribución de las imágenes generadas P_g en diferentes etapas del entrenamiento.	47
Figura 16. Agrupamiento no supervisado UMAP de los vectores embebidos condicionados. Un agrupamiento multiclase y biclase se llevó a cabo en diferentes experimentos.	48
Figura 17. En la parte superior: tarea de clasificación multiclase utilizando un Random Forest para clasificación multiclase. En la parte inferior: tareas de clasificación binaria NOR contra todos utilizando un Random Forest para clasificación binaria.	49

RESUMEN

TÍTULO: GENERACIÓN DE SECUENCIAS CINE-MRI CARDIACOS UTILIZANDO MODELOS GENERATIVOS DE APRENDIZAJE PROFUNDO.*

AUTOR: SANTIAGO GÓMEZ HERNÁNDEZ**

PALABRAS CLAVE: GANS PROGRESIVAS, EMULACIÓN DE PATRONES DE MOVIMIENTO, CINE-MRI.

DESCRIPCIÓN:

Una restricción fuerte en la generación de imágenes médicas es la presencia de artefactos que pueden alterar la estructura anatómica y causar la pérdida de interpretación clínica en estas imágenes médicas. Adicionalmente, enfoques actuales están dedicados a reproducir características visuales y similitudes entre imágenes sin prestar atención a variaciones en la estructura que sugieren comportamientos patológicos de estructuras vivientes. Particularmente para las secuencias cardiacas cine-MRI, la existencia de estos problemas afecta en gran medida el diagnóstico emitido por el experto, debido a que la descripción morfológica de los ventrículos es fundamental para llevar a cabo este. Este trabajo introduce una estrategia con una red generativa adversaria progresiva (PGAN) que permite emular secuencias cardiacas cine-MRI con sentido patológico, es decir, que tengan características de una patología cardiaca. Un conjunto etiquetado de secuencias cine-MRI, de cinco condiciones cardiacas diferentes, es usado como condición de entrenamiento para recuperar imágenes cardiacas con sentido clínico. La PGAN comienza generando imágenes con detalles generales de las secuencias cine-MRI de baja resolución, hasta imágenes detalladas de alta resolución MRI. En cada iteración progresiva, una etiqueta indicando la condición cardiaca, restringe la generación de nuevas imágenes MRI, permitiendo emular una patología particular. Utilizando la red generativa entrenada, las imágenes generadas son propagadas hacia atrás a través de la red para obtener su correspondiente descriptor embebido. El conjunto de vectores embebidos de diferentes condiciones cardiacas es proyectado a un espacio de baja dimensionalidad para llevar a cabo una tarea no supervisada de agrupamiento. Como era de esperarse, los vectores embebidos que emulan la misma condición cardiaca están más cerca. Estos vectores embebidos también fueron usados para construir un problema de clasificación supervisada, consiguiendo un puntaje promedio del 80% de precisión en la tarea de clasificación de imágenes cine-MRI.

* Trabajo de Grado

** Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Fabio Martínez Carrillo, Ph.D.

ABSTRACT

TITLE: CARDIAC CINE-MRI SEQUENCES SYNTHESIS USING DEEP LEARNING GENERATIVE MODELS.*

AUTHOR: SANTIAGO GÓMEZ HERNÁNDEZ**

KEYWORDS: PROGRESSIVE GANS, MOTION PATTERNS EMULATION, CINE-MRI.

DESCRIPTION:

A strong restriction on medical image synthesis is the presence of artifacts that might alter the anatomical structure, losing clinical interpretation. Additionally, current approaches are dedicated to reproducing visual features and image similarities without paying attention to structure variations that could suggest pathological behaviors on living structures. Particularly for cardiac-MRI sequences, these issues result very sensibly since the morphological ventricles description is fundamental to carry out a diagnosis. This work introduces a progressive generative adversarial network (PGAN) strategy that allows to emulate cardiac cine-MRI sequences with disease sense. A set of labeled cine-MRI, from five different heart conditions, are used as trained conditions to recover cardiac images with clinical sense. The PGAN starts by generating coarse images from low resolution to fine detailed images on high MRI resolution. At each progressive iteration, a label heart condition restricts the synthesis of new MRI images, allowing to emulate a particular pathology. From learned generated sub-network, the new generated image is back-propagated to obtain a corresponding embedding descriptor. The set of embeddings vectors from different heart conditions, are projected into a low-dimensional space to carried out a non-supervised clustering task. As expected, embeddings that emulate same heart conditions are closer. Also, these embeddings were used to built a supervised classification problem, achieving an average score of 80% on classification task of artificial cine-MRI images.

* Bachelor tesis.

** Faculty of Physicomechanical Engineering. School of Systems and Computer Engineering. Director: Fabio Martínez Carrillo, Ph.D.

INTRODUCCIÓN

Las enfermedades cardiovasculares (CVD, por sus siglas en inglés) son la mayor causa de muerte en el mundo, responsables de aproximadamente el 29% de las muertes globales totales¹. Hoy en día, las secuencias cardíacas cine-MRI son el mejor soporte diagnóstico observacional para analizar información estructural y dinámica de las paredes del miocardio a través del ciclo cardíaco. Los sistemas de diagnóstico asistidos por computadora (CADs, por sus siglas en inglés) ayudan a los médicos en la detección, diagnóstico y seguimiento de las enfermedades cardíacas, reduciendo la variabilidad entre expertos y facilitando el proceso de decisión. Para estas tareas, un modelamiento adecuado requiere una cantidad considerable de datos manualmente anotados. La recolección de estos datos cine-MRI cardíacos es una tarea desafiante, donde estos volúmenes tiene una alta variabilidad espaciotemporal natural y grandes variaciones asociadas con el protocolo de adquisición de los escáneres MRI.

Recientemente, diversos campos han adoptado métodos para aumentar los datos de manera sintética para incrementar la robustez de modelos supervisados y mejorar la representación aprendida. Estas estrategias van desde una simple transformación hasta estrategias que aumentan y resaltan patrones relevantes de un dominio en específico ^{2 3}. Hoy en día, existen muchas estrategias para generar datos sintéticos, incluyendo modelos autorregresivos, codificadores automáticos

¹ ROTH, Gregory A., et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, 2018, vol. 392, no 10159, p. 1736-1788.

² PEREZ, Luis; WANG, Jason. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*, 2017.

³ TAYLOR, Luke; NITSCHKE, Geoff. Improving deep learning using generic data augmentation. *arXiv preprint arXiv:1708.06020*, 2017.

variacionales (VAEs, por sus siglas en inglés), y redes generativas adversarias (GANs, por sus siglas en inglés). Los modelos autorregresivos predicen el comportamiento futuro en base a comportamientos pasados. Son utilizados para predecir cuando existe una correlación entre valores en una serie de tiempo y los valores que le preceden y suceden a estos. Solo se utilizan los datos pasados para modelar el comportamiento, por esto el nombre de autorregresivos. Por otro lado, los codificadores automáticos variacionales consisten en utilizar un codificador y decodificador como redes neuronales para aprender el mejor esquema de codificación-decodificación utilizando un problema iterativo. Estos VAEs pueden verse como reductores de dimensionalidad. Por último, las GANs están basadas en un juego adversario, donde dos redes neuronales, el generador y discriminador, mejoran al mismo tiempo⁴. Este proceso permite al generador aprender una distribución de probabilidad asociada con el conjunto de imágenes usadas en el entrenamiento. Trabajos recientes han utilizado las GANs para incrementar muestras en imágenes médicas⁵. Sin embargo, una advertencia importante es que las funciones de costo utilizadas en las redes generativas adversarias pueden generar imágenes sintéticas con artefactos notables. Estos artefactos pueden ser problemáticos en imágenes médicas, puesto que regiones clínicamente relevantes pueden ser distorsionadas. En el estado del arte se han propuesto algunos trabajos que tienen como base las resonancias magnéticas cardiacas y las redes generativas adversarias. Estos trabajos se enfocan en diversas tareas como la segmentación, superresolución, síntesis y cambio de modalidad (MRI a CT). Para lograr estas tareas proponen soluciones como utilizar variantes de las arquitecturas GANs, condicionar el entrenamiento con análisis de regiones de interés en contextos globales y/o locales, nuevas funciones de costo y procesamiento de los

⁴ GOODFELLOW, Ian, et al. Generative adversarial nets. En *Advances in neural information processing systems*. 2014. p. 2672-2680.

⁵ YI, Xin; WALIA, Ekta; BABYN, Paul. Generative adversarial network in medical imaging: A review. *Medical image analysis*, 2019, p. 101552.

datos para reducir la variabilidad en estos ⁶ ⁷. Sin embargo, ninguno de estos trabajos tiene explícitamente en cuenta las diversas condiciones cardiacas anotadas a las imágenes del conjunto de entrenamiento.

La contribución principal de este trabajo es un método que genera secuencias cine-MRI cardiacas con una condición particular cardiaca. Luego, las imágenes cardiacas resultantes son usadas como entradas para generar un espacio latente. Para lo anterior, una imagen es propagada hacia atrás del generador entrenado de la PGAN para obtener un vector embebido representativo de la imagen. Estos vectores embebidos resultantes son proyectados a un espacio de baja dimensionalidad. Imágenes generadas limitadas por la misma condición cardiaca están más cerca en este espacio embebido, mostrando la efectividad de la estrategia propuesta. Adicionalmente, estos vectores fueron evaluados en una tarea supervisada mostrando un buen desempeño en la tarea de clasificación de condiciones cardiacas. A lo mejor de nuestro conocimiento, este trabajo representa el primer esfuerzo para generar cine-MRI que reproduzca características morfológicas de varias enfermedades cardiacas. La estrategia propuesta abre la posibilidad de analizar nuevos patrones desde un espacio de baja dimensionalidad embebido, que puede ser relevante para encontrar nuevos biomarcadores relacionados a un desorden cardiaco particular.

⁶ ZHANG, Le; GOOYA, Ali; FRANGI, Alejandro F. Semi-supervised assessment of incomplete LV coverage in cardiac MRI using generative adversarial nets. En International Workshop on Simulation and Synthesis in Medical Imaging. Springer, Cham, 2017. p. 61-68.

⁷ MURUGESAN, Balamurali, et al. Recon-GLGAN: A Global-Local Context Based Generative Adversarial Network for MRI Reconstruction. En International Workshop on Machine Learning for Medical Image Reconstruction. Springer, Cham, 2019. p. 3-15.

1. PLANTEAMIENTO Y JUSTIFICACIÓN DEL PROBLEMA

Aumentar datos es una tarea fundamental para modelar la variabilidad en problemas complejos y con restricción de muestras en un dominio específico. Utilizando esta estrategia es posible definir fronteras robustas entre clases, y además conocer los principales descriptores que aportan a un fenómeno en particular. Este tema ha sido abordado ampliamente desde diferentes perspectivas que incluyen transformaciones rígidas en diferentes espacios, pero limitados en cuanto a descripción realista de muestras en el dominio. Particularmente, en aplicaciones clínicas, estas transformaciones no representan de forma fiable las secuencias clínicas de interés. Resultando, por lo tanto, en composiciones no-naturales y con información irrelevante que afectan la convergencia de algoritmos de aprendizaje de máquina. En cuanto a la generación de secuencias cardiacas cine-MRI existen múltiples desafíos relacionados con la alta-dimensionalidad de los datos (secuencias en 3D+t), estructuras complejas cardiacas, y movimientos bruscos, característicos, durante el ciclo cardiaco. Las aproximaciones propuestas, para producción de cine-MRI, aún tienen limitaciones computacionales y de resolución espacial^{8,9}. Tomando lo anterior en contexto, surge la siguiente pregunta de investigación:

- ¿Cuáles son los alcances de reproducción de las arquitecturas GANs para generar cines cardiacos MRI?

⁸ Ibít., p 14.

⁹ CHARTSIAS, Agisilaos, et al. Adversarial image synthesis for unpaired multi-modal cardiac data. En International Workshop on Simulation and Synthesis in Medical Imaging. Springer, Cham, 2017. p. 3-13.

2. OBJETIVOS

2.1 OBJETIVO GENERAL.

Desarrollar un método basado en modelos generativos de aprendizaje profundo (GANS) para generar secuencias sintéticas de cine-MRI cardiaco.

