Artículo Científico / Scientific Paper

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n34.2025.09
pISSN: 1390-650X / eISSN: 1390-860X

Exploración de modelos generativos profundos para una mejor generación de datos en la miocardiopatía hipertrófica

Exploring deep generative models for improved data generation in hypertrophic cardiomyopathy

Swarajya Madhuri Rayavarapu[1,*], Gottapu Sasibhushana Rao[1]

Recibido: 06-08-2024, Recibido tras revisión: 05-06-2025, Aceptado: 10-06-2025, Publicado: 01-07-2025

Resumen

Las estrategias de generación de datos son fundamentales para superar el desafío de los datos de entrenamiento limitados en el análisis de imágenes médicas basado en aprendizaje profundo, en particular para la miocardiopatía hipertrófica (HCM) mediante resonancia magnética (MRI). A diferencia de los métodos de aumento tradicionales, los modelos generativos profundos pueden sintetizar imágenes de MRI novedosas y diversas. Este estudio evalúa múltiples modelos generativos: autocodificadores variacionales (VAE), redes generativas adversarias (GAN), GAN convolucionales profundas (DCGAN), GAN con clasificador auxiliar (ACGAN), InfoGAN y modelos de difusión, utilizando el índice de similitud estructural (SSIM) y el coeficiente de correlación cruzada (CC) para evaluar la calidad de imagen y la fidelidad estructural. Si bien los VAE mostraron limitaciones como el ruido y la borrosidad, los modelos basados en GAN, especialmente DCGAN y ACGAN, produjeron imágenes de mayor calidad y precisión anatómica. Los modelos de difusión lograron la mayor fidelidad de imagen, aunque a expensas de tiempos de generación más prolongados. Estos resultados destacan la compensación entre la calidad de imagen y la eficiencia computacional, y demuestran el potencial de los modelos generativos para ampliar los conjuntos de datos de MRI, mejorando así las aplicaciones de aprendizaje profundo para el diagnóstico de HCM.

Palabras clave: generación de datos, modelos de difusión, redes generativas adversarias, codificadores automáticos variacionales

Abstract

Data generation strategies are essential for addressing the challenge of limited training data in deep learning-based medical image analysis, particularly for hypertrophic cardiomyopathy (HCM) using magnetic resonance imaging (MRI). Unlike traditional augmentation techniques, deep generative models can synthesize novel and diverse MRI images, enriching the training datasets. This study evaluates several generative models, including Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), Deep Convolutional GANs (DCGANs), Auxiliary Classifier GANs (ACGANs), InfoGANs, and Diffusion Models, using the Structural Similarity Index Measure (SSIM) and Cross-Correlation Coefficient (CC) to assess image quality and structural fidelity. While VAEs demonstrated limitations such as noticeable noise and blurriness, GAN-based models, particularly DCGANs and ACGANs, generated higher-quality and anatomically accurate images. Diffusion models achieved the highest image fidelity among all the methods evaluated, but required longer generation times. These findings underscore the trade-off between image quality and computational efficiency and highlight the potential of deep generative models to augment MRI datasets, thereby improving deep learning applications for HCM diagnosis.

Keywords: Data Generation, Diffusion models, Generative Adversarial networks, Variational autoencoders

[1,*]Departamento de Ingeniería Electrónica y de Comunicaciones, Universidad de Andhra, India.
Autor para correspondencia ✉: madhurirayavarapu.rs@andhrauniversity.edu.in.

Forma sugerida de citación: S. Madhuri Rayavarapu y G. Sasibhushana Rao, "Exploración de modelos generativos profundos para una mejor generación de datos en la miocardiopatía hipertrófica," *Ingenius, Revista de Ciencia y Tecnología*, N.° 34, pp. 116-125, 2025. DOI: https://doi.org/10.17163/ings.n34.2025.09.

1. Introducción

La inteligencia artificial (IA) se ha convertido en una herramienta cada vez más poderosa para analizar datos médicos, incluyendo imágenes, texto y señales, con el fin de apoyar la interpretación, identificación, clasificación y diagnóstico de enfermedades [1]. El desarrollo de software médico basado en IA depende en gran medida de grandes volúmenes de datos diversos, tales como electrocardiogramas, radiografías, imágenes por resonancia magnética (MRI), tomografía computarizada (CT), ecocardiografía e imágenes dermatológicas [2–8]. Sin embargo, la recopilación y anotación de estos conjuntos de datos sigue siendo costosa y requiere mucho tiempo, ya que a menudo se necesita que médicos expertos etiqueten manualmente los datos [9].

Aunque existen numerosos conjuntos de datos médicos, los datos de acceso público para afecciones cardíacas poco comunes, como la miocardiopatía hipertrófica (HCM), siguen siendo escasos. Esta limitación se debe a factores como la baja prevalencia de la enfermedad, las restricciones de privacidad y el desequilibrio de clases en los conjuntos de datos existentes. En consecuencia, existe una necesidad urgente de técnicas avanzadas capaces de generar imágenes médicas sintéticas de alta calidad para ampliar los conjuntos de datos actuales. Los modelos de aprendizaje profundo, en particular aquellos desarrollados para el análisis de imágenes médicas, se benefician en gran medida de conjuntos de datos amplios, equilibrados y etiquetados con precisión, lo cual mejora la precisión diagnóstica y la capacidad de generalización [10].

Además, la diversidad de datos de entrenamiento es esencial para mitigar el sobreajuste y aumentar la robustez del modelo.

