Aumento de imágenes usando una GAN condicional para clasificación de imágenes médicas.

Por:

Héctor Anaya Sánchez

Tesis sometida como requisito
para obtener el grado de:

Maestro en Ciencias en el Área de Ciencias Computacionales

por el

Instituto Nacional de Astrofísica,
Óptica y Electrónica

Noviembre, 2024

Tonantzintla, Puebla

Dirigida por:

Dr. Leopoldo Altamirano Robles

Índice general

Lista de figuras

Lista de tablas

Agradecimientos

Agradezco profundamente a mi familia por su apoyo incondicional durante este periodo, especialmente a mi compañera de vida, Jazmín, por los incontables momentos de motivación y aliento. También extiendo mi gratitud a mi amigo Migan y al resto de mis amigos por su constante respaldo a lo largo de este tiempo. Agradezco de manera especial a mi asesor, el doctor Leopoldo Altamirano Robles, por creer en mí, brindarme su apoyo y guiarme con su excelente coordinación. Al INAOE, por los valiosos conocimientos impartidos a lo largo de la maestría, y a CONACYT, por el apoyo otorgado mediante la beca No. 1148641.

Resumen

El problema de la falta de datos y el desbalance en la clasificación de imágenes médicas es un desafío importante en la actualidad. Los objetivos y metas de esta investigación se centran en desarrollar una metodología basada en una Red Adversarial Generativa condicional (cGAN, *conditional Generative Adversarial Networks*) para aumentar la cantidad y diversidad de imágenes retinales, incorporando información relevante en la definición de la condición. Esta línea de investigación tiene como justificación mejorar la precisión y rendimiento de los sistemas de clasificación, lo que podría tener un impacto significativo en el diagnóstico y tratamiento de enfermedades. En el desarrollo de esta investigación, se emplearon Redes Adversariales Generativas Wasserstein con penalidad de gradiente (WGAN-GP, *Wasserstein Generative Adversarial Network with Gradient Penalty*), junto con técnicas de extracción de lesiones y transferencia de estilo para generar imágenes sintéticas de alta calidad. Los resultados obtenidos fueron evaluados utilizando métricas estándar de redes GAN, como la Distancia Frechet-Inception (FID, *Frechet Inception Distance*), el Error Cuadrático Medio (MSE, *Mean Squeared Error*) y el Índice de Similitud Estructural (SSIM, *Structural Similarity Index Measure*). Los valores obtenidos de FID fueron bajos, indicando una alta similitud estadística entre las imágenes reales y generadas. El MSE y el SSIM también mostraron resultados favorables, sugiriendo que las imágenes generadas conservan la estructura y los detalles finos necesarios

para un análisis preciso. Además, las imágenes sintéticas fueron sometidas a evaluación por expertos, quienes clasificaron las imágenes con una precisión del 56.66 %, lo que subraya la calidad y realismo de las imágenes generadas. Estos resultados demuestran que la metodología propuesta no solo puede aumentar la cantidad de datos disponibles, sino también mejorar su calidad, contribuyendo significativamente a la formación de modelos de diagnóstico más precisos.

Abstract

The problem of data scarcity and imbalance in the classification of medical images is a significant challenge today. The objectives and goals of this research focus on developing a methodology based on conditional Generative Adversarial Networks (cGAN) networks to increase the quantity and diversity of fundus images, incorporating relevant information into the condition definition. This line of research is justified by the potential to improve the accuracy and performance of classification systems, which could have a significant impact on the diagnosis and treatment of diseases. In the development of this research, Wassertein Generative Adversarial Networks with Gradient Penalty (WGAN-GP) were employed along with lesion extraction and style transfer techniques to generate high-quality synthetic images. The obtained results were evaluated using standard GAN metrics, such as Frechet Inception Distance (FID), Mean Squared Error (MSE), and Structural Similarity Index Measure (SSIM). The FID values were low, indicating a high statistical similarity between the real and generated images. The MSE and SSIM also showed favorable results, suggesting that the generated images retain the structure and fine details necessary for precise analysis. Additionally, the synthetic images were evaluated by experts, who classified the images with an accuracy of 58 %, highlighting the quality and realism of the generated images. These results demonstrate that the proposed methodology not only increases the available data quantity, but also improves its

quality, significantly contributing to the development of more precise and robust diagnostic models.

Capítulo 1

Introducción

La clasificación precisa de imágenes médicas es de vital importancia para el diagnóstico y tratamiento de diversas enfermedades. Sin embargo, uno de los desafíos que enfrentan los sistemas de clasificación es la falta de datos o bases de datos desbalanceadas, lo que puede afectar su precisión y rendimiento. Para afrontar este desbalance en datos, se han propuesto varias técnicas en las que se le hacen modificaciones geométricas a las imágenes (aumento de imágenes tradicional, rotación, traslación, *zoom*). Con el auge de la inteligencia artificial y el aprendizaje profundo se han creado nuevos métodos con los cuales se pueden generar imágenes sintéticas lo suficientemente reales como para que un experto no pueda distinguirlas.