2.2 OBJETIVOS ESPECÍFICOS

- ❖ Seleccionar un conjunto de datos públicos de secuencias cine-MRI para el desarrollo del trabajo
- ❖ Desarrollar un modelo generativo que produzca secuencias sintéticas de cine-MRI cardiaco a partir de una distribución de datos de entrenamiento.
- ❖ Implementar el modelo generativo para la síntesis de secuencias cine-MRI.
- ❖ Validar las secuencias producidas utilizando métricas de similitud aparente y relacionadas con conservación de la estructura de las imágenes.

3. MARCO TEÓRICO Y ESTADO DEL ARTE

Hoy en día, tareas en el campo clínico en cuanto al análisis, clasificación de patologías y/o segmentación de regiones de interés se basan en estrategias de aprendizaje de máquina. Además, el demostrado comportamiento de arquitecturas de aprendizaje profundo, hacen que sea la alternativa ideal para resolver problemas complejos con variabilidad marcada en los objetos de interés. Sin embargo, la efectividad de estas arquitecturas radica en el ajuste hiperparamétrico de funciones de aprendizaje a partir de una gran cantidad de datos. En aplicaciones biomédicas, este hecho, es sin embargo una limitante, lo cual puede conducir a modelos que sobre aprenden y con limitaciones para diseñar funciones generalizadoras del problema.

3.1 AUMENTO DE UN CONJUNTO DE DATOS

Una alternativa para sobrellevar con el problema del reducido número de datos disponibles es la aplicación de estrategias de aumento de datos, a partir de un subconjunto reducido de muestras^{10 11 12}. En general, la creación de datos sintéticos aumenta la capacidad de modelar la variabilidad de los datos y permite producir funciones de aprendizaje con mayor flexibilidad y nivel de generalización. Los métodos geométricos han sido las estrategias clásicas para generar nuevos datos a partir de transformaciones que alteran la geometría de una imagen al transformar

¹⁰ *Ibít.*, p 12.

¹¹ MIKOLAJCZYK, Agnieszka; GROCHOWSKI, Michał. Data augmentation for improving deep learning in image classification problem. En 2018 international interdisciplinary PhD workshop (IIPhDW). IEEE, 2018. p. 117-122.

¹² SIMARD, Patrice Y., et al. Best practices for convolutional neural networks applied to visual document analysis. En *Icdar*. 2003.

cada píxel a un nuevo destino mientras se preserva la estructura global de la imagen. Estas transformaciones son las más usadas para aumentar conjuntos de datos para entrenar modelos de aprendizaje profundo ¹³ debido a su fácil aplicación, sin embargo, estas no aportan nuevas características visuales a las imágenes que podrían mejorar la capacidad de aprendizaje y generalización del modelo. En la figura 1 se puede observar el resultado de aplicar transformaciones geométricas clásicas a una imagen particular.

Así mismo, los métodos fotométricos alteran los espacios de color codificados de una imagen particular, cambiando factores de luminosidad e incluso usando diferentes espacios de

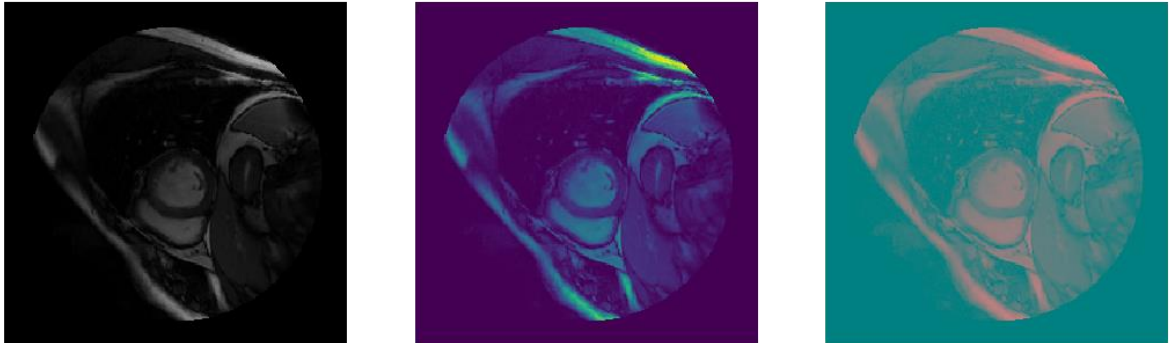
Figura 1. Algunas transformaciones geométricas aplicadas sobre imágenes de resonancias magnéticas cardiacas.



color. Estas transformaciones preservan la geometría de la imagen. En la figura 2 se puede observar las transformaciones obtenidas al introducir variaciones fotométricas. Tanto las transformaciones geométricas como las fotométricas deben ser usadas bajo un modelamiento de la distribución de probabilidad de los datos, donde los ejemplos generados, corresponden a dicha distribución.

¹³ KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. En Advances in neural information processing systems. 2012. p. 1097-1105.

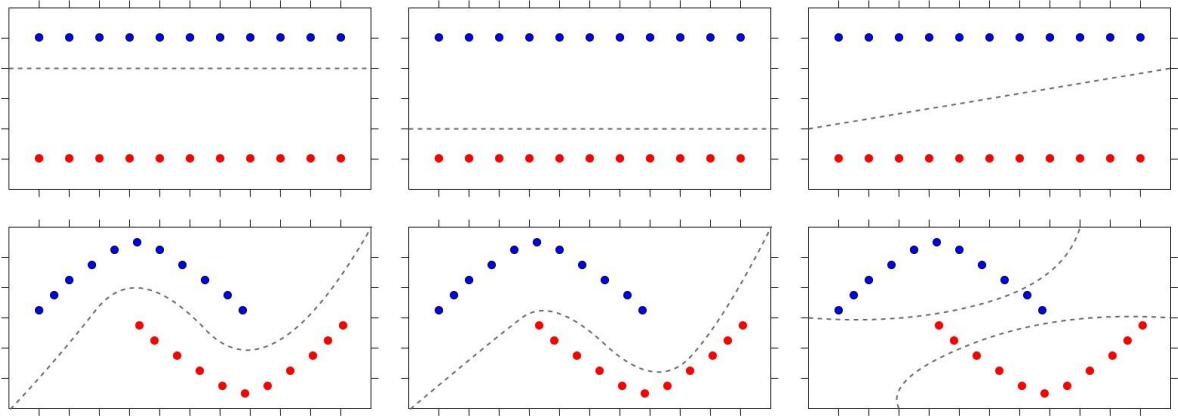
Figura 2. Algunas transformaciones fotométricas aplicadas sobre imágenes de resonancias magnéticas cardiacas.



3.2 MODELOS GENERATIVOS VS DISCRIMINATIVOS

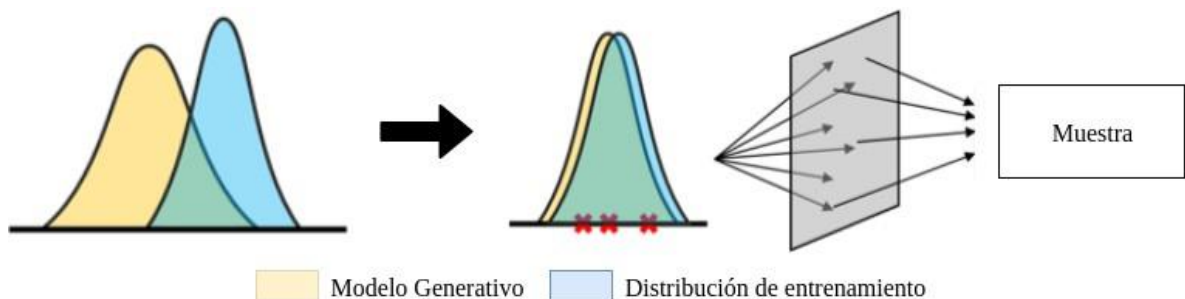
En el aprendizaje de máquina, existen los enfoques en modelos generativos y discriminativos. Estos enfoques calculan clasificadores utilizando diferentes técnicas. Los modelos discriminativos, comúnmente llamados modelos condicionales, son una clase de modelos utilizados en la clasificación probabilística, especialmente en métodos de aprendizaje supervisado de máquina. Un clasificador discriminativo trata de modelar un problema dependiendo exclusivamente de los datos observados, mientras aprende como hacer la clasificación con las estadísticas dadas. Un modelo discriminativo aprende la distribución de probabilidad condicional $P(y|x)$ o el mapeo directo de una variable no observada x a una etiqueta y , basándose en las variables observadas (muestras de entrenamiento) para crear una frontera divisoria entre puntos de una clase y otra (ver Figura 3).

Figura 3. Diferentes modelos discriminativos para clasificar un conjunto de datos.



Por otro lado, los modelos generativos hacen frente a una tarea más difícil que los modelos discriminativos análogos. Estos modelos aprenden la probabilidad conjunta de una distribución $P(x, y)$, es decir, dada una etiqueta y se pueden calcular sus correspondientes características x . A diferencia de los modelos discriminativos, estos modelos generativos dibujan una frontera que captura las características de una distribución de datos e indica qué tan probable es que un ejemplo pertenezca a esta luego de aproximar la distribución del modelo generativo a la distribución de entrenamiento (ver Figura 4).

Figura 4. Aproximación de una distribución a otra utilizando un modelo generativo.



El problema de generar datos puede ser reformulado en un problema de generar un vector aleatorio en el espacio n dimensional R^n que sigue la distribución de probabilidad de los datos que deseamos producir. Para lograr esto utilizando métodos de aprendizaje de máquina existen dos maneras propuestas para llevar el

entrenamiento: una directa y otra indirecta. El método de entrenamiento directo consiste en comparar las distribuciones de probabilidad del conjunto de datos de entrenamiento y generado retro propagar el error a través de la red. Por otro lado, el método de entrenamiento indirecto no compara ambas distribuciones. Este método indirecto entrena una red generativa haciendo que estas dos distribuciones pasen por una tarea posterior elegida, de modo que el proceso de optimización de la red generativa con respecto a la tarea posterior lleve la distribución generada a estar cerca de la distribución real. Esta última idea es la adoptada por los modelos generativos del estado del arte, las redes Generativas Adversarias (GANs) propuestas por Goodfellow et al ¹⁴.

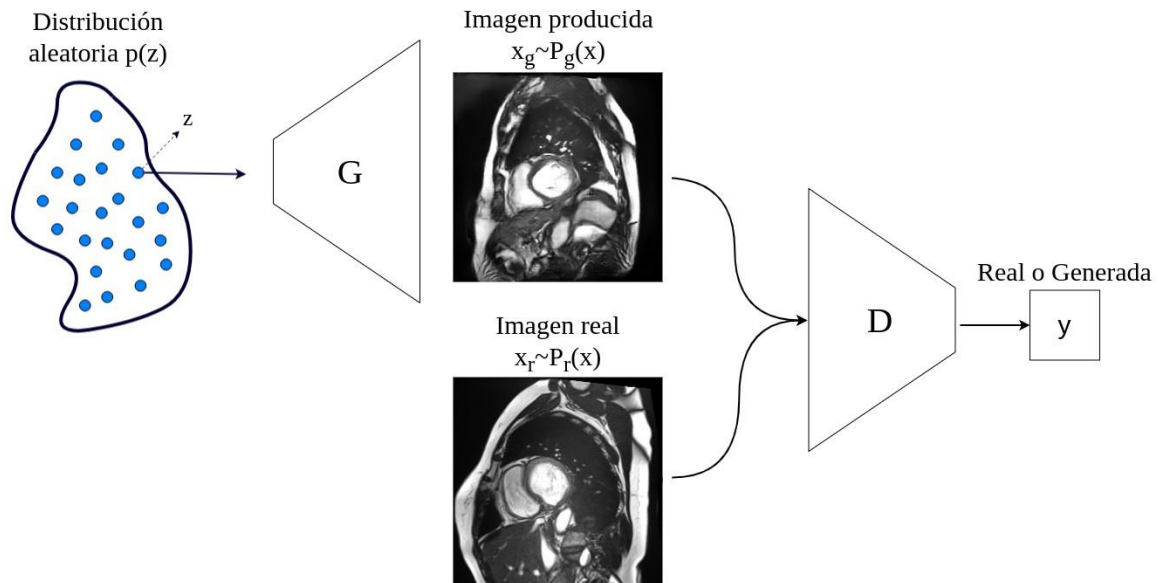
3.3 REDES GENERATIVAS ADVERSARIAS (GANS).