Desarrollar modelos efectivos de aprendizaje profundo para la imagenología médica presenta varios desafíos. En primer lugar, la adquisición de datos suficientes es difícil debido a las restricciones de confidencialidad de los pacientes y a la limitada compartición de datos entre instituciones. En segundo lugar, la anotación manual de imágenes médicas es intensiva en mano de obra, requiere mucho tiempo y es susceptible a la variabilidad entre diferentes modalidades de imagen. En tercer lugar, los conjuntos de datos a menudo presentan un desequilibrio de clases, ya que los casos patológicos suelen constituir una minoría en comparación con los casos normales o saludables [11]. En conjunto, estos desafíos complican el desarrollo de sistemas de diagnóstico automatizado confiables.

Para abordar estos problemas, los investigadores han adoptado técnicas de aumento de datos para incrementar el tamaño y la variabilidad de los conjuntos de datos. Los métodos tradicionales de aumento incluyen transformaciones geométricas, operaciones de convolución y ajustes en el espacio de color [12–14]. Aunque útiles, estos métodos generan solo una variación limitada, ya que manipulan imágenes existentes en lugar de crear nuevas imágenes desde cero. En contraste, los enfoques de aumento basados en aprendizaje profundo, como los autocodificadores variacionales (VAE), la transferencia de estilo neural, los modelos de difusión y las redes generativas adversarias (GAN), han demostrado una capacidad superior para generar imágenes sintéticas diversas y realistas [15–17]. A diferencia de los métodos convencionales, estos modelos generativos aprenden la distribución subyacente de los datos en espacios de alta dimensión, lo que permite la creación de muestras novedosas que se asemejan estrechamente a datos del mundo real. Como resultado, los modelos generativos profundos son particularmente eficaces para mitigar la escasez de datos y el desequilibrio de clases, especialmente en enfermedades raras como la HCM, mejorando así el entrenamiento y el desempeño de generalización de los modelos de diagnóstico.

La sección 2 revisa los modelos generativos profundos utilizados para la generación de imágenes sintéticas y describe la metodología, la sección 3 presenta los resultados y discusión y la sección 4 resume las conclusiones.

2. Materiales y métodos

2.1. Modelos generativos profundos

Esta sección proporciona una descripción concisa de los diferentes modelos generativos profundos empleados para la generación de imágenes médicas. Se centra en tres tipos principales de modelos: autocodificadores variacionales, redes generativas adversarias y modelos de difusión.

2.1.1. Autocodificador variacional

Welling y Kingma [18] introdujeron el concepto de autocodificador variacional (VAE) en 2013. Un VAE describe una observación en el espacio latente de manera probabilística. En lugar de producir un único valor para cada atributo del estado latente, el codificador genera una distribución de probabilidad para cada atributo. Las aplicaciones de los VAE incluyen la compresión de datos y la generación de datos sintéticos. La Figura 1 ilustra la arquitectura del VAE, con sus componentes principales detallados a continuación.

- **Entrada.** La entrada a un VAE depende de la aplicación específica y del dominio de interés. Para los VAE basados en imágenes, la entrada generalmente consiste en imágenes completas o fragmentos de imágenes.
- **Codificador.** El codificador transforma los datos de entrada en los parámetros del espacio latente, los cuales definen la distribución de

probabilidad asociada. Típicamente, consiste en múltiples capas de redes neuronales, como capas convolucionales o completamente conectadas, que trabajan en conjunto para reducir la dimensionalidad de la entrada. La salida del codificador es un conjunto de vectores de media y varianza que aproximan las características de una distribución gaussiana multivariada en el espacio latente.

- **Espacio latente.** Cada punto en el espacio latente representa un código latente, que es una descripción de dimensión reducida de los datos de entrada. Durante el entrenamiento, el codificador aprende a generar códigos latentes que capturan con precisión las características esenciales de la entrada.
- **Decodificador.** Antes de mapear los datos transformados de regreso al espacio de entrada, el decodificador extrae un subconjunto relevante del espacio latente. Al igual que el codificador, consiste en varias capas de redes neuronales que muestrean progresivamente el código latente y generan la salida. El objetivo del decodificador es reproducir los datos de entrada originales con la mayor precisión posible.

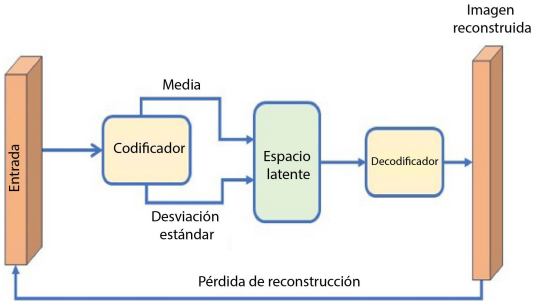

Figura 1. Autocodificador variacional

Los VAE se basan en la representación matemática del espacio latente aprendida por la red del codificador para aproximar la distribución de los datos, mientras que la red del decodificador utiliza esta representación para generar muestras similares a los datos de entrenamiento.