Las Redes Adversariales Generativas (GAN, *Generative Adversarial Network*, Goodfellow et al. (2014)) son capaces de crear imágenes sintéticas con un alto nivel de realismo, gracias a ellas podemos generar miles de imágenes diversas que nos ayuden a balancear bases de datos, o aumentar bases de datos con pocas imágenes. Recientemente, se han utilizado para aumentar bases de datos de retinopatía diabética (PathoGAN, Sampath et al. (2021), DR-GAN, Zhou et al. (2022)) con una baja distancia de Fretchet *Inception* (FID, Fretchet *Inception Distance*) la cual nos dice que tan realistas son sus imágenes (entre más bajo mejor). Con este tipo de redes planteamos generar imágenes de retinopatía diabética.

En cuanto a los aspectos específicos del problema, se pretende desarrollar una técnica que permita condicionar la generación de imágenes en función de información relevante, como la presencia de patologías específicas o características anatómicas particulares. Lo anterior permitirá generar conjuntos de datos sintéticos que reflejen la diversidad presente en las imágenes médicas reales, lo que a su vez mejorará el rendimiento de los sistemas de clasificación.

1.1. Planteamiento del problema

En el campo del análisis médico de imágenes, la calidad y la cantidad de los datos disponibles son cruciales para el desarrollo de modelos robustos de *machine learning*. Particularmente en la oftalmología, el uso de imágenes de retina para el diagnóstico y monitoreo de enfermedades tales como la retinopatía diabética, hipertensión, degeneración macular por edad, se basan fuertemente en algoritmos de *deep learning*. Sin embargo, la falta de imágenes médicas etiquetadas nos da un desafío significante debido a los costos y tiempos que consume el etiquetado de expertos.

Además, las imágenes retinales son altamente sensibles a variaciones en las condiciones de captura, lo que puede llevar a un sobreajuste y una pobre generalización de los modelos entrenados en bases de datos limitadas. Si bien el aumento de datos se ha reconocido como una solución para mitigar estos problemas al incrementar artificialmente el tamaño y la diversidad de un conjunto de datos de entrenamiento, las técnicas tradicionales como la rotación, el escalado y el reflejo no son suficientes para generar imágenes médicas que capturen de manera adecuada las patologías o los detalles intrínsecos esenciales para el diagnóstico.

Esta tesis aborda el problema de la escasez de datos al proponer el uso de una Red Adversarial Generativa de Wasserstein con Penalidad de Gradiente (WGAN-GP, *Wasserstein Generative Adversarial Network with Gradient Penalty*, Gulrajani

et al. (2017)), transferencia de estilo y extracción de lesiones. Este modelo es un nuevo método para la generación sintética de imágenes retinales que sean lo suficientemente realistas y diversas para aumentar efectivamente conjuntos de datos existentes. La habilidad de la WGAN-GP de producir imágenes de alta calidad sin el modo colapso presenta una dirección prometedora para enriquecer conjuntos de datos de imágenes retinales y, por lo tanto, mejorar el rendimiento y generalidad de modelos de *deep learning* en diagnóstico oftalmológico. Este trabajo tiene como objetivo evaluar la efectividad de utilizar un modelo WGAN-GP para aumento de datos en el contexto del análisis de imágenes retinales.

1.2. Objetivo general

Desarrollar una metodología que incorpore información en la condición de una Red Adversarial Generativa condicional (cGAN, *conditional Generative Adversarial Network*), permitiendo la reproducción de imágenes de alta calidad que sean comparables a las imágenes reales, para mejorar la precisión y la diversidad en los conjuntos de datos utilizados para entrenar métodos de diagnóstico médico.

1.3. Objetivos específicos

Los objetivos específicos se listan a continuación:

- Caracterizar las propiedades de las enfermedades para asegurarse que las imágenes sintéticas sean visualmente coherentes y se asemejen a imágenes médicas reales.
- Identificar las diferentes técnicas usadas para incorporar la condición en una cGAN.

- Implementar la metodología propuesta en un entorno de prueba, evaluando su efectividad y eficiencia en la generación de imágenes sintéticas de alta calidad.
- Desarrollar un método que permita evaluar las imágenes sintéticas generadas.

1.4. Organización de tesis

Los capítulos del presente trabajo de tesis se estructuran de la siguiente manera:

- Capítulo 2: Marco teórico. En esta sección se discuten los conceptos fundamentales y el estado del arte relacionado con el uso de redes tipo GAN en el ámbito de las imágenes médicas. Se detallan las técnicas de cGAN, WGAN-GP, extracción de lesiones y transferencia de estilo, además de explicar las métricas utilizadas para evaluar la calidad de las imágenes generadas.
- Capítulo 3: Trabajo relacionado. Se describe la metodología desarrollada para incorporar información específica en la condición de una cGAN. Se explican los pasos seguidos para la caracterización de las propiedades de las enfermedades, la identificación de técnicas de incorporación de condiciones, y el desarrollo de métodos de evaluación de imágenes sintéticas.
- Capítulo 4: Método propuesto. Se detallan los procedimientos llevados a cabo para implementar la metodología propuesta en un entorno de prueba. Este capítulo incluye la configuración del entorno de desarrollo, los algoritmos utilizados, y los desafíos enfrentados durante la implementación.
- Capítulo 5: Experimentos. Se presentan los resultados obtenidos de la implementación de la metodología. Se incluyen análisis detallados de las imágenes generadas, comparaciones con imágenes reales, y la evaluación de la efectividad de las imágenes sintéticas mediante métricas cuantitativas y cualitativas.

