Aumento de datos en Retinopatía Diabética implementando DCGAN y cWGAN-GP

Karina Cardozo
Redes Neuronales Generativas Profundas: Fundamentos y resolución de problemas (2025)
Universidad de la República - Facultad de Ingeniería

***Abstract*—La retinopatía diabética (DR) es una complicación de la diabetes que daña los vasos sanguíneos de la retina y puede provocar pérdida irreversible de visión cuando no se detecta de forma temprana. La necesidad de sistemas automáticos que detecten DR y clasifiquen sus estadios se ve limitada por el desbalance entre niveles de severidad y por la variabilidad de adquisición en las retinografías, lo que favorece errores de clasificación, en particular en los casos avanzados. Para mitigar esta limitación, en este trabajo se explora la generación sintética de retinografías como estrategia de aumento de datos, con énfasis en patrones asociados a enfermedad severa (severidad 4) que caracterizan la progresión desde formas no proliferativas hacia retinopatía proliferativa.**

Se realiza un estudio comparativo entre dos configuraciones generativas: una DCGAN no condicional a 256×256 píxeles entrenada exclusivamente sobre retinografías de severidad 4 del desafío APTOS 2019, y una arquitectura condicional tipo Wasserstein con penalización de gradiente (cWGAN-GP) que sintetiza imágenes RGB de 120×120 píxeles condicionadas por el nivel de severidad (0–4) a partir del conjunto Diabetic Retinopathy Detection, lo que define cinco condiciones discretas.

En ambos modelos se describe el preprocesamiento, la arquitectura y la configuración de entrenamiento, y se monitoriza la calidad mediante métricas de Fréchet Inception Distance (FID) e Inception Score (IS), además de inspección visual. Los resultados muestran que la DCGAN reproduce de forma global el campo retinal pero genera texturas ruidosas, con un FID muy elevado y un IS bajo. La cWGAN-GP condicional mejora la estabilidad del entrenamiento y la coherencia global de las imágenes, y alcanza un FID más bajo e IS mayores, aunque todavía se observa falta de detalle fino clínicamente relevante, en particular en la vasculatura y en las lesiones pequeñas.

***Index Terms*—GAN, DCGAN, WGAN-GP, cWGAN, diabetic retinopathy, fundus, data augmentation, conditional generation, FID, Inception Score.**

I. Introducción

La retinopatía diabética es una complicación microvascular de la diabetes y una de las principales causas de pérdida visual prevenible a nivel mundial. La enfermedad aparece cuando los vasos sanguíneos de la retina se dañan por niveles elevados y mantenidos de glucosa, lo que puede producir edema, hemorragias y, en estadios avanzados, pérdida irreversible de visión [9].

La evolución de la retinopatía diabética se describe mediante una escala de severidad que combina el tipo y la cantidad de lesiones visibles en retinografías de fondo de ojo. Esta escala, basada en los criterios del ETDRS y difundida por la American Academy of Ophthalmology [10], distingue cinco estadios clínicos. En etapas tempranas predominan los microaneurismas, las hemorragias intrarretinianas y los exudados, mientras que en estadios avanzados aparecen áreas extensas de isquemia y neovascularización con mayor riesgo de hemorragia vítrea y desprendimiento traccional de retina. En la Fig. 1 se ilustran ejemplos de retinografías correspondientes a los distintos estadios de la escala internacional: (a) No DR, (b) leve, (c) moderada, (d) severa y (e) proliferativa, que en el dataset utilizado se codifican como etiquetas enteras de 0 a 4 en el mismo orden.

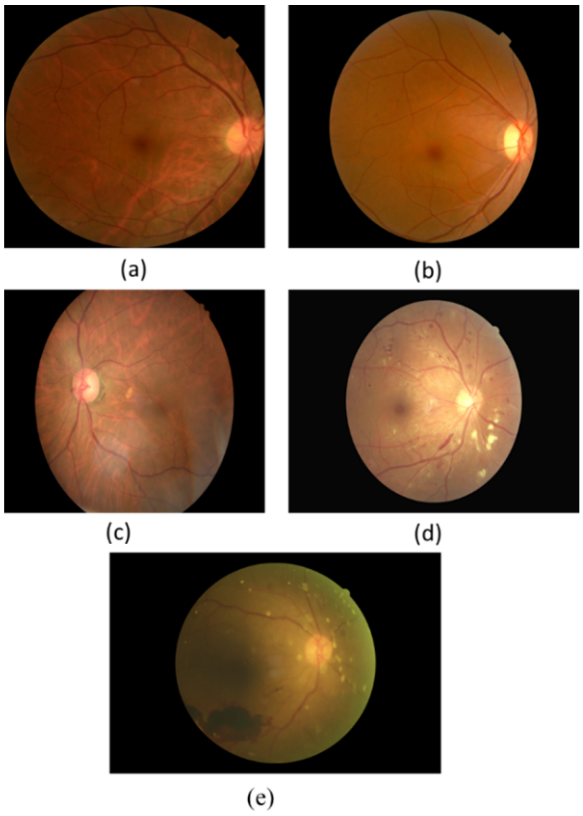

Fig. 1. Ejemplos de retinografías para distintos estadios de DR.