Las arquitecturas GANs son modelos generativos que fueron diseñados para producir muestras directamente de una distribución de datos deseada, sin necesidad de modelar explícitamente la función subyacente de densidad de probabilidad. Esta consiste en dos redes neuronales: un generador G y un discriminador D . La entrada de G es un vector de ruido aleatorio z que pertenece a una distribución aleatoria de ruido $P(z)$. La salida de G , x_g , se espera que tenga una similitud visual con las imágenes reales x_r que pertenecen al conjunto de entrenamiento $P_r(x)$. Para esto G debe aprender un mapeo no lineal parametrizado por θ_g que se denota $x_g = G(z; \theta_g)$. La entrada de D son imágenes reales o generadas por G . La salida de D , y , es un valor escalar indicando la probabilidad de que una entrada pertenezca a las muestras reales o sintéticas. El mapeo aprendido por D es parametrizado por θ_d y se denota como $y = D(x; \theta_d)$. Las imágenes producidas forman una distribución $P_g(x)$ que va cambiando en el proceso de entrenamiento para parecerse a la distribución del conjunto de las imágenes reales $P_r(x)$. En la figura 5 se ilustra una GAN formada por dos redes convolucionales que

¹⁴ lbit., p 13.

utiliza datos de resonancias magnéticas cardiacas para ajustar la distribución del modelo generativo. El objetivo de D es diferenciar entre estos dos grupos de imágenes, mientras que el generador G es entrenado para confundir el discriminador D . De esta manera, el generador recibe una señal del discriminador indicando la calidad de las imágenes generadas. La información del gradiente es retropropagada de D hasta G , para que los parámetros de G sean modificados para producir una imagen que pueda engañar a D . El objetivo de entrenamiento de D y G

Figura 5. Marco de referencia GANs modelan la distribución de datos de resonancias magnéticas cardiacas.



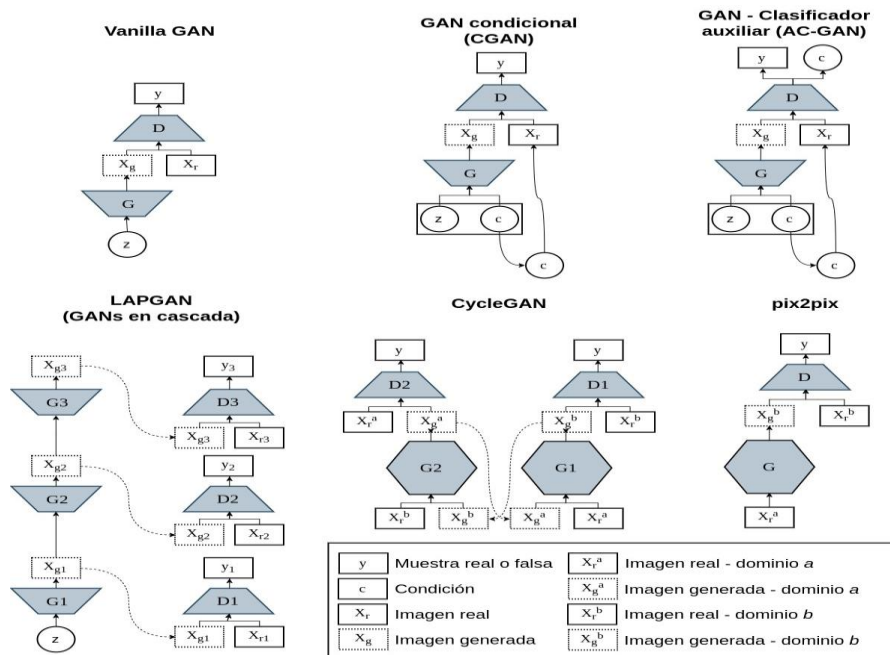
puede ser expresado matemáticamente en la ecuación 3.1.

$$\begin{aligned} \mathcal{L}_D^{GAN} &= \max_D E_{x_r \sim P_r(x)} [\log(D(x_r))] + E_{x_g \sim p_g(x)} [\log(1 - D(x_g))], \\ \mathcal{L}_G^{GAN} &= \min_G E_{x_g \sim P_g(x)} [\log(1 - D(x_g))] \end{aligned} \quad (3.1)$$

De la figura y ecuaciones anteriores se puede inferir que D es un clasificador binario con un objetivo de máxima probabilidad logarítmica. Si este es optimizado antes de que G modifique sus parámetros, minimizar L^{GAN}_G es equivalente a minimizar la divergencia Jensen-Shannon (JS) entre $P_r(x)$ y $P_g(x)$. El resultado final del entrenamiento de estas redes adversarias es que las muestras x_g que conforman el conjunto de entrenamiento, aproximen la arquitectura hacia la distribución de los datos reales $P_r(x)$.

En la actualidad existen múltiples variaciones a la primera arquitectura de GANs propuesta, la cual consiste en múltiples capas totalmente conectadas. Las variaciones presentadas en estas arquitecturas permiten a las GANs ser uno de los métodos más prometedores de los modelos generativos en diferentes campos de interés. En la figura 6 se encuentran algunas de las arquitecturas más relevantes de GANs.

Figura 6. Algunas de las arquitecturas más relevantes de las GANs. Imagen adaptada de *Generative Adversarial Network in Medical Imaging: A Review*



Fuente. YI, Xin; WALIA, Ekta; BABYN, Paul. Generative adversarial network in medical imaging: A review. *Medical image analysis*, 2019, p. 101552.

3.4 GANS EN SECUENCIAS CINE-MRI CARDIACAS

En términos de generación de secuencias cine-MRI cardíacas con GANs caben destacar trabajos como el propuesto por Zhang et al. En este se propone un método semi-supervisado Semi-Coupled-GAN, SCGANs¹⁵ que utiliza dos generadores (G_1, G_2) para aprender una representación de características para manejar un amplio rango de atributos visuales en las resonancias magnéticas cardíacas e identificar imágenes con una cobertura incompleta del ventrículo izquierdo. Estos dos generadores comparten un discriminador D y se entrenan de una manera adversaria utilizando la retropropagación. G_1 se encarga de generar los datos positivos, esto es, que pertenecen a cortes basales o apicales, mientras que G_2 se encarga de generar los datos negativos, que son imágenes que no pertenecen a un corte basal, apical o pertenecen a la distribución de imágenes sintéticas. D como en otros modelos de GANs, se encarga de mapear cada imagen de entrada a una probabilidad que indica si la resonancia cardíaca pertenece a los datos positivos o negativos. Este trabajo se centra en la importante tarea de generación de imágenes de resonancias magnéticas cardíacas de cortes apicales y basales, pero deja de lado el modelamiento de las diferentes condiciones cardíacas que son de gran interés. Además, esta arquitectura SCGANs es una arquitectura fija, definida desde el principio del entrenamiento, que acarrea problemas de convergencia y un costo computacional elevado. Por otro lado, Chartsias et al. propone una estrategia para la generación de nuevas imágenes de resonancias magnéticas y tomografías computadas cardíacas ¹⁶ utilizando una arquitectura CycleGAN que permite el

¹⁵ ZHANG, Le; GOOYA, Ali; FRANGI, Alejandro F. Semi-supervised assessment of incomplete LV coverage in cardiac MRI using generative adversarial nets. En International Workshop on Simulation and Synthesis in Medical Imaging. Springer, Cham, 2017. p. 61-68.

¹⁶ CHARTSIAS, Agisilaos, et al. Adversarial image synthesis for unpaired multi-modal cardiac data. En International Workshop on Simulation and Synthesis in Medical Imaging. Springer, Cham, 2017. p. 3-13.

cambio de dominio entre dos modalidades no emparejadas X y Y . Primero se aplican unas transformaciones en el conjunto de entrenamiento de la primera modalidad para asegurarse que la escala, posición y ángulo sean parecidos al de las imágenes en el conjunto de entrenamiento de la segunda modalidad. Luego, estas imágenes son usadas para entrenar la CycleGAN que, mediante un entrenamiento adversario con consistencia cíclica, logra transformar imágenes de una resonancia magnética a tomografía computada y viceversa. Todos los datos son posteriormente convertidos a una modalidad en específico y son utilizados para tareas de segmentación en busca de validar los resultados. La CycleGAN logra producir imágenes sintéticas de gran calidad utilizando como condición la imagen de un dominio fuente, mientras conserva características importantes como la estructura global del corazón. Sin embargo, hay propiedades que no se preservan y existe la posibilidad de que nuevos artefactos aparezcan con estos cambios de dominio¹⁷. Savioli et al. propone una arquitectura ROI-GAN¹⁸ que utiliza dos GANs que cooperan en dos campos de la imagen como la totalidad de esta y una región de interés. Al combinar características globales y locales, se preservan características más relevantes como la morfología de los ventrículos y una adecuada segmentación del corazón. Este método no introduce condiciones de entrada que limitan el aprendizaje de la arquitectura propuesta y da lugar a la producción o segmentación de imágenes cine-MRI cardiacas clínicamente incoherentes. Existen trabajos que abordan el problema de la superresolución de resonancias magnéticas cardiacas como el propuesto por Mahapatra et al., este

¹⁷ COHEN, Joseph Paul; LUCK, Margaux; HONARI, Sina. Distribution matching losses can hallucinate features in medical image translation. En International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2018. p. 529-536.

¹⁸ SAVIOLI, Nicolás, et al. A generative adversarial model for right ventricle segmentation. arXiv preprint arXiv:1810.03969, 2018.

propone una arquitectura P-SRGAN¹⁹ que toma una imagen en baja resolución (LR) como entrada para ser procesada a través de múltiples generadores, cada uno de estos escalando la imagen por 2. El trabajo del discriminador D es similar al de las GANs convencionales. Al generar cada imagen nueva en super resolución, los autores introducen un término adicional al cálculo del costo ya conformado por un error cuadrático medio y una función de costo CNN. Este término nuevo llamado la función de costo de Triplet se asegura de mejorar la imagen en super resolución mientras mantiene la nueva imagen distante de la de baja resolución. Esta arquitectura hace de las imágenes producidas en super resolución, cuantitativa y cualitativamente mejores. Este trabajo busca transformar imágenes de resonancias cardíacas afectadas por diferentes fuentes de ruido en imágenes más óptimas para tareas de clasificación y segmentación. Sin embargo, esta propuesta no está aumentando los datos, si no haciendo de los existentes, más adecuados para otras tareas canónicas. En ²⁰ se propone una arquitectura GAN llamada Reconstruction Global-Local GAN (Recon-GLGAN) para la reconstrucción de imágenes de resonancias magnéticas. El modelo propuesto contiene un generador U-Net y un discriminador de contexto que incorpora información global y local contextual de las imágenes que ofrecen un mejor desempeño en la tarea de reconstrucción. El discriminador de contexto consiste en un extractor global de características que opera en la totalidad de la imagen, un extractor local de características que recupera información de una región de interés (ROI) y un clasificador que toma como entrada las salidas de ambos extractores para tomar ventaja de la información contextual global y local de la imagen. Este discriminador incluye una función de costo que es

¹⁹ MAHAPATRA, Dwarikanath; BOZORGTABAR, Behzad; GARNAVI, Rahil. Image superresolution using progressive generative adversarial networks for medical image analysis. *Computerized Medical Imaging and Graphics*, 2019, vol. 71, p. 30-39.

²⁰ MURUGESAN, Balamurali, et al. Recon-GLGAN: A Global-Local Context Based Generative Adversarial Network for MRI Reconstruction. En *International Workshop on Machine Learning for Medical Image Reconstruction*. Springer, Cham, 2019. p. 3-15.

una combinación lineal de la función de costo adversaria y la función de costo L1 en el dominio de la imagen. El método propuesto permite una mejor reconstrucción en general, condicionando esta con información previa de un área de interés. Este método, al igual que los anteriores, dejan de lado información relevante de la patología cardiaca.