La red del codificador mapea una muestra de entrada x a una representación latente z, de acuerdo con la ecuación (1), mientras que la red del decodificador utiliza la ecuación (2) para reconstruir el espacio de entrada a partir de la representación latente z. Las funciones f y g denotan las redes del codificador y del decodificador, respectivamente.

$$z = f(x) \tag{1}$$

$$x' = g(z) \tag{2}$$

La función de pérdida del VAE se define en la ecuación (3). En esta ecuación, el primer término representa la pérdida de reconstrucción, mientras que el segundo término corresponde a la divergencia de Kullback–Leibler (KL).

$$Pérdida = E\left[L\left(x, x'\right)\right] - D_{KL}\left(q\left(z \mid x\right) || \, p\left(z\right)\right) \tag{3}$$

$$Donde \; L\left(x, x'\right) = logp\left(x \mid x'\right) \tag{4}$$

2.1.2. Redes generativas adversarias

Las redes generativas adversarias (GAN) son un método de aprendizaje no supervisado que aprovecha el marco bien establecido de la teoría de juegos de suma cero para dos jugadores. Este concepto fue introducido por Goodfellow et al. en 2014 [19]. En una GAN, el generador crea nuevas muestras basadas en datos reales, mientras que el discriminador estima la distribución subyacente de los datos distinguiendo entre muestras reales y generadas, ver la figura 2.

- **Generador.** El componente generador de una GAN crea datos sintéticos transformando ruido aleatorio en muestras que se asemejan a datos reales.
- **Discriminador.** El componente discriminador de una GAN actúa como un clasificador que distingue entre datos reales y datos generados artificialmente por el generador.

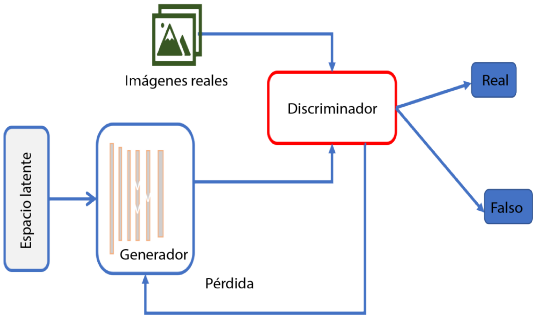

Figura 2. Arquitectura de una red generativa adversaria (GAN)

Las redes generativas adversarias (GAN) operan sobre la base de un marco matemático en el cual la red del discriminador proporciona retroalimentación sobre el realismo de las muestras generadas, mientras que la red del generador mapea representaciones del espacio latente al espacio de datos original. Este proceso adversarial permite que las GAN aprendan un modelo generativo de los datos y produzcan muestras sintéticas diversas y realistas. En comparación con otros métodos generativos, las GAN ofrecen ventajas

notables, incluyendo la capacidad de manejar distribuciones de datos complejas y generar imágenes de alta resolución [20].

La función de pérdida de la GAN se define en la ecuación (5).

$$min_G max_D V(D,G) = E_x log(D(x)) + E_z log(1 - D(G(z))) \quad (5)$$

En la ecuación (5), Z representa el espacio latente o entrada de ruido, que se suministra al generador. El discriminador se denota como D, y el generador se denota como G. El discriminador recibe tanto las imágenes generadas como las muestras de datos reales. La representación de los datos reales se expresa como $D(x)$, mientras que la representación de los datos generados se expresa como $D(G(z))$. Las redes del discriminador y del generador se entrenan simultáneamente: el discriminador busca minimizar la puntuación de clasificación de las muestras generadas, mientras que el generador intenta maximizar dicha puntuación.

En los últimos años, se han introducido varias variantes de las GAN. Algunas de estas variantes se describen en las secciones siguientes.

2.1.3. Redes generativas adversarias convolucionales profundas (DCGAN)

Las redes generativas adversarias convolucionales profundas (DCGAN) son una variante novedosa de las redes neuronales convolucionales (CNN), diseñada con restricciones arquitectónicas específicas, tal como se introdujo en [21]. Para cumplir con estos requisitos, las DCGAN incorporan tres modificaciones arquitectónicas clave.

En primer lugar, reemplazan las capas ocultas completamente conectadas y las capas de agrupamiento por capas convolucionales, utilizando convoluciones fraccionarias con paso *(fractional-strided convolutions)* en el generador y convoluciones con paso *(strided convolutions)* en el discriminador para mejorar el rendimiento de la red.

En segundo lugar, aplican funciones de activación ReLU a todas las capas del generador, excepto a la capa de salida, mientras que en el discriminador se emplean funciones de activación LeakyReLU en todas las capas.

En tercer lugar, se aplica normalización por lotes *(batch normalization)* de forma consistente tanto en el generador como en el discriminador.

2.1.4. Redes generativas adversarias con clasificador auxiliar (AC-GAN)

Odena et al. [22] propusieron la red generativa adversaria con clasificador auxiliar (AC-GAN), que incluye un clasificador adicional. El discriminador de AC-GAN incorpora un clasificador para categorizar las muestras en clases discretas. La incorporación de este clasificador mejora la estabilidad del entrenamiento.

En una AC-GAN, el generador G utiliza tanto el ruido z como una etiqueta de categoría c, muestreada de una distribución, para generar una muestra sintética, denotada como $x_{fake} = G(c, z)$. El discriminador D distingue entre muestras reales y falsas, considerando tanto la autenticidad como las etiquetas de clase.

Las funciones objetivo de la AC-GAN se definen en las ecuaciones (6) y (7).

$$L_s = E\left[log(P(S = real \mid Xreal))\right] + E\left[log(P(S = fake \mid Xfake))\right] \quad (6)$$

$$L_c = E\left[log(P(C = c \mid Xreal))\right] + E\left[log(P(C = c \mid Xfake))\right] \quad (7)$$

Los términos L_s y L_c denotan las probabilidades de identificar correctamente la fuente y la clase, respectivamente. X representa la imagen de entrada, C es la etiqueta de clase y S es la fuente.

Al entrenar el discriminador D, el objetivo principal es maximizar la pérdida total $(L_s + L_c)$. En contraste, el generador G se entrena para maximizar la diferencia $(L_s - L_c)$ entre las pérdidas.