En *computer vision*, la clasificación automática de DR a partir de imágenes de fondo de ojo ha sido ampliamente estudiada mediante modelos de *deep learning*. No obstante, persisten desafíos relevantes: desbalance entre clases con menor representación de estadios severos, variabilidad en adquisición debida a diferencias de iluminación, centrado, escala y dispositivos, y dificultad para capturar lesiones pequeñas y patrones

vasculares finos. Estos factores afectan el rendimiento y la robustez de los clasificadores [9].

Diversos trabajos han explorado estrategias de aumento de datos para mitigar el desbalance. En particular, Peña-Reyes et al. [8] investigan el uso de redes generativas adversariales para generación de retinografías sintéticas y comparan aumento tradicional con aumento basado en modelos generativos. Sus resultados sugieren que la incorporación de imágenes sintéticas puede contribuir a mejorar el desempeño en clasificación cuando las clases de mayor severidad están subrepresentadas.

En paralelo, otros trabajos han abordado la clasificación multiestadio y la localización de lesiones mediante arquitecturas profundas como redes convolucionales y detectores tipo YOLO. Estos enfoques muestran que la detección explícita de biomarcadores puede complementar la clasificación global por estadio [9]. Sin embargo, la mayoría de estas propuestas se centran en tareas discriminativas mientras que la síntesis de imágenes por severidad ha recibido menos atención sistemática.

Desde la perspectiva generativa, la síntesis realista de retinografías implica capturar tanto estadísticas globales como coherencia anatómica fina. Trabajos de síntesis *end-to-end* señalan que la incorporación de condicionamientos explícitos es clave para mejorar la plausibilidad anatómica [7], lo que resalta la dificultad de modelar imágenes médicas con alta estructura interna.

En este contexto, el objetivo de este trabajo es estudiar el uso de redes generativas adversariales para aumento de datos en retinopatía diabética y comparar una configuración no condicional con una configuración condicional más estable. En particular, se analizan dos modelos representativos: una DCGAN entrenada como línea base sobre casos de alta severidad y una arquitectura condicional del tipo Wasserstein GAN con penalización de gradiente (cWGAN-GP) que incorpora la severidad como condición. La comparación se realiza mediante métricas cuantitativas estándar en GANs, como Fréchet Inception Distance (FID) e Inception Score (IS), junto con una evaluación cualitativa por inspección visual, con el fin de discutir el potencial y las limitaciones de estas arquitecturas en el contexto de imágenes médicas.

II. Datos y preprocesamiento

A. Conjuntos de datos

En este trabajo se utilizan retinografías de dos competiciones públicas de Kaggle, ambas basadas en fotografías de fondo de ojo en color con la misma escala de severidad 0–4.

Para la DCGAN se emplea el conjunto APTOS 2019 Blindness Detection [1]. Este dataset contiene imágenes etiquetadas con un diagnóstico ordinal de retinopatía diabética, sin indicar explícitamente la lateralidad del ojo. La distribución por severidad es marcadamente desbalanceada: clase 0 (No DR) con 1800 imágenes, clase 1 con 370, clase 2 con 999, clase 3 con 193 y clase 4 con 295 imágenes. En este trabajo la DCGAN se entrena únicamente sobre las 295 retinografías de severidad 4, de modo que el tamaño efectivo de datos para el modelo generativo es reducido.

Para la cWGAN-GP se utiliza el conjunto Diabetic Retinopathy Detection [2], donde las imágenes se nombran como `{id}_(left|right).jpeg`. Cada archivo incluye el identificador del paciente y la lateralidad del ojo. El archivo `trainLabels.csv` mantiene la misma escala de severidad 0–4. La distribución es también desbalanceada y depende de la combinación nivel–lado, por ejemplo, para nivel 0 hay 10392 retinografías de ojo derecho y 10249 de ojo izquierdo, mientras que para nivel 4 hay 291 y 272 respectivamente (osea que se entrena con 28100 imágenes en total). Este es el dataset utilizado en el repositorio de referencia `retinal-data-augmentation-GAN` [3].

A pesar de las diferencias en organización y nomenclatura, las retinografías de ambas competiciones muestran características visuales muy similares. En los dos casos se observan variaciones de iluminación y contraste, dispersiones en el centrado del fondo de ojo, diferencias en la distancia de captura y casos en los que la retina aparece parcialmente recortada por el borde del sensor. Esta heterogeneidad se mantiene en los preprocesamientos y refleja condiciones de adquisición realistas.

Una extensión natural habría sido entrenar también la cWGAN-GP directamente sobre el conjunto APTOS 2019 para maximizar la comparabilidad con la DCGAN. Por restricciones de tiempo del proyecto esa variante no se llegó a implementar. Dado que las imágenes de ambas competiciones comparten una calidad y un patrón de adquisición muy similares, se espera que la dinámica de entrenamiento y el tipo de resultados cualitativos hubieran sido comparables.