4. MÉTODO PROPUESTO

La estructura del corazón es esencial para una evaluación adecuada de la función cardiaca. Sin embargo, la compleja geometría y variabilidad de los ventrículos dificulta el modelamiento incluso en poblaciones con la misma afección cardiaca. En este trabajo se presenta una estrategia automática para generar imágenes cardiacas cine-MRI utilizando métodos de aprendizaje profundo. El enfoque aquí propuesto modela adecuadamente una distribución de entrenamiento formada por imágenes de resonancias magnéticas cardiacas diagnosticadas con diferentes patologías particulares. El novedoso entrenamiento adversario en las redes generativas adversarias (GANs) permite aprender características de un conjunto de datos para que una red neuronal convolucional produzca imágenes similares a las suministradas durante el entrenamiento. El enfoque condicional impone restricciones al entrenamiento de las GANs que ayuda a conservar atributos importantes de diferentes patologías cardíacas, y además ayudan a la convergencia de una mejor solución estabilizando así el modelo en el entrenamiento. El enfoque progresivo es capaz de modelar con gran precisión la estructura global y local de una resonancia magnética cardiaca, generando imágenes con coherencia clínica de alta resolución y ayudando a reducir el tiempo computacional. El desarrollo del enfoque propuesto es descrito en detalle en las siguientes subsecciones.

4.1 CONDICIONANDO LA GENERACIÓN DE IMÁGENES CON CONDICIONES CARDIACAS.

Las GANs son el estado del arte de los modelos generativos para una estimación de verosimilitud. Sin embargo, un problema grande en el entrenamiento de estas GANs es el colapso a una distribución específica, lo que limita la diversidad de las

muestras generadas ²¹. En general, los modelos adversarios consisten en dos redes opuestas (G , D) que representan funciones de mapeo no lineales. Durante el entrenamiento, el generador $G(Z)$ produce una imagen de un vector aleatorio de ruido z , mientras que el discriminador estima la probabilidad $P(x_{\text{new}} | X_{\text{tr}}, X_g)$ de que una imagen x_{new} pertenezca a los datos de entrenamiento X_{tr} o a los generados X_g . Esta estrategia adversaria maximiza la probabilidad $\min_G \max_D E_{x \sim p_{\text{data}}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(x)))]$.

Desafortunadamente, optimizar la función mencionada no garantiza la generación de imágenes realistas. Además, preservar la morfología asociada a enfermedades cardíacas específicas es de gran interés. Las GANs condicionales han emergido para tratar con estas limitaciones integrando información previa y en una representación oculta conjunta con el vector z , y asignada a la función discriminativa. En este caso, cada muestra entrenada X_{tr_i} tiene asociada una etiqueta de una patología diagnosticada y_{tr_i} , formando la tupla $(\mathbf{X}_{\text{tr}}, \mathbf{Y}_{\text{tr}})$. Un clasificador auxiliar (AC-GAN) fue implementado para recuperar volúmenes cine-MRI con sentido morfológico. La AC-GAN genera distribuciones independientes para imágenes de entrada x_{new} y etiqueta asociada y_{new} , para el generador $G(\mathbf{X}_{\text{tr}}, \mathbf{Y}_{\text{tr}})$, mientras que el discriminador integra ambas fuentes como probabilidades independientes: $P(x_{\text{new}} | X_{\text{tr}}, X_g)$, $P(y_{\text{tr}_i} | X_{\text{tr}}, X_g)$, siendo X_g las imágenes generadas parcialmente. En tal sentido, la función de estimación de verosimilitud es reescrita para depender de la tupla $G(\mathbf{X}_{\text{tr}}, \mathbf{Y}_{\text{tr}})$, como: $\min_{G_{I,y}} \max_{D_{I,y}} \underbrace{V(D_{I,y}, G_{I,y})}_{\hat{B}(D,G)}$. De tal suposición, la AC-GAN logra un espacio latente independiente de cada etiqueta de clase, ayudando a la estabilización del entrenamiento.

²¹ GOODFELLOW, Ian. NIPS 2016 tutorial: Generative adversarial networks. arXiv preprint arXiv:1701.00160, 2016.

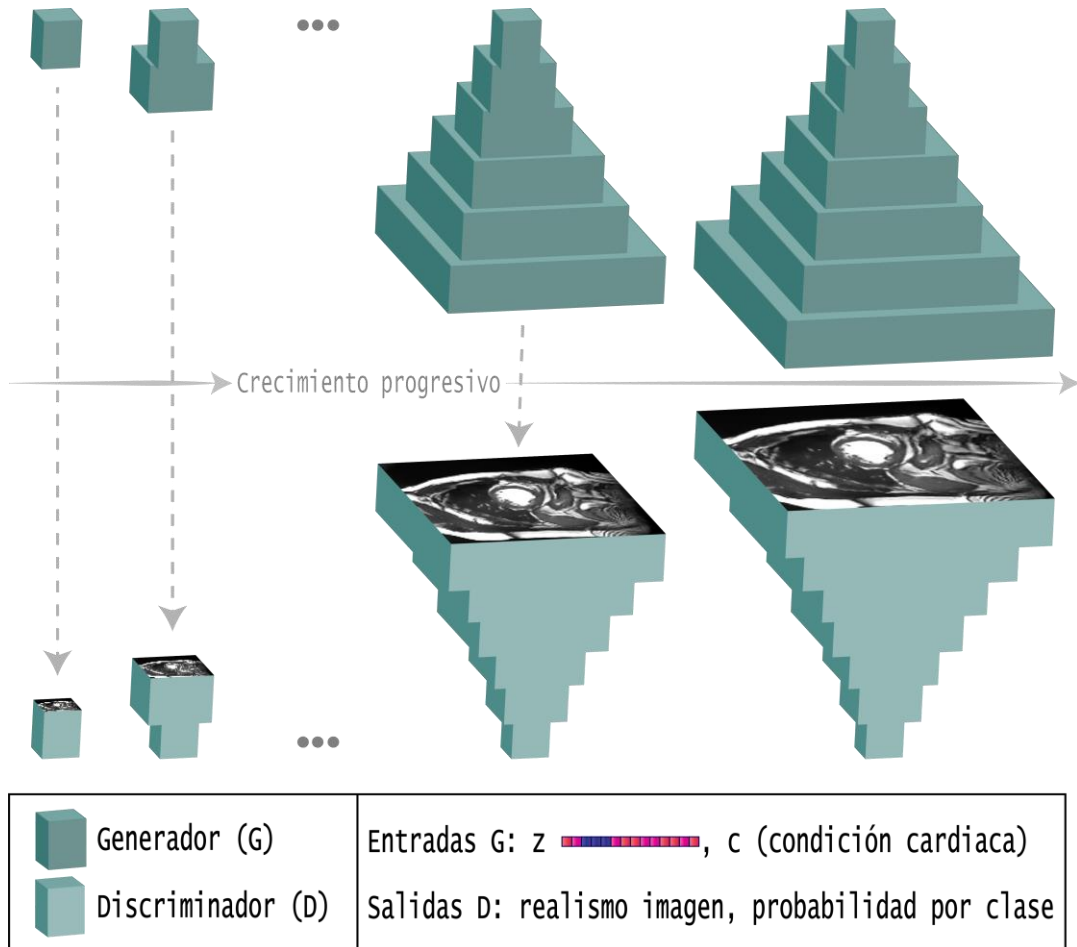
4.2 ESTRATEGIA PROGRESIVA ADVERSARIA: PGAN

Últimamente, una estrategia progresiva de entrenamiento fue adoptada en este trabajo para aprender la GAN condicional²². Esta estrategia empieza resolviendo la generación de imágenes de baja resolución cine-MRI, y progresivamente añade capas a la red para incrementar la resolución espacial de las imágenes resultantes (ver Figura 7).

De este esquema, la gran variación de la representación estructural del corazón es tratada en un marco multiescala, donde primero surge la gran escala de la imagen y luego se enfoca la atención en detalles de escala más finos. En cada escala de la tupla (G, D) , en cada paso progresivo, son enfrentados a un entrenamiento adversario independiente, donde G aprende el mapeo $z \rightarrow x_g$. La salida del generador, x_g se espera que sea cercana a la distribución real $p_r(x)$. La PGAN implementada empieza con una resolución espacial de la imagen de 4×4 y va incrementando su tamaño hasta lograr imágenes de 256×256 . Esta estrategia es más estable que las estrategias clásicas, recuperando la variabilidad del corazón y siendo más eficiente en tiempo computacional. El algoritmo 1 resume el proceso de entrenamiento progresivo.

²² KARRAS, Tero, et al. Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196, 2017.

Figura 7. Estructura del enfoque propuesto para generación de resonancias magnéticas cardiacas. La arquitectura convolucional de ambas redes neuronales va creciendo progresivamente a medida que cada nivel de detalle llega a un punto óptimo.



Algorithm 1 Progressive AC-adversarial (PGAN) learning

Input: $(L_0, \hat{\mathbf{B}}(D^{L_0}, G^{L_0}))$

Output: $\hat{\mathbf{B}}(D^{\mathbf{L}}, G^{\mathbf{L}})$ with $\mathbf{L} = L_0 \cup L_1 \dots \cup L_T$

- 1: function PGAN_Learn($L_i, \hat{\mathbf{B}}(D^{L_i}, G^{L_i})$)
 - 2: if $L == L_{T+1}$ then
 - 3: return $\hat{\mathbf{B}}(D^{\mathbf{L}}, G^{\mathbf{L}})$
 - 4: return PGAN_Learn($L_{i+1}, \hat{\mathbf{B}}(D^{L \cup L_{i+1}}, G^{L \cup L_{i+1}})$)
-

La implementación de esta estrategia requiere que el generador y discriminador (G , D) crezcan en sincronía y que todas las capas puedan entrenarse aún con la

introducción de nuevas capas convolucionales. La introducción de estos nuevos bloques debe hacerse gradualmente para evitar cambios bruscos en las capas previamente entrenadas. Esta introducción gradual de un nuevo bloque de capas convolucionales requiere usar una conexión de omisión (skip connection) para conectar un nuevo bloque de la entrada del discriminador o la salida del generador y añadir este a la entrada o salida existentes con una ponderación. En las siguientes subsecciones se detalla el proceso de crecimiento de ambas redes convolucionales junto con sus parámetros de entrenamiento.

4.2.1 Añadiendo capas al generador. Para el generador G , el proceso de crecimiento consiste en añadir un nuevo bloque de capas convolucionales que generen una imagen de resolución espacial más grande. La salida de la capa agregada G^L , es combinada con el resultado obtenido de aumentar espacialmente la salida de la capa anterior G^{L-1} utilizando el método de interpolación por el vecino más cercano (nearest-neighbor interpolation en inglés). La contribución de la capa redimensionada está ponderada por un factor $(1 - \alpha)$, mientras que la contribución de la nueva capa está ponderada por α . Alpha (α) es un valor pequeño inicialmente, dando al inicio más importancia a la versión redimensionada de la capa G^{L-1} y con el paso de las iteraciones dando más importancia a la capa añadida G^L . Durante la transición se tratan las capas que operan en más alta resolución como un bloque residual, cuya importancia α incrementa linealmente desde 0 hasta 1. En la figura 8 se aprecia el proceso de crecimiento del generador.

4.2.2 Añadiendo capas al discriminador. Para el discriminador D , el proceso de crecimiento involucra añadir nuevos bloques de capas convolucionales a la entrada del modelo para admitir las nuevas imágenes de mejor resolución producidas por el generador G . La imagen de entrada es reducida espacialmente utilizando una agrupación por promedio (average pooling) para que esta pueda pasar por la siguiente capa D^{L-1} . La salida del bloque D^L también es reducida espacialmente utilizando agrupación por promedio para que esta pueda ser tomada en cuenta como entrada para el ya existente bloque D^{L-1} . Las dos versiones reducidas espacialmente de la entrada son combinadas en una suma ponderada, comenzando con una importancia total para la imagen de entrada y transicionando linealmente hasta llegar a dar total importancia al nuevo bloque de entrada D^L . En la figura 9 se aprecia el proceso de crecimiento del discriminador.

4.2.3 Incrementando la variación de los datos producidos. Las GANs tienden a capturar un subconjunto pequeño de la variación de los datos de entrenamiento.

Figura 8. Introducción gradual de nuevas capas al generador. Antes de la transición (1), durante la transición (2) y después de la transición (3). El bloque 2x significa aumentar espacialmente al doble. El bloque toRGB representa una capa que proyecta vectores d .