2.1.5. GAN de maximización de la información (InfoGAN)

InfoGAN fue propuesto por Chen et al. [23] como un enfoque basado en la teoría de la información para mejorar la interpretabilidad en las redes generativas adversarias (GAN) mediante el aprendizaje de variables latentes significativas. El término "info" hace referencia a la información mutua compartida entre la distribución generada $G(z, c)$ y el código latente c.

Para aumentar la correlación entre x y c, InfoGAN maximiza la información mutua $I(c; G(z, c))$. Se añade un término de regularización que incorpora este objetivo de información a la función de pérdida estándar de las GAN.

Para estimar una cota inferior manejable de la información mutua $P(c \mid x)$, se introduce una distribución auxiliar $Q(c \mid x)$ que aproxima $P(c \mid x)$.

La función objetivo de InfoGAN se define en la ecuación (8).

$$\frac{min}{G,Q}\frac{max}{D} L(D,G,Q) = V(D,G) - \lambda I(G,Q) \quad (8)$$

Donde λ es la constante de regularización, típicamente establecida en uno.

2.1.6. Modelos de difusión

Sohl-Dickstein et al. fueron los primeros en introducir los modelos de difusión [24]. Basándose en esta idea, Ho et al. [25] propusieron los modelos probabilísticos de difusión con eliminación de ruido (DDPM, por sus siglas en inglés), marcando la primera demostración de que los modelos de difusión pueden alcanzar un rendimiento comparable al de otros modelos generativos en tareas de síntesis de imágenes.

Los modelos de difusión son algoritmos avanzados de aprendizaje automático que generan datos de alta calidad mediante la adición gradual de ruido a un conjunto de datos y el posterior aprendizaje de cómo invertir este proceso [26]. Como una subcategoría de los modelos generativos basados en aprendizaje profundo, su objetivo principal es producir datos sintéticos que sean realistas y plausibles dadas unas condiciones de entrada.

Debido a sus numerosas ventajas sobre otros métodos generativos —como la capacidad de generar datos altamente diversos, manejar conjuntos de datos de alta dimensión y aprender distribuciones complejas de manera eficaz—, los modelos generativos basados en difusión han ganado recientemente una popularidad significativa en diversas disciplinas científicas [27].

Un modelo de difusión es un marco generativo probabilístico que implica dos procesos de múltiples pasos: difusión directa y difusión inversa. En el proceso de difusión directa, se añade ruido gradualmente a los datos de entrada hasta que la información original queda completamente oculta. En contraste, el proceso de difusión inversa emplea una red neuronal entrenable para eliminar progresivamente el ruido y reconstruir la distribución original de los datos. Las salidas sintéticas se generan aplicando esta red neuronal entrenada a muestras ruidosas, ver la figura 3.

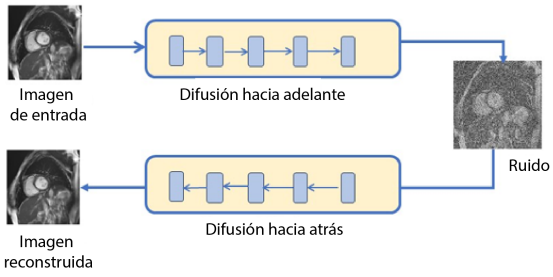

Figura 3. Arquitectura de un modelo de difusión

Para iniciar el proceso de difusión directa, se extraen muestras de una distribución simple, típicamente una distribución gaussiana. Esta muestra inicial se somete luego a una secuencia de pequeñas transformaciones reversibles. A través de una cadena de Markov, cada paso incrementa gradualmente la complejidad de la muestra, lo cual puede interpretarse como la adición progresiva de ruido estructurado.

En un proceso de difusión directa, pequeñas cantidades de ruido gaussiano se añaden progresivamente a un punto de datos x, extraído de la distribución real de datos $q(x)$, produciendo una serie de muestras cada vez más ruidosas, denotadas por $x_1, x_2, x_3, \ldots, x_T$. Este proceso se define matemáticamente mediante las ecuaciones (9) y (10).

$$q(x_t \mid x_{t-1}) = N(x_i, \sqrt{1-\beta_t}x_{t-1,}\beta_t I) \tag{9}$$

$$q(x_{1:T} \mid x_0) = \prod_{t}^{T} = q(x_t \mid x_{t-1}) \tag{10}$$

El proceso de difusión inversa se define mediante las ecuaciones (11) y (12).

$$p\theta(x_{0:T}) = p(x_T) \cdot \prod_{l=1}^{\bar{I}} p\theta(x_{t-1} \mid x_t) \tag{11}$$

$$p\theta(x_{t-1} \mid x_t) = W(x_{t-1}; u_\theta(x_t, t), \sum \theta(x_t, t)) \tag{12}$$

2.2. Metodología

Las imágenes de miocardiopatía hipertrófica (HCM) se generan utilizando diversos modelos generativos profundos, siendo la imagen por resonancia magnética (IRM) la modalidad principal para la detección y evaluación de la enfermedad. La IRM proporciona imágenes detalladas y de alta resolución de la estructura y función del corazón, sin el uso de radiación ionizante, lo que la hace particularmente adecuada para la evaluación cardíaca.

Características clave de la HCM, como el engrosamiento de la pared ventricular y la fibrosis miocárdica, pueden evaluarse de manera efectiva mediante escaneos de IRM. Esta técnica de imagen también permite la evaluación de la gravedad y progresión de la enfermedad, lo cual respalda la toma de decisiones clínicas.