B. Preprocesamiento para la DCGAN (256×256, severidad 4)

Para la DCGAN se utiliza un pipeline centrado en imágenes de severidad 4 del desafío APTOS 2019 [1]. A partir del archivo `train.csv` se seleccionan únicamente los casos con diagnóstico igual a 4 y se cargan sus retinografías RGB desde `train_images`. El preprocesamiento implementado en la clase `AptosDataset` comprende:

- Conversión a espacio de color RGB.
- Centrado automático y reencuadre de la retina sobre un lienzo cuadrado de 256×256 píxeles con fondo negro mediante la transformación `FundusMaskCropPad`, que busca aislar la región retinal y normalizar el encuadre sin intervención manual.
- Conversión a tensor y normalización canal a canal con media $(0.5, 0.5, 0.5)$ y desvío estándar $(0.5, 0.5, 0.5)$, de modo que los valores queden en el rango $[-1, 1]$ en coherencia con la activación `Tanh` del generador.

En algunas corridas se incluyó además una transformación geométrica aleatoria suave (`RandomAffine`) con pequeñas traslaciones y cambios de escala, como forma simple de aumentar ligeramente la variabilidad de posición y tamaño del fondo de ojo. Esta componente de aumento no tuvo un papel central en el análisis de resultados.

El modelo DCGAN opera entonces sobre imágenes de tamaño $(3, 256, 256)$ y aprende a sintetizar retinografías proliferativas manteniendo la viñeta retinal global y la colorimetría dominante de los casos severos.

C. Preprocesamiento para la cWGAN-GP (120×120)

Para la cWGAN-GP se genera primero el conjunto `dataset_120_square` a partir del desafío Diabetic Retinopathy Detection [2]. A grandes rasgos, el script de preprocesamiento recorre el archivo `trainLabels.csv`, localiza cada retinografía y construye una versión estandarizada de menor resolución adecuada para el entrenamiento del modelo generativo.

El pipeline aplicado puede resumirse en tres pasos:

- Conversión de todas las imágenes originales a espacio de color RGB.
- Redimensionado uniforme a 120×120 píxeles sobre un lienzo cuadrado, sin recortes manuales ni segmentación explícita de la retina.
- Almacenamiento de las imágenes procesadas en la carpeta `dataset_120_square`, junto con un archivo `trainLabels.csv` simplificado que asocia cada nombre de archivo con su nivel de severidad.

Durante el entrenamiento, la clase `Dataset` carga estas imágenes de 120×120 y aplica una transformación estándar formada por `ToTensor` y `Normalize` con media $(0.5, 0.5, 0.5)$ y desvío estándar $(0.5, 0.5, 0.5)$, de forma análoga a la DCGAN.

Este pipeline de preprocesamiento se mantiene alineado con el del repositorio de referencia `retinal-data-augmentation-GAN` [3], donde se reporta un entrenamiento estable de modelos cGAN y cWGAN sobre el mismo conjunto de datos. La decisión de no modificarlo facilita la comparabilidad metodológica con ese trabajo.

D. Condición: severidad y lateralidad

El código original del repositorio `retinal-data-augmentation-GAN` explora modelos condicionales con diez clases, que corresponden a la combinación de severidad y lateralidad:

$$(0r, 1r, 2r, 3r, 4r, 0l, 1l, 2l, 3l, 4l) \mapsto (0, 1, 2, 3, 4, 5, 6, 7, 8, 9),$$

donde r indica ojo derecho y l indica ojo izquierdo. En este esquema el generador recibe una condición que especifica simultáneamente el nivel de retinopatía diabética y el lado del ojo, lo que permite muestrear imágenes por combinación severidad y lateralidad.

En este trabajo se replican de forma exploratoria entrenamientos con esta codificación de diez clases. Posteriormente, para la comparación final entre la DCGAN y la cWGAN-GP se adopta una codificación simplificada de cinco clases basada sólo en la severidad 0–4. En la implementación, la clase `Dataset` sigue leyendo la columna `side`, pero la función `processcondition` mapea tanto `_left` como `_right` al mismo identificador de severidad. De esta manera se reutiliza el mismo pipeline de datos del código condicional original, mientras que la condición efectiva que ve la cWGAN-GP en los experimentos comparativos corresponde únicamente al nivel 0–4.

Las pruebas con diez clases y las de cinco clases muestran un comportamiento cualitativo similar en cuanto a la calidad global de las muestras generadas. Sin embargo, desde un punto de vista clínico la distinción entre ojo izquierdo y derecho resulta relevante, sobre todo en modelos que trabajan por parches. Por esta razón el modelado explícito de la lateralidad se considera una extensión natural para trabajos futuros, una vez estabilizada la comparación principal entre las dos arquitecturas seleccionadas.