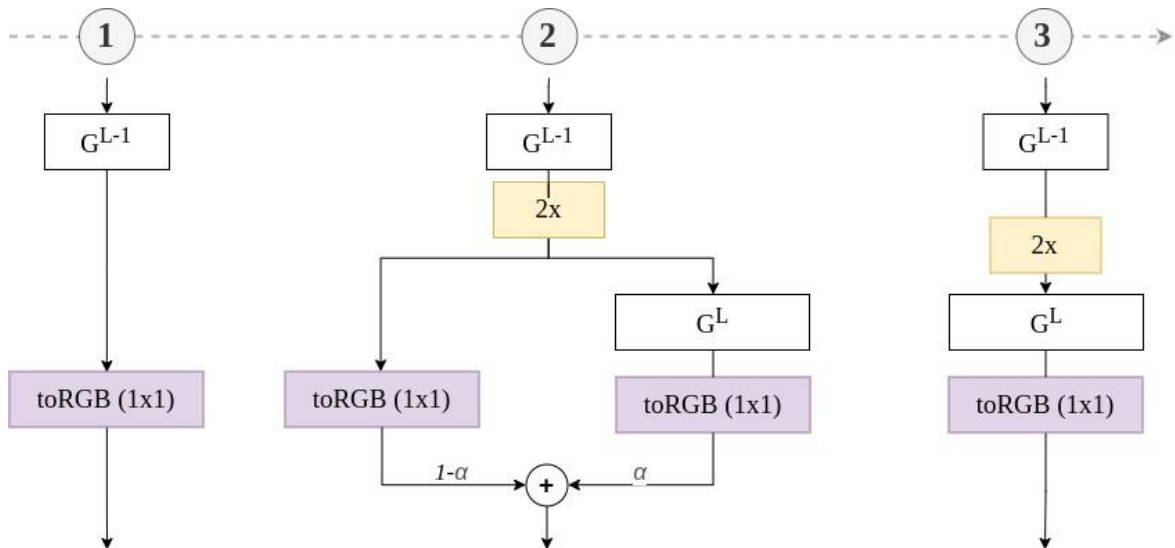
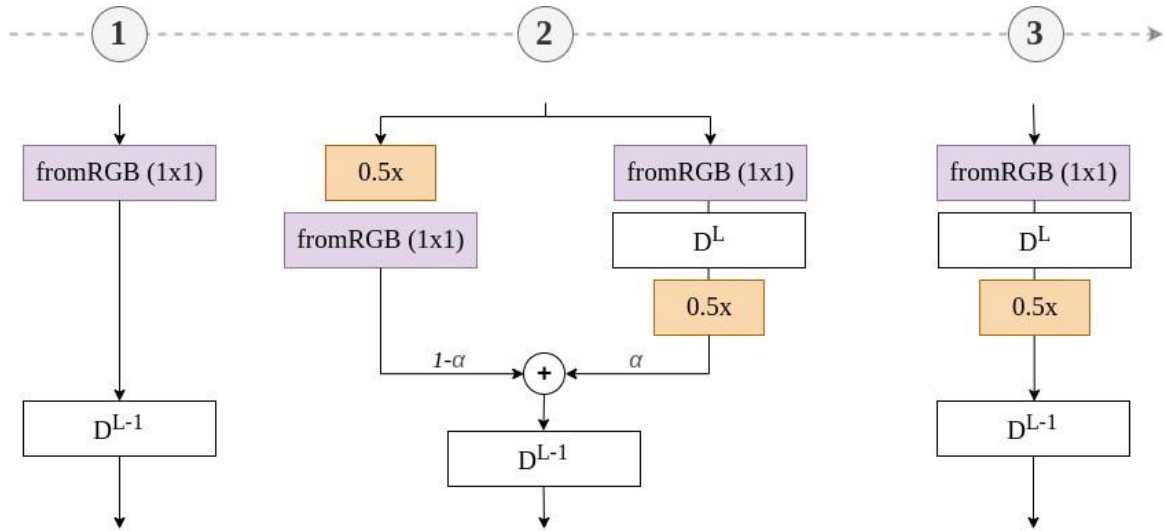


Figura 9. Introducción gradual de nuevas capas al discriminador. Antes de la transición (1), durante la transición (2) y después de la transición (3). El bloque 0.5x significa reducir espacialmente a la mitad. El bloque fromRGB representa una capa que proyecta



cionar esto se implementó una solución que no tiene parámetros a aprender o hiperparámetros que deban ser ajustados por el usuario. Primero se calcula una desviación estándar para cada una de las características en cada localización espacial a través de un minibatch, luego se promedian estos estimados a través de todas las características y localizaciones espaciales para obtener un único valor. Este valor se replica y concatena a todas las localizaciones espaciales y a través del minibatch, produciendo un nuevo mapa de características que se inserta al final del discriminador.

4.2.4 Aprendizaje en el generador y discriminador. Existe una necesidad por limitar las magnitudes de los gradientes y la competencia en el entrenamiento de las GANs. Para esto se utilizó una inicialización con una distribución normal aleatoria $N(0,1)$ para los pesos de las diferentes capas de la red y luego escalar estos pesos durante el tiempo de ejecución. El beneficio de hacer esto dinámicamente en lugar de en la inicialización se relaciona con la invariancia de escala en el descenso de gradiente estocástico adaptivo comúnmente usado en métodos como RMSProp y Adam²³. Junto con el método de inicialización mencionado anteriormente, se incluyó en el generador una normalización del vector de características en cada píxel para evitar el escenario donde las magnitudes en el generador y discriminador se salen de control como resultado de la competencia entre ambas redes.

4.3 GENERANDO ESPACIO LATENTE

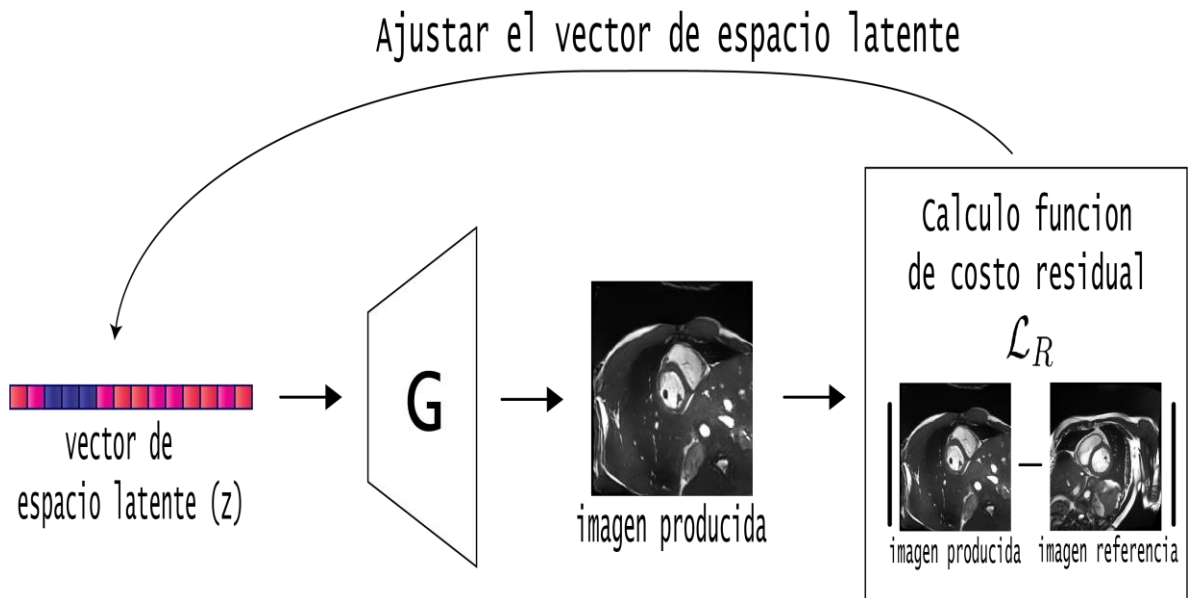
Después del proceso de entrenamiento, el generador $G(z): z \rightarrow (x_n, y_n)$ aprende un mapeo adecuado de un vector aleatorio a una imagen cine-MRI con una etiqueta asociada. En este sentido, el esquema adversario aprende la variedad χ subyacente a la estructura de variabilidad de las imágenes de diferentes patologías. Sin embargo, un problema grande es que el mapeo inverso (de una imagen particular a el espacio latente: $(x_n, y_n) \rightarrow z$) es desconocido. Para obtener un vector embebido de una imagen específica, se implementó una función de costo residual que determina el mejor vector z_j ²⁴. Este proceso utiliza el algoritmo de retropropagación con un total de pasos κ . El método calcula una función de costo definida entre la imagen sintética $G(z_0)$, para un vector aleatorio z_0 , y una imagen objetivo x_n . Esta función de costo es minimizada utilizando la retropropagación al proveer los gradientes para actualizar los coeficientes del vector latente ($z_0 \rightarrow z_1$). Después de

²³ Ibít., p. 30

²⁴ RADFORD, Alec; METZ, Luke; CHINTALA, Soumith. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434, 2015.

varios pasos de retropropagación, el vector embebido z_i , está lo suficientemente cerca a la imagen de representación x_n en la variedad χ . La función de costo implementada es descrita como $L_R(z_\gamma) = \rho \|x_n - G(z_\gamma)\|^2$,²⁵ donde γ son la cantidad de pasos de retropropagación necesarios para obtener el vector embebido. En la Figura 10, el proceso de generación de vectores embebidos durante κ pasos iterativos es ilustrado.

Figura 10. Estructura general del enfoque propuesto para la recuperación del vector de espacio latente embebido z a través de la retropropagación utilizando una función de costo residual.



4.4 CONFIGURACIÓN EXPERIMENTAL

²⁵ SCHLEGL, Thomas, et al. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. En International Conference on Information Processing in Medical Imaging. Springer, Cham, 2017. p. 146-157.

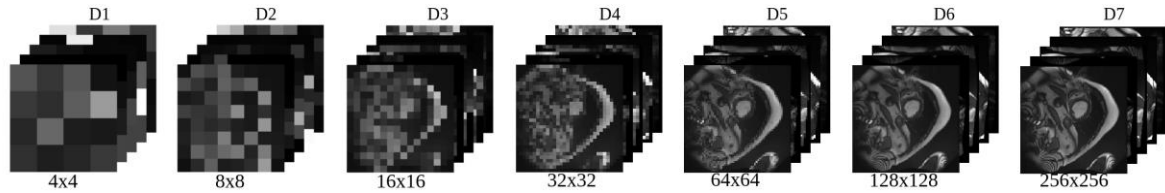
4.4.1 Dataset. El enfoque propuesto fue entrenado y validado utilizando datos del dataset público del desafío (The MICCAI 2017 automated cardiac diagnosis challenge (ACDC)) con un total de 150 volúmenes cine-MRI con vistas de eje corto (SA, por sus siglas en inglés)²⁶ [20]. Cada volumen es asociado a una de las siguientes cinco condiciones cardíacas diferentes: cardiomiopatía dilatada (DCM), cardiomiopatía hipertrófica (HCM), normal (NOR), falla cardíaca con infarto (MINF), o con un ventrículo derecho anormal (RV). Para cada condición cardíaca presente en el conjunto de datos existen 30 volúmenes cine-MRI.

Los datos fueron divididos en tres conjuntos diferentes: 70% para el entrenamiento de la PGAN, 10% para la validación de la PGAN y 20% fueron dejados para el conjunto de prueba. Cada conjunto de datos fue cuidadosamente seleccionado para asegurar la misma distribución de afecciones cardíacas. El objetivo principal del enfoque presentado es generar imágenes cineMRI condicionadas por una caracterización confiable de la patología. Por lo tanto, solo cortes basales y medios fueron considerados. Un conjunto total de 14515, 3691 y 2296 fueron usadas para los conjuntos de entrenamiento, validación y prueba respectivamente. Cada imagen de los tres conjuntos de datos se procesó para que coincidiera con cada uno de los niveles de detalle de la arquitectura progresiva GAN como se muestra en la Figura 11.

²⁶ BERNARD, Olivier, et al. Deep learning techniques for automatic MRI cardiac multistructures segmentation and diagnosis: Is the problem solved?. IEEE transactions on medical imaging, 2018, vol. 37, no 11, p. 2514-2525.

4.4.2 Configuración de entrenamiento PGAN. El entrenamiento de la estrategia PGAN fue llevado a cabo utilizando niveles de resolución de: $\{(4 \times 4), (8 \times 8), (16 \times 16), (256 \times 256)\}$ pixeles.