Además, los datos de IRM sirven como base para el desarrollo y validación de técnicas automatizadas para la detección y análisis de la HCM.

El conjunto de datos para la miocardiopatía hipertrófica se obtuvo del conjunto de datos cardíacos ACDC (Cardiac ACDC Dataset).

2.2.1. Conjunto de datos

Las imágenes por resonancia magnética cardíaca (IRM) utilizadas en este estudio provienen del conjunto de datos de código abierto Cardiac ACDC [28]. Los exámenes clínicos reales que conforman el conjunto de datos ACDC fueron proporcionados por el Hospital Universitario de Dijon.

Para proteger la privacidad de los pacientes, todos los datos fueron sometidos a un proceso exhaustivo

de anonimización y procesamiento, conforme a los criterios establecidos por el comité de ética local del hospital francés de Dijon.

El conjunto de datos incluye un número suficiente de ejemplos para entrenar algoritmos de aprendizaje automático y evaluar de manera confiable los cambios en parámetros fisiológicos clave derivados de cine-IRM, como el volumen diastólico y la fracción de eyección. Abarca una variedad de patologías cardíacas y se divide en cinco categorías con una proporción equitativa de casos en cada una.

En total, el conjunto de datos comprende 150 exámenes, cada uno obtenido de un paciente diferente.

2.2.2. Configuración experimental y proceso de entrenamiento

El lenguaje de programación Python y el entorno de trabajo PyTorch se utilizaron para implementar los diversos modelos generativos profundos. El entrenamiento de los modelos se llevó a cabo en un entorno de Anaconda Navigator y Jupyter Notebook, con aceleración por unidad de procesamiento gráfico (GPU) habilitada en un ordenador portátil con procesador Intel Core i7.

Cada modelo generativo, incluyendo VAE, GAN, GAN convolucional profunda (DCGAN), InfoGAN, ACGAN y modelos de difusión, fue entrenado utilizando el optimizador Adam con una tasa de aprendizaje inicial de $0,0002, \beta_1 = 0,5$ y $\beta_2 = 0,999$. El tamaño del lote se estableció en 32, y los modelos se entrenaron durante un máximo de 100 épocas, a menos que se activara una detención anticipada basada en la pérdida de validación. Para los modelos de difusión, el entrenamiento se extendió aproximadamente hasta las 115 épocas.

Para las arquitecturas basadas en GAN, se aplicaron las funciones de pérdida adversarial correspondientes; InfoGAN además optimizó un término de pérdida de información mutua, y ACGAN incorporó una pérdida de clasificación auxiliar. Los VAE se entrenaron minimizando una combinación de la pérdida de reconstrucción y la divergencia de Kullback–Leibler (KL).

La estabilidad del entrenamiento se mejoró mediante técnicas como penalización de gradiente y suavizado de etiquetas, cuando fue aplicable. Los puntos de control de los modelos se guardaron regularmente para preservar los pesos de mejor desempeño según las métricas de validación.

Al finalizar el entrenamiento, se generaron aproximadamente 1000 imágenes sintéticas. Estas imágenes fueron posteriormente sometidas a una evaluación cualitativa para analizar su similitud con respecto al conjunto de datos original.

2.2.3. Métricas de evaluación

Se utilizaron las métricas SSIM, correlación cruzada y error cuadrático medio (MSE) para evaluar los modelos generativos profundos, ya que capturan diferentes aspectos de la calidad y similitud de las imágenes, los cuales son cruciales en contextos de imagenología médica.

2.2.4. Índice de similitud estructural (SSIM)

El índice de similitud estructural, también conocido como SSIM *(Structural Similarity Index)*, es una métrica ampliamente utilizada para cuantificar la similitud entre dos imágenes [29]. El SSIM evalúa la similitud de la información estructural considerando la luminancia, el contraste y los patrones estructurales dentro de las imágenes. Funciona realizando tres comparaciones entre fragmentos correspondientes: comparación de luminancia, comparación de contraste y comparación estructural. Estos resultados se combinan para producir un índice SSIM general, el cual varía entre –1 y 1, donde un valor de 1 indica una similitud estructural perfecta.

Para dos imágenes, x e y, el SSIM se calcula utilizando la ecuación (13):

$$(x, y) = \frac{(2 \cdot \mu_x \mu_y + c_1) \cdot (2 \cdot \sigma_{xy} + c_2)}{(c_1 2 + \mu_x 2 + \mu_y 2) \cdot (c_2 2 + \sigma_x 2 + \sigma_y 2)} \quad (13)$$

$\sigma_y 2$, $\sigma_x 2$ son las varianzas de las imágenes y y x, respectivamente.

$\sigma_{x,y}$ es la covarianza entre y y x.

μ_x, μ_y son los valores medios de las imágenes x e y, respectivamente.

Las constantes c_1 y c_2 se determinan en función del rango dinámico de los valores de los píxeles. El valor de SSIM es igual a uno si, y solo si, x e y son idénticas.

2.2.5. Coeficiente de correlación cruzada

En el procesamiento de imágenes, la correlación cruzada es una técnica utilizada para medir la similitud entre dos señales o imágenes [30]. Consiste en deslizar una imagen (o señal) sobre otra y calcular una medida de similitud en cada posición.

2.2.6. Error cuadrático medio (MSE)

Una de las métricas de referencia completa más utilizadas y sencillas es el error cuadrático medio (MSE), que se calcula elevando al cuadrado las diferencias de intensidad entre los píxeles correspondientes en las imágenes distorsionadas y de referencia [31].