E. Estandarización geométrica y normalización

En ambos pipelines las imágenes originales presentan variación en distancia de captura, centrado del disco óptico y la mácula, relación de aspecto y tamaño relativo del campo retinal dentro del marco. Para homogeneizar el espacio de entrada y facilitar el aprendizaje de las arquitecturas convolucionales y lineales descritas se fuerza un formato cuadrado con tamaño fijo: 256×256 píxeles en la DCGAN y 120×120 píxeles en la cWGAN-GP. Las diferencias específicas de cada pipeline de preprocesamiento se detallan en las subsecciones previas.

En todas las configuraciones, las imágenes se normalizan canal a canal al rango $[-1, 1]$, en coherencia con la salida `Tanh` de los generadores en ambos modelos.

III. Arquitecturas y configuración de entrenamiento

A. DCGAN

La primera arquitectura es una DCGAN no condicional entrenada sobre retinografías de severidad 4 preprocesadas a tamaño 256×256 y normalizadas a $[-1, 1]$. El vector latente $z \in \mathbb{R}^{256}$ se muestrea de una normal estándar y se proyecta mediante una pila de capas convolucionales transpuestas hasta obtener imágenes RGB de tamaño $(3, 256, 256)$ con activación `Tanh`.

El discriminador recibe imágenes reales o generadas y devuelve un logit escalar que se interpreta como evidencia de "real vs falsa". La arquitectura usa convoluciones estratificadas, `BatchNorm` en capas intermedias y activaciones `ReLU` en el generador y `LeakyReLU` en el discriminador. La Tabla I resume la arquitectura concreta usada en los experimentos.

Se empleó `BCEWithLogitsLoss` para generador y discriminador, optimizador Adam con tasa de aprendizaje $2 \cdot 10^{-4}$ y parámetros $(\beta_1, \beta_2) = (0{,}5, 0{,}999)$. Para estabilizar el entrenamiento se usó *label smoothing* en el discriminador (etiquetas reales en el rango $[0{,}9, 1]$ y falsas en $[0, 0{,}1]$)..

B. cWGAN-GP

La segunda arquitectura es una cWGAN-GP condicional entrenada sobre el conjunto `dataset_120_square`, con retinografías RGB de 120×120. La condición codifica la severidad en cinco clases (0-4) y se incorpora mediante una capa `Embedding` tanto en el generador como en el crítico.

TABLE I
ARQUITECTURA DCGAN

Generador	Discriminador
Entrada: $z \in \mathbb{R}^{256}$	Entrada: imagen $(3, 256, 256)$
ConvT $(256 \to 512)$, BN, ReLU	Conv $(3 \to 32)$, LeakyReLU
ConvT $(512 \to 256)$, BN, ReLU	Conv $(32 \to 64)$, BN, LeakyReLU
ConvT $(256 \to 128)$, BN, ReLU	Conv $(64 \to 128)$, BN, LeakyReLU
ConvT $(128 \to 64)$, BN, ReLU	Conv $(128 \to 256)$, BN, LeakyReLU
ConvT $(64 \to 32)$, BN, ReLU	Conv $(256 \to 512)$, BN, LeakyReLU
ConvT $(32 \to 16)$, BN, ReLU	Conv $(512 \to 1024)$, BN, LeakyReLU
ConvT $(16 \to 3)$, Tanh	Conv $(1024 \to 1)$, salida escalar

1) Objetivo WGAN-GP: En lugar de optimizar una entropía cruzada, la cWGAN-GP optimiza una aproximación a la distancia de Wasserstein entre la distribución real p_r y la generada p_g. El discriminador se reemplaza por un crítico $C(x, c)$ que asigna un score real a cada par imagen/condición. Dado $z \sim p_z = \mathcal{N}(0, I)$ y condición c, el generador produce

$$\hat{x} = G(z, c).$$

La pérdida del crítico se define como

$$\mathcal{L}_C = -\mathbb{E}_{x\sim p_r}[C(x, c)] + \mathbb{E}_{z\sim p_z}[C(G(z, c), c)] + \lambda_{gp}\,\mathbb{E}_{\tilde{x}}\Big[(\|\nabla_{\tilde{x}} C(\tilde{x}, c)\|_2 - 1)^2\Big], \quad (1)$$

donde $\tilde{x} = \alpha x + (1 - \alpha)\hat{x}$ con $\alpha \sim U(0, 1)$ y λ_{gp} controla el peso de la penalización de gradiente.

La pérdida del generador es

$$\mathcal{L}_G = -\mathbb{E}_{z\sim p_z}[C(G(z, c), c)]. \quad (2)$$

Durante el entrenamiento se monitoriza también

$$\widehat{W} = \mathbb{E}_{x\sim p_r}[C(x, c)] - \mathbb{E}_{z\sim p_z}[C(G(z, c), c)],$$

que actúa como estimador empírico de la distancia de Wasserstein entre datos reales y generados.

2) Arquitectura MLP y condicionamiento: El generador y el crítico son redes totalmente conectadas que operan sobre vectores aplanados. La condición se representa mediante un embedding discreto y se concatena tanto al ruido latente como a la imagen aplanada. La Tabla II resume la arquitectura utilizada.