Figura 11. Ejemplos de imágenes de resonancias magnéticas en diferentes dimensiones espaciales para el entrenamiento de la GAN progresiva.



Un vector de espacio latente de dimensión 512 fue usado como entrada para el generador. El tamaño del batch utilizado va decreciendo progresivamente: un tamaño de 128 fue usado para los tres primeros niveles, 64 para el segundo grupo de tres niveles y 8 para la última capa de alta resolución. La función de costo para el entrenamiento de la GAN progresiva es la función de costo Wasserstein utilizando una penalización de gradientes para hacer cumplir la restricción de Lipschitz ²⁷.

²⁷ GULRAJANI, Ishaan, et al. Improved training of wasserstein gans. En Advances in neural information processing systems. 2017. p. 5767-5777.

4.4.3 Evaluación con métricas de similitud aparente y similitud entre distribuciones de entrenamiento y sintética. La calidad de las imágenes de resonancias magnéticas cardiacas producidas por la GAN es determinante para la posterior aplicación de métodos de segmentación, detección, registro y reconstrucción. Esta calidad puede ser medida por un cardiólogo experto para determinar la coherencia de las estructuras globales y locales de las imágenes producidas. Sin embargo, este método resulta ineficiente y poco escalable cuando se tienen grandes volúmenes de datos. Como solución al problema anterior de medición, se propone utilizar diferentes métricas para cuantificar la similitud estructural de las imágenes reales y producidas y también medir la distancia entre las distribuciones de entrenamiento P_r y la distribución generadas P_g . En las siguientes subsecciones se explican las métricas utilizadas y el porqué de su aplicación.

4.4.3.1 MS-SSIM. La métrica de similitud estructural multiescala²⁸ (MS-SSIM, por sus siglas en inglés) es una variante multiescala de la métrica de similitud estructural (SSIM). Esta métrica calcula un índice para cuantificar la similitud entre dos imágenes. Este enfoque es un modelo basado en la percepción que considera la degradación de la imagen como el cambio percibido en la información estructural. La información estructural es la idea de que los pixeles tienen dependencias fuertes cuando estos están espacialmente cerca uno de otro. Estas dependencias contienen información importante acerca de la estructura de los objetos y suelen tener más importancia en análisis multiescala como el efectuado por MS-SSIM. La aplicación de esta métrica en imágenes de resonancias magnéticas cardiacas es sensato, ya que estas imágenes son altamente estructuradas. Los valores del MS-SSIM están entre 0.0 y 1.0, donde valores cercanos a cero corresponden a imágenes perceptualmente más parecidas. Intuitivamente, los valores cercanos a 0 indican que las imágenes conservan la estructura global y local de los múltiples objetos encontrados en una resonancia magnética cardiaca.

²⁸ WANG, Zhou, et al. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 2004, vol. 13, no 4, p. 600-612.

4.4.3.2 Sliced Wasserstein Distance. La Wasserstein Distance (WD) propuesta por Villani et al.²⁹, es una métrica utilizada en la optimización de las GANs durante el entrenamiento y también en la evaluación de la distribución que forman las imágenes generadas $P_g(x)$. Esta métrica es comúnmente conocida como Earth Mover's Distance (EMD), que intuitivamente es el mínimo desplazamiento de masa para transformar una distribución en otra. La distancia Wasserstein funciona cuando la distancia base es calculada a partir de espacios de baja dimensionalidad, de otro modo esta resulta intratable. Por lo anterior surge como alternativa la Sliced Wasserstein Distance (SWD) que aproxima la distancia Wasserstein-1 entre las imágenes reales y sintéticas. Esta SWD es calculada como la similitud estadística entre parches locales de la imagen, extraídos de representaciones de la pirámide laplaciana de estas imágenes. Los valores para esta métrica están dentro del rango $[0, \infty]$ y son indicadores de que tan cerca está la distribución de imágenes sintéticas de la distribución de imágenes reales. Los valores cercanos a cero de una métrica SWD indican que las distribuciones son cercanas, mientras que los valores grandes indican que las distribuciones están alejadas, es decir, las imágenes generadas tienen pocas características de las reales.

²⁹ VILLANI, Cédric. Optimal transport: old and new. Springer Science & Business Media, 2008.

4.4.4 Evaluación con tareas de clasificación. En este trabajo, el interés no está dirigido exclusivamente a evaluar la calidad en general de las imágenes generadas, también en la propiedad de retención de información del espacio latente asociado a diferentes condiciones cardíacas. Para hacer esto, se definieron algunas tareas de clasificación actuando como una evaluación sustituta de esta propiedad. Esta evaluación fue llevada a cabo en dos diferentes tareas supervisadas utilizando el conjunto de datos de prueba: (a) Tarea de clasificación multiclase (5 clases) y (b) Cuatro diferentes tareas de clasificación binarias de condiciones del corazón (Normal contra todos). El conjunto de datos de prueba se dividió en dos grupos, para el entrenamiento supervisado se asignó un 70% y para la evaluación un 30%. No hubo superposición a nivel de paciente entre estos grupos.

La evaluación comenzó generando el conjunto de datos de vectores embebidos como fue descrito en la subsección 2.3. Para cada imagen, la recuperación del vector de espacio latente asociado fue llevada a cabo utilizando 5000 iteraciones. Luego, la estrategia de reducción de dimensionalidad UMAP fue aplicada [24]. UMAP es una técnica de reducción de dimensionalidad no supervisada que genera una función de mapeo T desde la alta dimensión R^n hasta el espacio de baja dimensión R^2 limitado para preservar las distancias locales de los vectores de características de acuerdo con una métrica definida por el usuario en el espacio de alta dimensionalidad. Un conjunto de clasificadores Random Forest fueron entrenados en el espacio de baja dimensionalidad para las cuatro tareas de clasificación binarias y para la tarea de clasificación multiclase. Luego, los puntos del espacio bidimensional (2D) fueron usados para entrenar el clasificador en un entrenamiento supervisado. Luego del entrenamiento, los puntos sobrantes del espacio bidimensional fueron usados para la evaluación. Finalmente, la matriz de confusión y el puntaje de precisión fueron calculados para las cuatro tareas de clasificación binarias y la tarea de clasificación multiclase.

5. EVALUACIÓN Y RESULTADOS

Una vez finalizado el entrenamiento de la PGAN, se utilizó el generador para producir algunos ejemplos para cada condición cardíaca NOR, DCM, HCM, MINF y RV (ver Figura 12). La GAN progresiva y condicionada logra resultados visuales confiables, que preservan la estructura cardíaca de los ventrículos. Además, se debe notar que, para diferentes patologías, notables diferencias se pueden observar que pueden asociarse a la naturaleza de las condiciones cardíacas. Algunas imágenes ruidosas y de intensidades alteradas también fueron producidas como se puede observar en la sexta fila. Sin embargo, este tipo de imágenes también se encuentran en el conjunto de datos de entrenamiento debido a protocolos de adquisición, lo que sugiere un buen modelamiento de la distribución de los datos de entrenamiento por parte de la PGAN. Como comparación, imágenes cardíacas sintéticas fueron generadas utilizando una estrategia no condicional (ver Figura 13). Las imágenes generadas contienen artefactos y en algunos casos falta información anatómica. En otros casos, imágenes totalmente incoherentes son generadas que contienen estructuras cardíacas adicionales como ventrículos se pueden observar. Estas incoherencias pueden ser producto de no limitar el aprendizaje de las resonancias magnéticas con una condición.

Para evaluar la calidad visual de las imágenes producidas por la PGAN, utilizamos la métrica MS-SSIM que realiza un análisis multiescala de la estructura de las imágenes producidas con respecto a las imágenes reales. Para esto, se generaron 20000 imágenes por medio del generador entrenado de la PGAN y se compararon estas imágenes sintéticas contra las imágenes reales del conjunto de entrenamiento. Para evaluar el progreso de nuestro método, replicamos el proceso de producción de las 20000 imágenes por cada uno de los checkpoints del entrenamiento de la PGAN. En la figura 14 se pueden visualizar los nombres de los checkpoints de la PGAN sobre el eje X junto con su índice MS-SSIM en el eje Y. El

MS-SSIM de las imágenes reales significa que tan estructuralmente relacionadas están las imágenes del conjunto de entrenamiento entre ellas. Los valores del eje X simbolizan la cantidad de imágenes en miles que han pasado por la arquitectura en ese momento, es decir, un valor de 12000 significa que han pasado 12000000

Figura 12. Ejemplos de resonancias magnéticas generadas.

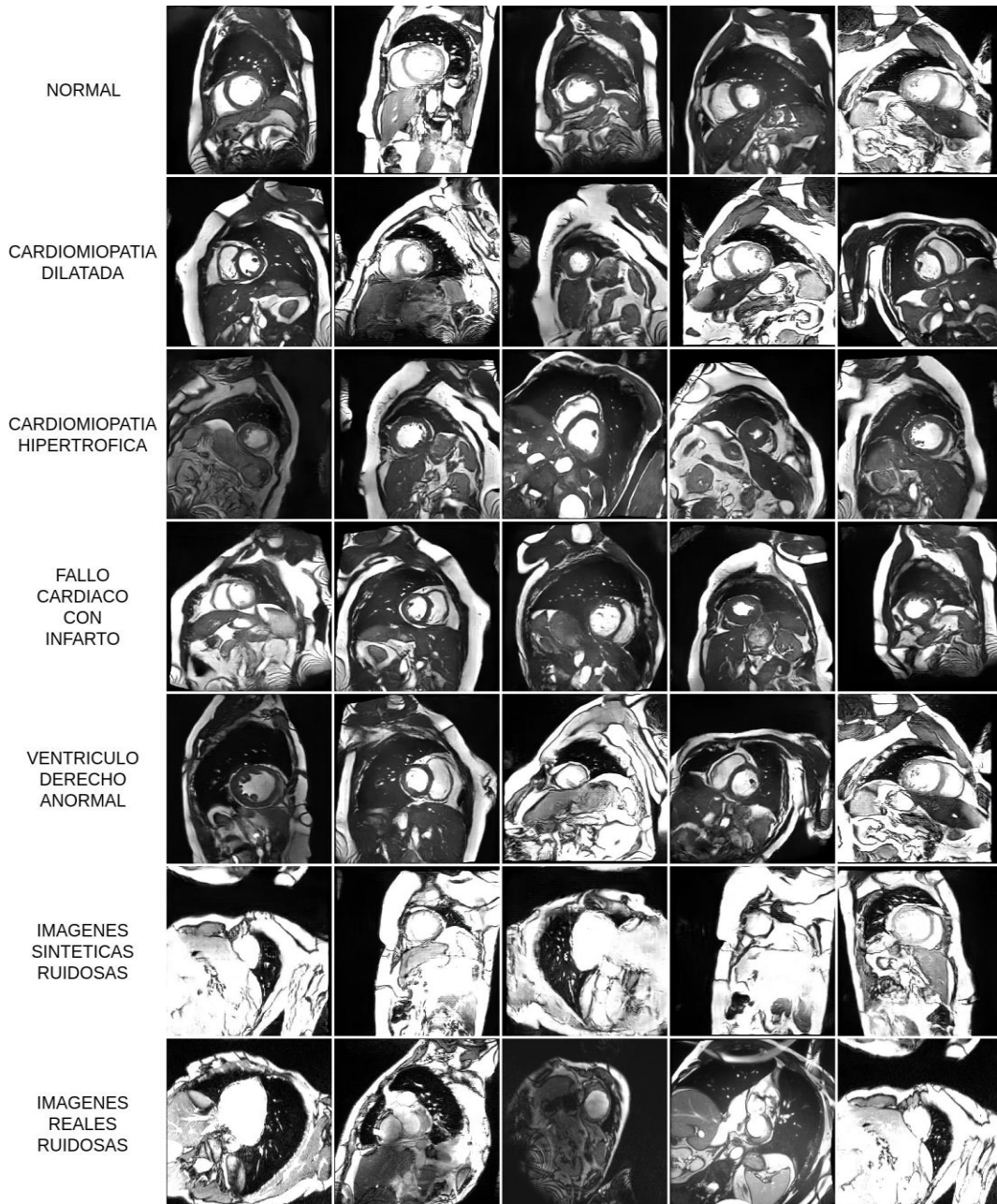
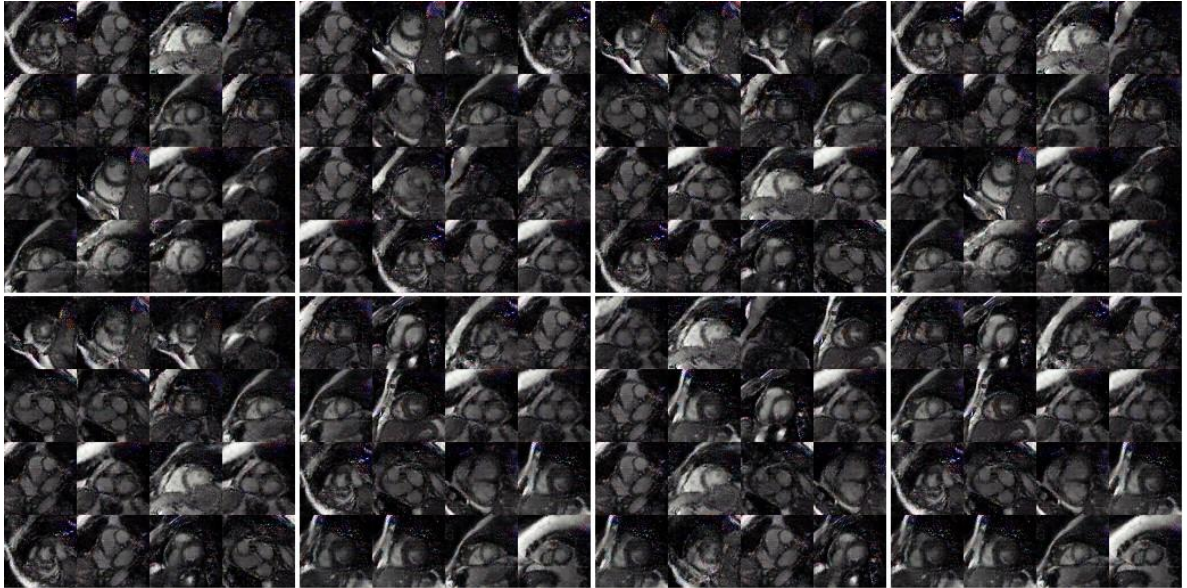