3. Resultados y discusión

La comparación entre las imágenes reales y las imágenes generadas por diferentes modelos generativos profundos se llevó a cabo utilizando tres métricas: el índice de similitud estructural (SSIM), el coeficiente de correlación cruzada y el error cuadrático medio (MSE). Estas métricas cuantifican el grado de similitud entre las imágenes generadas y las originales.

El rendimiento de las imágenes generadas para la miocardiopatía hipertrófica (HCM) utilizando diferentes modelos generativos profundos se resume en la Tabla 1. El índice de similitud de las imágenes generadas por el VAE es inferior al de todos los modelos basados en GAN, así como al de los modelos de difusión.

Tabla 1. Análisis de rendimiento entre imágenes reales y generadas utilizando diferentes modelos generativos profundos (VAE, GAN, DCGAN, ACGAN y modelos de difusión)

Método	SSIM	Coeficiente de correlación cruzada	MSE
VAE	0.903	0.8421	0.39
GAN	0.941	0.8821	0.1
DCGAN	0.958	0.9204	0.1
ACGAN	0.969	0.9371	0.09
Modelos de difusión	0.972	0.9403	0.06

Las imágenes generadas por los distintos modelos generativos profundos para la HCM se presentan en la Figura 4. La primera imagen muestra la IRM original, seguida por la imagen generada por el VAE, la imagen generada por el GAN, las imágenes generadas por DCGAN e InfoGAN, y finalmente la imagen generada mediante el modelo de difusión.

Además de estos resultados de similitud, también se presentan los gráficos de pérdida durante el entrenamiento para los distintos modelos generativos profundos. Los resultados de la función de pérdida para el VAE se muestran en la Figura 5, donde se puede observar que la pérdida disminuye a medida que aumenta el número de épocas.

La Figura 6 muestra que la pérdida del generador en la GAN aumenta al inicio, mientras que la pérdida del discriminador disminuye a medida que aumenta el número de épocas, hasta aproximadamente la época 18. A partir de ese punto, ambas funciones de pérdida convergen alrededor de la época 62, lo que produce los mejores resultados. Las curvas de pérdida durante el entrenamiento para los modelos DCGAN y de difusión se muestran en las Figuras 7 y 8, respectivamente.

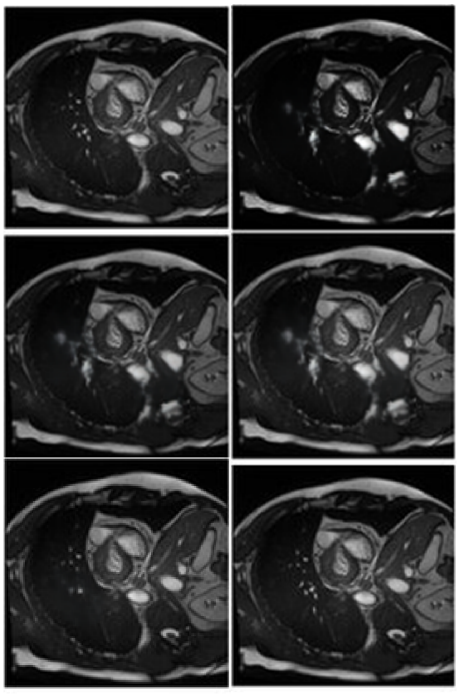

Figura 4. Imagen original e imágenes cardíacas sintetizadas generadas utilizando VAE, GAN, DCGAN, ACGAN y modelo de difusión, respectivamente.

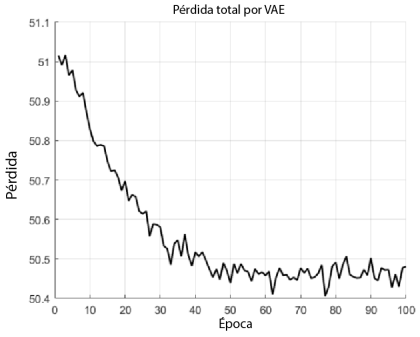

Figura 5. Pérdida de entrenamiento del VAE

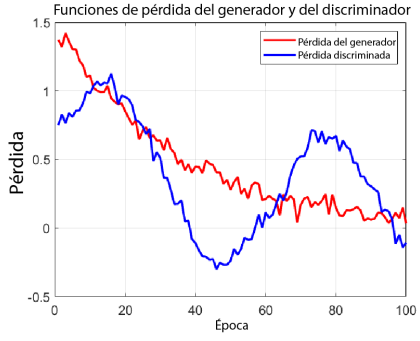

Figura 6. Pérdida de entrenamiento de la GAN

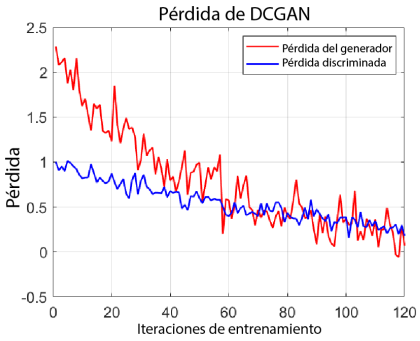

Figura 7. Pérdida de entrenamiento del DCGAN

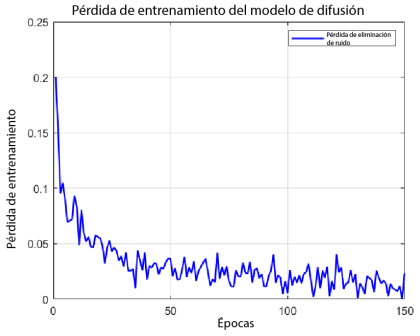

Figura 8. Pérdida de entrenamiento del modelo de difusión

En comparación con las GAN, los VAE presentan un mejor desempeño durante el entrenamiento, debido a su resistencia al colapso de modo y a su capacidad para producir salidas más diversas. Sin embargo, una desventaja importante es que las imágenes generadas suelen ser borrosas y carecen de detalles nítidos.