C. Resumen de hiperparámetros

En la Tabla III se resumen los hiperparámetros de entrenamiento más relevantes de cada arquitectura. Se incluyen sólo aquellos valores que permiten comparar directamente la configuración de la DCGAN y de la cWGAN-GP.

IV. RESULTADOS

A. DCGAN

Se entrenó una DCGAN no condicional para generar retinografías de severidad 4 a resolución 256×256 píxeles. La inspección visual evidencia que, aunque el generador aprende la forma global de la viñeta retinal, las muestras presentan texturas repetitivas y artefactos de rejilla, sin estructuras anatómicas consistentes (por ejemplo: disco óptico o patrón vascular). Además, la diversidad intramuestra es limitada, con salidas muy similares entre sí hacia el final del entrenamiento.

TABLE II
ARQUITECTURA DE LA CWGAN-GP (MLP CONDICIONAL).

Generador $G(z, c)$	**Crítico** $C(x, c)$
Entrada: $z \in \mathbb{R}^{20}$, $c \in \{0, \ldots, 4\}$	Entrada: imagen $(3, 120, 120)$, c
Embedding de condición $e(c) \in \mathbb{R}^{20}$	Embedding de condición $e(c) \in \mathbb{R}^{20}$
Concat $[e(c); z] \in \mathbb{R}^{40}$	Aplanado de imagen $x \mapsto \mathbb{R}^{3\cdot120\cdot120}$
Linear $40 \to 64$, LeakyReLU	Concat $[\mathrm{vec}(x); e(c)] \in \mathbb{R}^{3\cdot120\cdot120+20}$
Linear $64 \to 128$, BatchNorm1d, LeakyReLU	Linear $(3 \cdot 120 \cdot 120 + 20) \to 64$, LeakyReLU
Linear $128 \to 128$, BatchNorm1d, LeakyReLU	Linear $64 \to 64$, Dropout(0,4), LeakyReLU
Linear $128 \to 256$, BatchNorm1d, LeakyReLU	Linear $64 \to 1$, salida escalar
Linear $256 \to 128$, BatchNorm1d, LeakyReLU	
Linear $128 \to 64$, BatchNorm1d, LeakyReLU	
Linear $64 \to 3 \cdot 120 \cdot 120$, Tanh	

TABLE III
HIPERPARÁMETROS CLAVE DE ENTRENAMIENTO

	DCGAN 256×256, sev 4	cWGAN-GP 120×120, sev 0–4
Condicional	No	Sí (sev.)
Dimensión del ruido d_z	256	20
Dimensión del embedding d_e	–	20
Batch size	16	16
Épocas	100	100
LR generador	$2 \cdot 10^{-4}$	$2 \cdot 10^{-4}$
LR discriminador/crítico	$2 \cdot 10^{-4}$	$2 \cdot 10^{-3}$
n_{critic}	-	5
λ_{gp}	–	10

La Fig. 2 muestra un minibatch de imágenes reales y generadas, junto con la distribución de intensidades de píxel (reescaladas a $[0, 1]$), para la última época de entrenamiento, que coincide con la época seleccionada como "best epoch". Aunque el modelo no llega a capturar detalles clínicos relevantes, en esta época las muestras resultan ligeramente más coherentes visualmente que en etapas tempranas, por lo que se tomó como punto de referencia para la inferencia y el cálculo de FID/IS. Se observa que las imágenes sintéticas concentran más masa de probabilidad en intensidades bajas y exhiben un balance distinto de tonos medios, coherente con el aspecto más oscuro y con artefactos de la salida del generador.

La evolución de las pérdidas del discriminador y del generador se muestra en la Fig. 3. Tras un pico inicial, la pérdida del generador desciende rápidamente pero se mantiene en valores relativamente altos y oscilantes, mientras que la del discriminador cae pronto y permanece baja.

Desde el punto de vista del entrenamiento, la DCGAN no alcanza un equilibrio estable: el discriminador se vuelve muy fuerte muy rápido y el generador no logra aprender lo suficiente, lo que se refleja en las imágenes generadas,

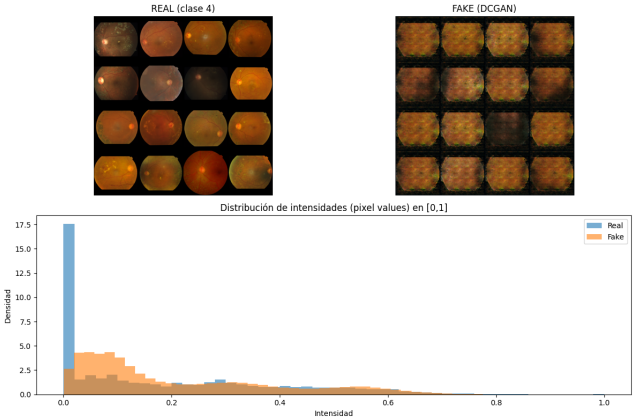

Fig. 2. DCGAN (best epoch): Retinografías Reales vs Generadas para severidad 4, junto con el histograma de intensidades de píxel en $[0, 1]$.