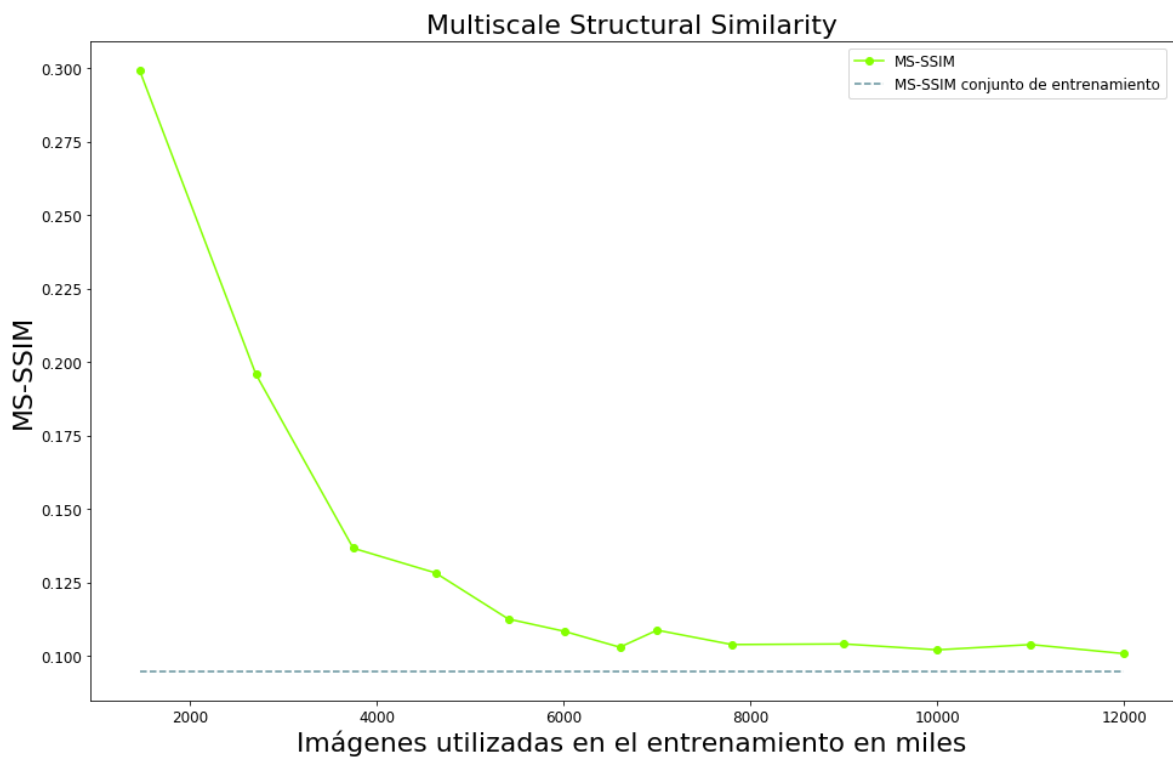
Figura 13. Ejemplos de resonancias magnéticas generadas utilizando una estrategia GAN sin condición.



(doce millones) de imágenes durante el entrenamiento. Por último, los índices MS-SSIM para cada uno de los checkpoints significan qué tan estructuralmente correlacionadas están las imágenes generadas por la PGAN en ese instante del entrenamiento. Se espera correctamente que los coeficientes MS-SSIM bajen a medida que el número de imágenes utilizadas para el entrenamiento aumenta, indicando así un aumento en la variedad de las muestras producidas. La calidad estructural y visual son buenas medidas para comparar las imágenes producidas con las reales. Sin embargo, estas medidas dejan por fuera conceptos como la variación de las imágenes producidas y distribución de ambos conjuntos de datos. Para obtener un valor que indique la proximidad de la distribución de datos de entrenamiento y la distribución de las imágenes producidas, se utilizó la métrica Sliced Wasserstein Distance (SWD). Para el cálculo de esta métrica se utilizaron 16384 imágenes producidas por el generador. Con las imágenes producidas por el generador, se creó un conjunto de datos nuevo para los niveles de detalle 128, 64,

32, 16. Estos conjuntos de datos nuevos, junto con el de las imágenes producidas por el generador, suman 5 conjuntos de datos que se compararon con las imágenes de entrenamiento. Para evaluar el progreso de nuestro método, replicamos este proceso de generación de los 5 conjuntos de datos por cada checkpoint del entrenamiento de la PGAN. En la Figura 15 se muestran los valores SWD en los diferentes niveles de detalle para cada checkpoint registrado en el proceso de entrenamiento. Se espera de manera acertada que a medida que la PGAN procese más imágenes en el entrenamiento, los valores de SWD se acerquen a cero. Algo a notar

Figura 14. Similitud estructural entre las imágenes generadas en diferentes etapas del entrenamiento y las imágenes del conjunto de entrenamiento.

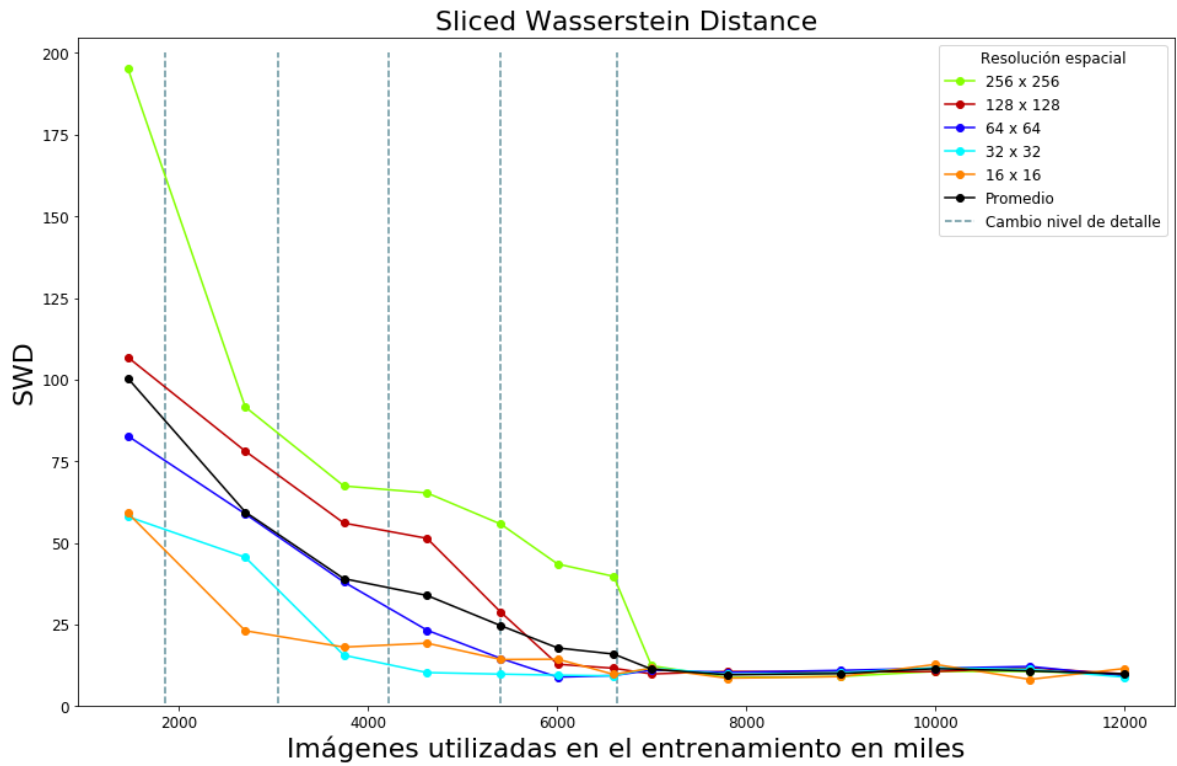


es que, aunque en las primeras etapas del entrenamiento no se tienen todos los niveles de detalle para producir imágenes en la resolución final, la optimización de estas capas inferiores ayuda a reducir el puntaje SWD de las resoluciones mayores,

lo cual confirma que el aprendizaje progresivo es útil y comienza por los detalles más generales de las resonancias magnéticas. Luego de evaluar la calidad estructural de las imágenes sintéticas y la distancia entre ambas distribuciones, se ilustran las visualizaciones UMAP 2D para el conjunto de prueba. Bajo la suposición que el espacio latente tiene transiciones suaves, se espera de manera acertada que los puntos más cercanos en el espacio latente generen imágenes cine-MRI parecidas. Además, es probable que el componente condicional haga énfasis en este comportamiento. La Figura 16 también muestra que las diferentes clases son adecuadamente mapeadas en el espacio de agrupamiento con algunos valores naturales atípicos que corresponden a condiciones de adquisición de la imagen que también son forzadas a aprender y que pierden algunas caracterizaciones particulares de la morfología del corazón.

Finalmente, la Figura 17 ilustra los resultados para la tarea de clasificación multiclase y la tarea de clasificación binaria. El clasificador Random Forest (RF) entrenado en los vectores embebidos 2D, extraídos del vector de características de dimensión 512, preservó suficiente

Figura 15. Distancia entre la distribución de entrenamiento P_r y la distribución de las imágenes generadas P_g en diferentes etapas del entrenamiento.



información para distinguir entre diferentes condiciones cardiacas. En algunos casos, el modelo RF entrenado, arroja medidas de precisión por encima del 90% para las diferentes tareas de clasificación binaria. Mientras que, el modelo RF entrenado para la tarea multiclase logró una precisión del 80%.

Figura 16. Agrupamiento no supervisado UMAP de los vectores embebidos condicionados. Un agrupamiento multiclase y biclase se llevó a cabo en diferentes experimentos.