Aunque los modelos de difusión pueden generar resultados altamente realistas y mantener un entrenamiento estable, el extenso proceso de difusión requiere un período de muestreo prolongado, lo cual puede limitar su viabilidad para la generación eficiente de imágenes. Esta compensación se refleja en las métricas de rendimiento presentadas en la Tabla 1.

4. Conclusiones

Las estrategias de generación de datos son fundamentales para superar el desafío de la disponibilidad reducida de datos de entrenamiento en el análisis de imágenes médicas basado en aprendizaje profundo. A diferencia de las técnicas convencionales de aumento de datos, comúnmente aplicadas en el diagnóstico cardíaco, los modelos generativos profundos pueden sintetizar muestras completamente nuevas y diversas.

En este estudio, se evaluó el rendimiento de varios enfoques de generación de datos utilizando el índice de similitud estructural (SSIM) y el coeficiente de correlación cruzada (CC), ambos considerados métricas estándar para valorar la calidad de imagen y la fidelidad estructural

El enfoque del autoencoder variacional (VAE) alcanzó un valor de SSIM de 0,9028 y un coeficiente de correlación cruzada (CC) de 0,8421; sin embargo, las imágenes generadas presentaron ruido significativo y desenfoque, lo que evidencia limitaciones en cuanto al realismo visual.

En contraste, las redes generativas antagónicas (GAN) demostraron un mejor rendimiento, alcanzando un SSIM de 0,9428 y el mismo valor de CC (0,8421). Entre las variantes de GAN, tanto las GAN convolucionales profundas (DCGAN) como las GAN con clasificador auxiliar (ACGAN) produjeron resultados superiores, con valores de SSIM de 0,9576 y 0,9687, respectivamente, lo que indica una mayor capacidad para generar imágenes de alta calidad y precisión estructural.

Los modelos de difusión superaron tanto a las GAN como a los VAE en términos de métricas de similitud, al alcanzar los valores más altos de SSIM; sin embargo, su aplicabilidad práctica se ve limitada por tiempos de muestreo significativamente más prolongados. Esta compensación entre la calidad de imagen y la eficiencia computacional debe considerarse cuidadosamente al seleccionar un modelo generativo adecuado para la ampliación de datos médicos en el diagnóstico cardíaco, en particular en afecciones como la miocardiopatía hipertrófica.

Rol de autores

- **Swarajya Madhuri Rayavarapu:** conceptualización, metodología, software.
- **Gottapu Sasibhushana Rao:** investigación, supervisión.

Referencias

[1] C. González García, E. Núñez-Valdez, V. García-Díaz, C. Pelayo G-Bustelo, and J. M. Cueva-Lovelle, "A review of artificial intelligence in the internet of things," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 4, p. 9, 2019. [Online]. Available: http://dx.doi.org/10.9781/ijimai.2018.03.004

[2] Y. Shen, L. Chen, J. Liu, H. Chen, C. Wang, H. Ding, and Q. Zhang, "PADS-Net: GAN-based radiomics using multi-task network of denoising and segmentation for ultrasonic diagnosis of Parkinson disease," *Computerized Medical Imaging and Graphics*, vol. 120, p. 102490, Mar. 2025. [Online]. Available: https://doi.org/10.1016/j.compmedimag.2024.102490

[3] H. Zhang and Y. Qie, "Applying deep learning to medical imaging: A review," *Applied Sciences*, vol. 13, no. 18, p. 10521, Sep. 2023. [Online]. Available: https://doi.org10.3390/app131810521

[4] M. Rana and M. Bhushan, "Machine learning and deep learning approach for medical image analysis: diagnosis to detection," *Multimedia Tools and Applications*, vol. 82, no. 17, pp. 26 731–26 769, Dec. 2022. [Online]. Available: https://doi.org/10.1007/s11042-022-14305-w

[5] X. Liu, H. Wang, Z. Li, and L. Qin, "Deep learning in ECG diagnosis: A review," *Knowledge-Based Systems*, vol. 227, p. 107187, Sep. 2021. [Online]. Available: https://doi.org/10.1016/j.knosys.2021.107187

[6] S. K. Mathivanan, S. Srinivasan, M. S. Koti, V. S. Kushwah, R. B. Joseph, and M. A. Shah, "A secure hybrid deep learning framework for brain tumor detection and classification," *Journal of Big Data*, vol. 12, no. 1, Mar. 2025. [Online]. Available: https://doi.org/10.1186/s40537-025-01117-6

[7] C. Chola, P. Mallikarjuna, A. Y. Muaad, J. V. Bibal Benifa, J. Hanumanthappa, and M. A. Al-antari, "A hybrid deep learning approach for COVID-19 diagnosis via CT and X-ray medical images," in *The 1st International Electronic Conference on Algorithms*, ser. IOCA 2021. MDPI, Sep. 2021, p. 13. [Online]. Available: https://doi.org/10.3390/IOCA2021-10909

[8] F. Y. Shih and H. Patel, "Deep learning classification on optical coherence tomography retina images," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 34, no. 08, p. 2052002, Oct. 2019. [Online]. Available: https://doi.org/10.1142/S0218001420520023