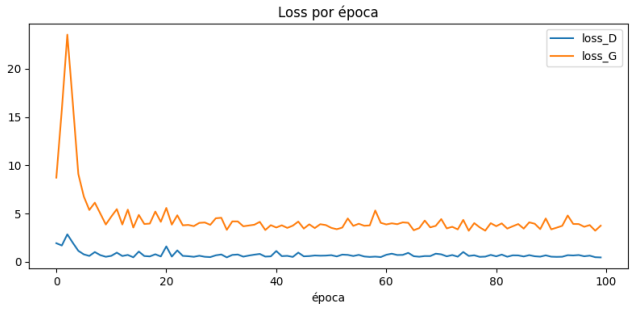

Fig. 3. Evolución de las pérdidas del discriminador ($\mathcal{L}_D$) y del generador ($\mathcal{L}_G$) durante el entrenamiento de la DCGAN.

ruidosas y muy similares entre sí. Por restricciones de tiempo, aunque se ensayaron distintas configuraciones, no se logró un ajuste fino ni de la arquitectura ni de los hiperparámetros. En consecuencia, esta DCGAN se interpreta como una línea base más que como el mejor modelo posible dentro de esta familia de arquitecturas.

Cuantitativamente, se obtuvo un FID global de 347.321, indicando una separación marcada entre la distribución de imágenes reales y generadas en el espacio de características de Inception. El Inception Score (IS) resultó bajo (1.208 ± 0.052); sin embargo, esta métrica es poco informativa en retinografías debido a que el clasificador subyacente está entrenado en ImageNet y no está calibrado para el dominio médico. Por ello, en el resto del trabajo se prioriza FID como indicador principal en la comparación entre modelos.

B. cWGAN-GP

Aunque el modelo condicional se entrena con severidades 0–4, en las figuras se muestran principalmente ejemplos de severidad 4 para compararlo directamente con la DCGAN.

A nivel global el modelo genera: campo circular, fondo negro y distribución general de color/iluminación. Al comparar condiciones extremas (por ejemplo: severidad 0 vs. 4) se observan diferencias globales de contraste y tonalidad, sin embargo, los detalles finos clínicamente relevantes (microaneurismas,

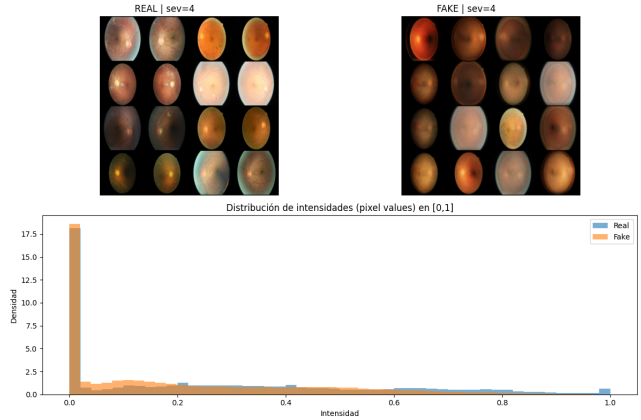

Fig. 4. cWGAN-GP (epoch 80): Retinografías Reales vs Generadas para severidad 4, junto con el histograma de intensidades de píxel en $[0, 1]$.

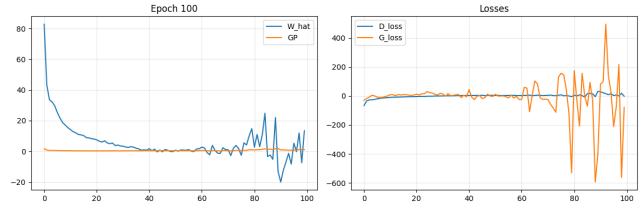

Fig. 5. Evolución de la pérdida del crítico, la pérdida del generador, la estimación $\widehat{W}$ y la penalización de gradiente durante el entrenamiento de la cWGAN-GP.

exudados puntuales, hemorragias pequeñas) no aparecen de forma consistente, y la vasculatura tiende a no verse.

La Fig. 4 muestra la comparación real vs. generado para severidad 4 en la época 80, junto con el histograma de intensidades de píxel. Esta época se seleccionó como punto de inferencia porque todavía producía muestras visualmente coherentes antes de que el entrenamiento comenzara a volverse inestable en épocas posteriores. A diferencia del caso DCGAN, las distribuciones de intensidades se asemejan más a nivel global, aunque esta coincidencia no implica equivalencia anatómica, ya que el histograma ignora la estructura espacial y la topología vascular.