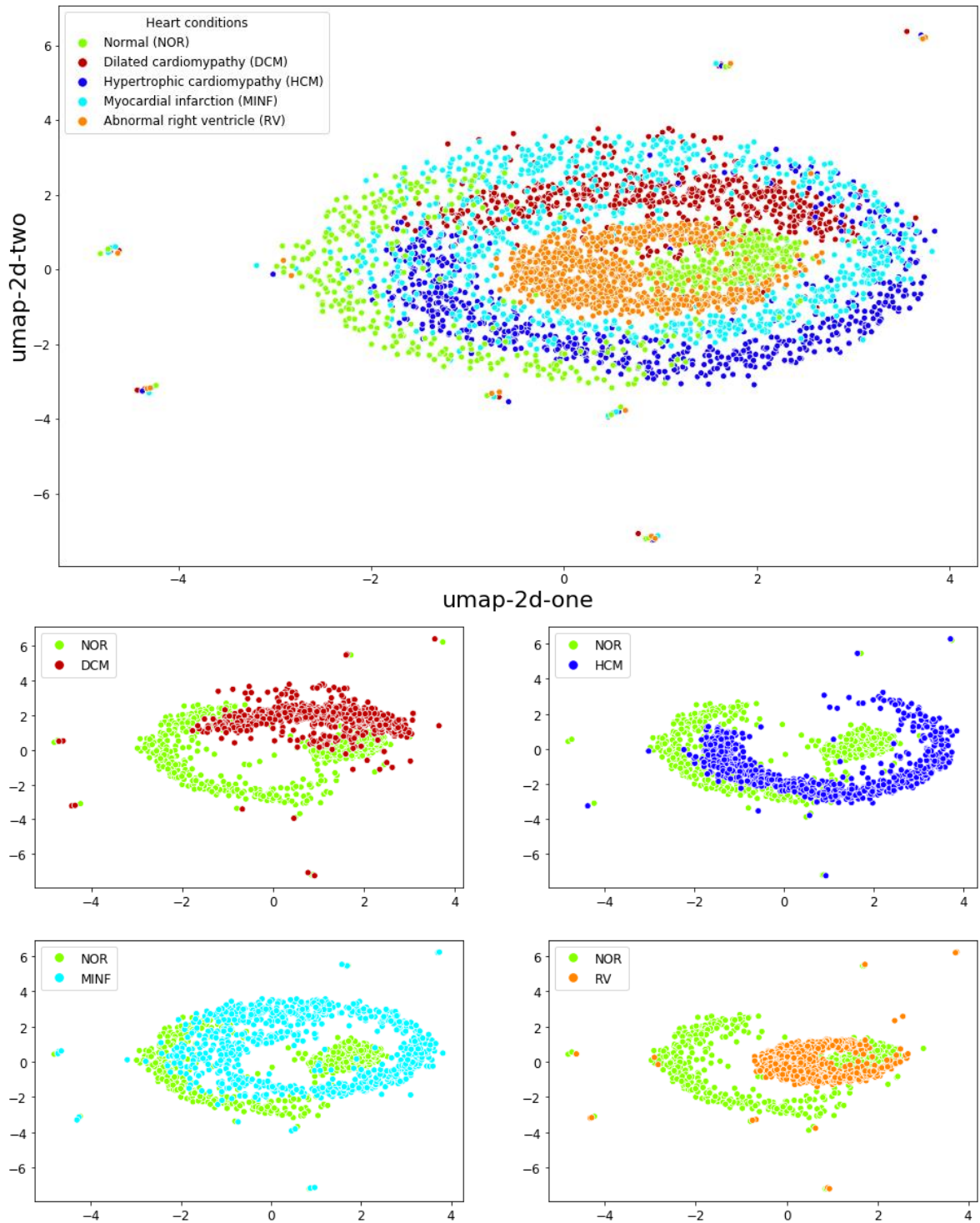
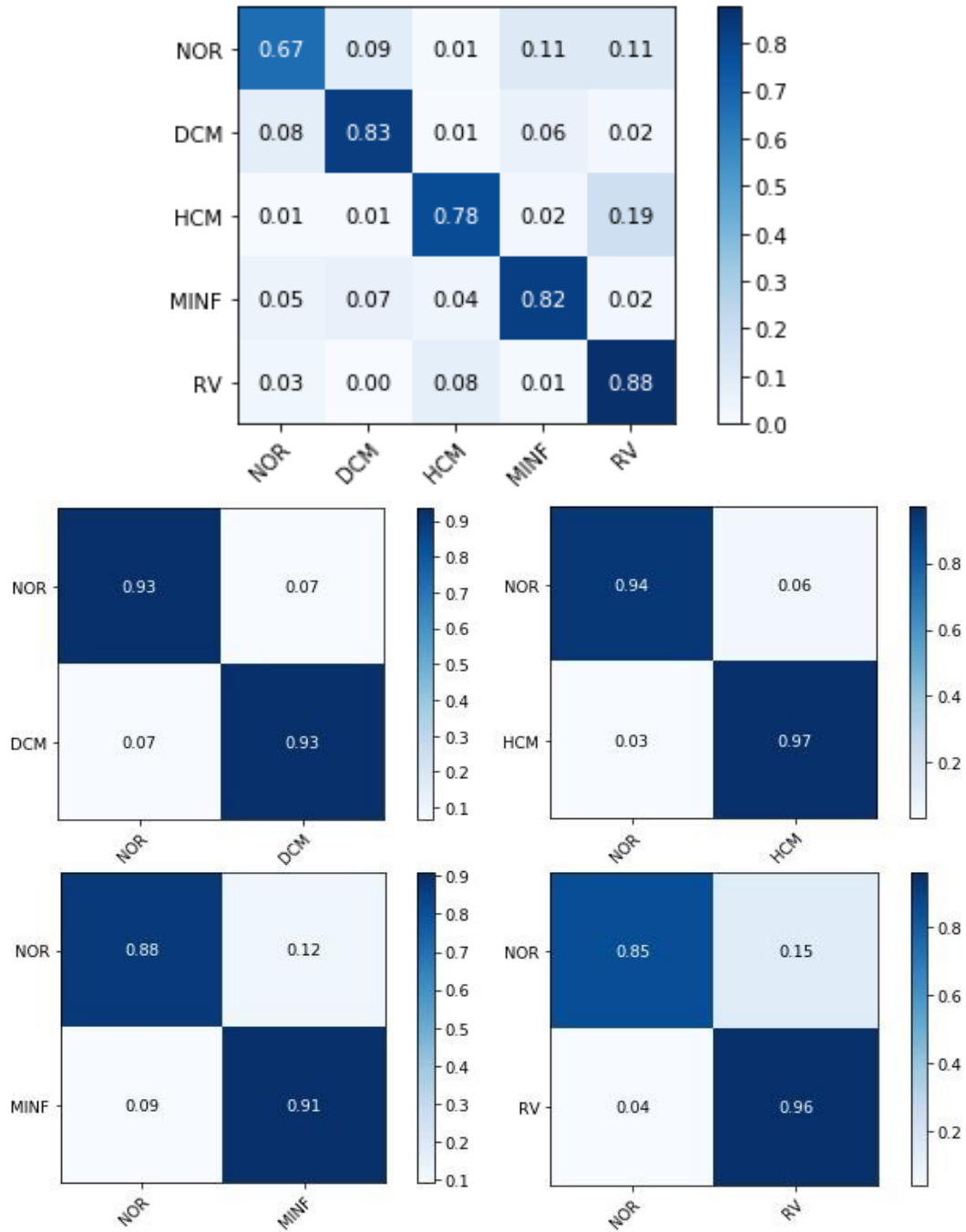


Figura 17. En la parte superior: tarea de clasificación multiclase utilizando un Random Forest para clasificación multiclase. En la parte inferior: tareas de clasificación binaria NOR contra todos utilizando un Random Forest para clasificación binaria.



6. CONCLUSIONES Y PERSPECTIVAS

- ❖ La estrategia propuesta en este trabajo, que consiste en una arquitectura GAN progresiva condicionada para la caracterización es capaz de generar imágenes cardíacas MRI realistas con un índice de similitud estructural (MS-SSIM) similar al del conjunto de datos de entrenamiento.
- ❖ Esta estrategia permite capturar de forma apropiada la variedad en el conjunto de datos de entrenamiento como sugiere la métrica SWD. En contraste, las imágenes generadas desde estrategias no condicionadas colapsan hacia unos pocos modos.
- ❖ El condicionamiento durante el entrenamiento de la GAN ayudó a conservar características relevantes de la estructura global y local del corazón como se evidenció en la comparación visual de la estrategia GAN condicional y no condicional.
- ❖ Condicionar el entrenamiento de la GAN con una patología cardíaca es capaz de reproducir en cierta medida, características de distintas condiciones cardíacas en imágenes de resonancias magnéticas como sugieren los vectores del espacio latente embebido en las tareas de clasificación y visualización.
- ❖ Este trabajo abre la posibilidad de explorar nuevos marcadores relevantes asociados a enfermedades cardíacas particulares que pueden ayudar en la evaluación de la progresión de la enfermedad cardíaca y el monitoreo del tratamiento.

BIBLIOGRAFÍA

BERNARD, Olivier, et al. Deep learning techniques for automatic MRI cardiac multistructures segmentation and diagnosis: ¿Is the problem solved? IEEE transactions on medical imaging, 2018, vol. 37, no 11, p. 2514-2525.

CHARTSIAS, Agisilaos, et al. Adversarial image synthesis for unpaired multi-modal cardiac data. En International Workshop on Simulation and Synthesis in Medical Imaging. Springer, Cham, 2017. p. 3-13.

COHEN, Joseph Paul; LUCK, Margaux; HONARI, Sina. Distribution matching losses can hallucinate features in medical image translation. En International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2018. p. 529-536.

GOODFELLOW, Ian, et al. Generative adversarial nets. En Advances in neural information processing systems. 2014. p. 2672-2680.

GOODFELLOW, Ian. NIPS 2016 tutorial: Generative adversarial networks. arXiv preprint arXiv:1701.00160, 2016.

GULRAJANI, Ishaan, et al. Improved training of wasserstein gans. En Advances in neural information processing systems. 2017. p. 5767-5777.

KARRAS, Tero, et al. Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196, 2017.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. En Advances in neural information processing systems. 2012. p. 1097-1105.

MAHAPATRA, Dwarikanath; BOZORGTABAR, Behzad; GARNAVI, Rahil. Image superresolution using progressive generative adversarial networks for medical image analysis. Computerized Medical Imaging and Graphics, 2019, vol. 71, p. 30-39.

MCINNES, Leland; HEALY, John; MELVILLE, James. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426, 2018.

MIKOLAJCZYK, Agnieszka; GROCHOWSKI, Michał. Data augmentation for improving deep learning in image classification problem. En 2018 international interdisciplinary PhD workshop (IIPhDW). IEEE, 2018. p. 117-122.

MURUGESAN, Balamurali, et al. Recon-GLGAN: A Global-Local Context Based Generative Adversarial Network for MRI Reconstruction. En International Workshop on Machine Learning for Medical Image Reconstruction. Springer, Cham, 2019. p. 3-15.

ODENA, Augustus; OLAH, Christopher; SHLENS, Jonathon. Conditional image synthesis with auxiliary classifier gans. En Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 2017. p. 2642-2651.

PEREZ, Luis; WANG, Jason. The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621, 2017.

RADFORD, Alec; METZ, Luke; CHINTALA, Soumith. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434, 2015.

ROTH, Gregory A., et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, 2018, vol. 392, no 10159, p. 1736-1788.

SAVIOLI, Nicolás, et al. A generative adversarial model for right ventricle segmentation. arXiv preprint arXiv:1810.03969, 2018.

SCHLEGL, Thomas, et al. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. En *International Conference on Information Processing in Medical Imaging*. Springer, Cham, 2017. p. 146-157.

SIMARD, Patrice Y., et al. Best practices for convolutional neural networks applied to visual document analysis. En *Icdar*. 2003.

TAYLOR, Luke; NITSCHKE, Geoff. Improving deep learning using generic data augmentation. arXiv preprint arXiv:1708.06020, 2017.

VILLANI, Cédric. *Optimal transport: old and new*. Springer Science & Business Media, 2008.

WANG, Zhou, et al. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 2004, vol. 13, no 4, p. 600-612.
YI, Xin; WALIA, Ekta; BABYN, Paul. Generative adversarial network in medical imaging: A review. *Medical image analysis*, 2019, p. 101552.

ZHANG, Le; GOOYA, Ali; FRANGI, Alejandro F. Semi-supervised assessment of incomplete LV coverage in cardiac MRI using generative adversarial nets. *En International Workshop on Simulation and Synthesis in Medical Imaging*. Springer, Cham, 2017. p. 61-68.