[9] P. Gupta, S. Nandakumar, M. Gupta, and G. Panda, "Data programming enabled weak supervised labeling for ECG time series," *Biomedical Signal Processing and Control*, vol. 87, p. 105540, Jan. 2024. [Online]. Available: https://doi.org10.1016/j.bspc.2023.105540

[10] S. U. Amin, A. Hussain, B. Kim, and S. Seo, "Deep learning based active learning technique for data annotation and improve the overall performance of classification models," *Expert Systems with Applications*, vol. 228, p. 120391, Oct. 2023. [Online]. Available: https://doi.org/10.1016/j.eswa.2023.120391

[11] T. Liu, W. Fan, and C. Wu, "A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset," *Artificial Intelligence in Medicine*, vol. 101, p. 101723, Nov. 2019. [Online]. Available: https://doi.org/10.1016/j.artmed.2019.101723

[12] T. Islam, M. S. Hafiz, J. R. Jim, M. M. Kabir, and M. Mridha, "A systematic review of deep learning data augmentation in medical imaging: Recent advances and future research directions," *Healthcare Analytics*, vol. 5, p. 100340, Jun. 2024. [Online]. Available: https://doi.org/10.1016/j.health.2024.100340

[13] N. Nonaka and J. Seita, "Data augmentation for electrocardiogram classification with deep neural network," *arXiv*, 2020. [Online]. Available: https://doi.org/10.48550/arXiv.2009.04398

[14] M. M. Rahman, M. W. Rivolta, F. Badilini, and R. Sassi, "A systematic survey of data augmentation of ECG signals for AI applications," *Sensors*, vol. 23, no. 11, p. 5237, May 2023. [Online]. Available: http://doi.org/10.3390/s23115237

[15] F. J. Moreno-Barea, J. M. Jerez, and L. Franco, "Improving classification accuracy using data augmentation on small data sets," *Expert Systems with Applications*, vol. 161, p. 113696, Dec. 2020. [Online]. Available: https://doi.org/10.1016/j.eswa.2020.113696

[16] J. Saldanha, S. Chakraborty, S. Patil, K. Kotecha, S. Kumar, and A. Nayyar, "Data augmentation using variational autoencoders for improvement of respiratory disease classification," *PLOS ONE*, vol. 17, no. 8, p. e0266467, Aug. 2022. [Online]. Available: https://doi.org/10.1371/journal.pone.0266467

[17] D. Bhattacharya, S. Banerjee, S. Bhattacharya, B. Uma Shankar, and S. Mitra, *GAN-Based Novel Approach for Data Augmentation with Improved Disease Classification.* Springer Singapore, Dec. 2019, pp. 229–239. [Online]. Available: https://doi.org/10.1007/978-981-15-1100-4_11

[18] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv*, 2013. [Online]. Available: https://doi.org/10.48550/arXiv.1312.6114

[19] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *arXiv*, 2014. [Online]. Available: https://doi.org/10.48550/arXiv.1406.2661

[20] Y. Skandarani, P.-M. Jodoin, and A. Lalande, "GANs for medical image synthesis: An empirical study," *Journal of Imaging*, vol. 9, no. 3, p. 69, Mar. 2023. [Online]. Available: https://doi.org10.3390/jimaging9030069

[21] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv*, 2015. [Online]. Available: https://doi.org/10.48550/arXiv.1511.06434

[22] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," *arXiv*, 2016. [Online]. Available: https://doi.org/10.48550/arXiv.1610.09585

[23] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "InfoGAN: interpretable representation learning by information maximizing generative adversarial nets," *arXiv*, 2016. [Online]. Available: https://doi.org/10.48550/arXiv.1606.03657

[24] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," *arXiv*, 2015. [Online]. Available: https://doi.org/10.48550/arXiv.1503.03585

[25] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *arXiv*, 2020. [Online]. Available: https://doi.org/10.48550/arXiv.2006.11239

[26] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, "Diffusion models in vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 9, pp. 10 850–10 869, Sep. 2023. [Online]. Available: https://doi.org/10.1109/TPAMI.2023.3261988

[27] Z. Guo, J. Liu, Y. Wang, M. Chen, D. Wang, D. Xu, and J. Cheng, "Diffusion models in bioinformatics: A new wave of deep learning revolution in action," *arXiv*, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2302.10907

[28] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. Gonzalez Ballester, G. Sanroma, S. Napel, S. Petersen, G. Tziritas, E. Grinias, M. Khened, V. A. Kollerathu, G. Krishnamurthi, M.-M. Rohé, X. Pennec, M. Sermesant, F. Isensee, P. Jäger, K. H. Maier-Hein, P. M. Full, I. Wolf, S. Engelhardt, C. F. Baumgartner, L. M. Koch, J. M. Wolterink, I. Išgum, Y. Jang, Y. Hong, J. Patravali, S. Jain, O. Humbert, and P.-M. Jodoin, "Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved?" *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2514–2525, Nov. 2018. [Online]. Available: http://doi.org/10.1109/TMI.2018.2837502

[29] H. Sheikh, M. Sabir, and A. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006. [Online]. Available: http://doi.org/10.1109/TIP.2006.881959

[30] G. Prieto, E. Guibelalde, M. Chevalier, and A. Turrero, "Use of the cross-correlation component of the multiscale structural similarity metric (R* metric) for the evaluation of medical images," *Medical Physics*, vol. 38, no. 8, pp. 4512–4517, Jul. 2011. [Online]. Available: https://doi.org/10.1118/1.3605634

[31] A. Borji, "Pros and cons of GAN evaluation measures: New developments," *arXiv*, 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2103.09396