La Fig. 5 resume la dinámica de entrenamiento de la cWGAN-GP, mostrando la pérdida del crítico, la pérdida del generador, la estimación $\widehat{W}$ de la distancia de Wasserstein y la penalización de gradiente (GP). Tras un período de calentamiento, $\widehat{W}$ disminuye y luego oscila en un rango acotado mientras la GP se mantiene cercana a 1, lo que sugiere que el entrenamiento es más estable que en la DCGAN.

Se calculó FID y se obtuvo:

- **FID global: 205.821**
- **FID por severidad**:
  - sev 0: 211.877
  - sev 1: 207.615
  - sev 2: 196.485
  - sev 3: 203.948
  - sev 4: 185.512

TABLE IV
RESUMEN DE CONFIGURACIONES Y MÉTRICAS PARA LOS DOS MODELOS CONSIDERADOS.

Modelo	Resolución	Condicional	FID	IS
DCGAN	256×256	No	347.3	1.21 ± 0.05
cWGAN-GP	120×120	Sí (sev 0-4)	205.8	2.03 ± 0.11

Valores aún altos de FID son consistentes con la observación visual de falta de textura fina y diferencias en estructuras locales. No obstante, la comparación con la DCGAN muestra una reducción sustancial de la distancia distribucional (de 347.3 a 205.8), lo que indica que el modelo condicional se acerca más a la distribución real. Además, InceptionV3 fue entrenada en ImageNet, por tanto, su embedding puede no capturar adecuadamente patrones clínicos de retina y puede inflar la distancia en este dominio.

Se calculó Inception Score (IS) con InceptionV3:

- **IS global:** 2.034 ± 0.106
- **IS por severidad**:
  - sev 0: 2.063 ± 0.154
  - sev 1: 2.014 ± 0.147
  - sev 2: 1.996 ± 0.077
  - sev 3: 1.840 ± 0.101
  - sev 4: 1.859 ± 0.109

Los valores se mantienen en un rango estrecho, lo que sugiere ausencia de un colapso fuerte en una severidad particular bajo esta métrica. El incremento respecto a la DCGAN (de ≈ 1.21 a ≈ 2.03) es coherente con muestras más variadas y mejor reconocibles por Inception. No obstante, IS depende de un clasificador entrenado en ImageNet, por lo que debe interpretarse como comparación interna entre corridas, no como medida de realismo clínico.

C. Resumen comparativo DCGAN vs. cWGAN-GP

La Tabla IV resume las configuraciones y métricas cuantitativas de ambos modelos.

V. DISCUSIÓN

Los resultados muestran un contraste claro entre ambos modelos, pero también entre los regímenes de datos en los que operan. La DCGAN se entrena sólo con 295 retinografías de severidad 4 a 256×256, mientras que la cWGAN-GP utiliza unas 28 mil imágenes de severidad 0–4 a 120×120. Esta diferencia en cantidad y diversidad de ejemplos condiciona fuertemente la dinámica adversarial y ayuda a interpretar por qué uno de los modelos resulta mucho más estable que el otro.

En la DCGAN el entrenamiento no alcanza un equilibrio estable. La pérdida del discriminador cae rápido y se mantiene baja, mientras que la del generador permanece alta y oscilante. Con tan pocas imágenes reales y una arquitectura convolucional profunda a 256×256, el discriminador puede sobreajustar con facilidad y aprender una frontera casi perfecta entre datos reales y sintéticos. En ese escenario la señal de gradiente que recibe el generador se debilita y aparecen fenómenos descritos en la literatura de GANs como "discriminador demasiado fuerte", gradiente casi nulo y colapso hacia patrones ruidosos. Esto se traduce en imágenes con artefactos de rejilla, texturas poco realistas y baja diversidad, junto con un FID muy elevado y un IS bajo. Por tanto, los resultados de esta DCGAN deben interpretarse como una línea base en un régimen de datos especialmente exigente, más que como un veredicto definitivo sobre la familia DCGAN.

En la cWGAN-GP la dinámica de entrenamiento es más estable. La evolución conjunta de la pérdida del crítico, la pérdida del generador, la estimación $\widehat{W}$ de la distancia de Wasserstein y la penalización de gradiente coincide con el comportamiento esperado de WGAN-GP. El crítico sigue proporcionando gradientes útiles incluso cuando las distribuciones real y generada están alejadas, y la restricción 1-Lipschitz mediante *gradient penalty* evita que el modelo entre en un régimen de saturación. Además, el modelo ve muchas más muestras reales y con mayor variabilidad de severidad, lo que reduce el riesgo de sobreajuste extremo del crítico. Esta combinación de función de pérdida y mayor disponibilidad de datos se refleja en muestras visualmente más coherentes y en mejores valores de FID e IS respecto de la DCGAN.

El condicionamiento explícito por severidad también contribuye a mejorar el comportamiento del modelo. En la cWGAN-GP la información de clase se incorpora mediante *embeddings* en generador y crítico, lo que reduce la ambigüedad del problema generativo. A nivel cualitativo esto se observa en diferencias de contraste y tonalidad entre severidades, y a nivel cuantitativo en una reducción sustancial de la distancia distribucional y un aumento del Inception Score. La condición actúa como una guía adicional que facilita que el modelo cubra mejor los modos de la distribución real y, al entrenar simultáneamente con imágenes de severidad 0–4 en lugar de una sola clase, el generador recibe ejemplos más variados que enriquecen la representación aprendida y aumentan la diversidad de las muestras sintéticas.

A pesar de estas mejoras, los dos modelos comparten una limitación importante. Ninguno consigue reproducir de forma consistente las estructuras clínicamente más relevantes como la vasculatura fina, los microaneurismas, los exudados puntuales o las hemorragias pequeñas. Las resoluciones utilizadas (256×256 y 120×120) y las arquitecturas empleadas favorecen que el generador modele bien las estadísticas globales, como la forma del campo, el fondo oscuro y la distribución general de intensidades, pero no garantizan la síntesis de detalles de alta frecuencia espacial. Esto produce retinografías que resultan razonables a primera vista, pero con suavizado y pérdida de texturas finas, lo que limita su uso directo como sustituto de imágenes clínicas reales.

VI. CONCLUSIONES

Se implementó un *pipeline* reproducible para síntesis condicional de retinografías mediante cWGAN-GP (MLP) a 120×120 RGB, condicionando por severidad (5 condiciones), y se comparó su desempeño con una DCGAN no condicional a

256×256 entrenada sólo sobre severidad 4. El modelo condicional logra resultados coherentes a nivel global y permite un monitoreo estable mediante $\widehat{W}$ y la penalización de gradiente, mejorando de forma clara las métricas FID/IS respecto al modelo base DCGAN.

Sin embargo, las imágenes generadas por la cWGAN-GP siguen sin reproducir de forma consistente las estructuras clínicas finas (vasculatura, microaneurismas, exudados puntuales), por lo que todavía están lejos de un "ojo sintético" indistinguible de uno real. En este contexto, los resultados son prometedores como generador de muestras adicionales para tareas de clasificación (en línea con trabajos previos de aumento de datos basados en GANs), pero no son suficientes si el objetivo es emular con precisión el aspecto de una retinografía clínica individual.

Como trabajo futuro se propone:

- Incorporar guías estructurales (segmentación de vasos o mapas de lesiones) como condición adicional, para reforzar la coherencia anatómica.
- Modelar explícitamente severidad y lateralidad (ojo izquierdo/derecho) como condición conjunta, extendiendo la codificación a las diez clases originales del *pipeline* condicional y evaluando su impacto en FID, IS y en la utilidad para aumento de datos.
- Explorar arquitecturas multiescala y convolucionales específicas para retina, capaces de preservar detalles de alta frecuencia espacial.
- Investigar otras familias de modelos más allá de las GANs clásicas, por ejemplo modelos de difusión, modelos multimodales de gran escala o el *fine-tuning* de redes preentrenadas en imágenes naturales y médicas, con el objetivo de acercarse a retinografías sintéticas que resulten clínicamente indistinguibles de las reales.

REFERENCES

[1] Kaggle, "APTOS 2019 Blindness Detection," *Kaggle Competition*, 2019. [Online]. Available: https://www.kaggle.com/competitions/aptos2019-blindness-detection/data. Accessed: 2025.

[2] Kaggle, "Diabetic Retinopathy Detection," *Kaggle Competition*, 2015. [Online]. Available: https://www.kaggle.com/competitions/diabetic-retinopathy-detection. Accessed: 2025.

[3] Icebearbear, "retinal-data-augmentation-GAN," GitHub repository. [Online]. Available: https://github.com/Icebearbear/retinal-data-augmentation-GAN/tree/main. Accessed: 2025.

[4] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved Training of Wasserstein GANs," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[5] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017. (Introduces FID.)

[6] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved Techniques for Training GANs," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2016. (Introduces Inception Score.)

[7] P. Costa, A. Galdran, M. I. Meyer, M. Niemeijer, M. D. Abràmoff, A. Mendonça, and A. Campilho, "End-to-End Adversarial Retinal Image Synthesis," *IEEE Transactions on Medical Imaging*, vol. 37, no. 3, pp. 781–791, 2018, doi:10.1109/TMI.2017.2759102.

[8] M. Heenaye-Mamode Khan, N. Z. Mungloo, Z. Mungloo-Dilmohamud, C. Peña-Reyes, and K. Jhumka, "Investigating on Data Augmentation and Generative Adversarial Networks (GANs) for Diabetic Retinopathy," 2022.

[9] W. L. Alyoubi, M. F. Abulkhair, and W. M. Shalash, "Diabetic Retinopathy Fundus Image Classification and Lesions Localization System Using Deep Learning," *Sensors*, vol. 21, no. 11, p. 3704, May 2021.

[10] D. W. Parke III, "How to Classify the Diabetic Eye," American Academy of Ophthalmology, Young Ophthalmologists (YO Info), 2020. [Online]. Available: https://www.aao.org/young-ophthalmologists/yo-info/article/how-to-classify-diabetic-eye