- Capítulo 6: Conclusiones y Trabajo futuro. Esta sección incluye las conclusiones generales de la investigación, destacando las contribuciones significativas del estudio, y propone posibles direcciones para futuros trabajos que podrían mejorar y expandir los hallazgos de esta investigación.

CAPÍTULO 2

MARCO TEÓRICO

Este capítulo presenta los fundamentos para el entendimiento de los aspectos teóricos sobre las WGAN-GP en el aumento de conjuntos de datos de imágenes retinales. La discusión está estructurada en 6 secciones principales, (1) una breve introducción a la clasificación de imágenes, (2) seguido de una introducción más general de la generación de imágenes, (3) una introducción a las redes tipo GAN, y sus variaciones, (4) sobre la transferencia de estilo, (5) donde se detalla la extracción de características y (6) donde se hablan de las métricas de evaluación.

2.1. Clasificación de imágenes

La clasificación de imágenes, en su esencia, implica la asignación de etiquetas o categorías a las imágenes según sus características visuales. En el ámbito médico, esta tarea contiene una significativa relevancia, ya que permite la identificación y diagnóstico preciso de diversas patologías. Este proceso, clave en la interpretación de imágenes médicas, se vuelve especialmente crítico en el análisis de imágenes de retinas, donde la detección temprana de anomalías puede influir directamente en la atención y tratamiento oftalmológico. Enfocándonos en la retina, una estructura anatómica compleja y vital, la clasificación de imágenes médicas se enfrenta a

desafíos únicos. La identificación de características sutiles, como microaneurismas o exudados, demanda una precisión excepcional. Además, la clasificación efectiva de retinas normales y patológicas se convierte en un componente crucial para la detección temprana de enfermedades como la retinopatía diabética. La capacidad de clasificar imágenes de retinas de manera precisa se traduce directamente en una mejora en la atención al paciente, permitiendo intervenciones tempranas que pueden ser decisivas para preservar la salud ocular.

2.2. Generación de imágenes

En esencia, la generación de imágenes representa un avance crucial en el procesamiento de datos médicos, permitiendo la creación de representaciones visuales que pueden ampliar y diversificar conjuntos de datos limitados. En este contexto, las GAN emergen como una herramienta revolucionaria. Estas redes neuronales compuestas por generadores y discriminadores trabajan en conjunto, aprendiendo y mejorando continuamente para producir imágenes que, a nivel visual, son indistinguibles de aquellas presentes en conjuntos de datos reales.

2.2.1. GAN

Una GAN es un tipo de arquitectura de red neuronal utilizada en el campo del aprendizaje profundo para generación de datos sintéticos. Propuesta en 2014 por Goodfellow et al. (2014), representa un cambio en el paradigma de los modelos generativos al utilizar un enfoque adversarial. Esta sección provee una discusión sobre el trasfondo de la mecánica, evolución y variantes de las redes tipo GAN.

Una GAN consta de 2 redes neuronales, generador y discriminador, las cuales son entrenadas simultáneamente a través de un proceso adversarial. El generador

toma como entrada un vector de ruido y genera datos sintéticos, ya sea imágenes, audio, texto. Su objetivo es engañar al discriminador al generar datos sintéticos que imiten la distribución de datos reales, mientras que el discriminador evalúa dada una instancia de datos si es "real" (de un conjunto de datos reales) o "falso" (generado por el generador). La GAN alcanza su equilibrio cuando el generador es tan competente para crear datos falsos que el discriminador ya no puede distinguir entre datos falsos y reales. En este punto, la GAN ha aprendido a generar datos que son indistinguibles de los datos reales.

Formalmente, el discriminador y el generador se involucran en un proceso de minimax con la función de valor V(D,G):

$$\min_G \max_D V(D,G) = \mathop{\mathbb{E}}_{x \sim p_{data}(x)} [\log D(x)] + \mathop{\mathbb{E}}_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

donde x es un dato tomado del conjunto de datos reales p_{data} y z es un dato tomado del conjunto de datos de ruido $p_z(z)$.

El potencial de las GAN en la interpretación de imágenes médicas radica en su capacidad para generar datos que reflejen la variabilidad intrínseca de las patologías. Esto no solo mejora la capacidad de los modelos para generalizar a nuevas situaciones, sino que también facilita el entrenamiento de modelos de clasificación en condiciones clínicas específicas.

Sin embargo, es importante destacar que el desafío en el entrenamiento de GAN radica en encontrar un equilibrio óptimo entre el generador y el discriminador. El éxito de estas redes depende de su capacidad para evolucionar sin caer en estados de colapso, donde la generación de imágenes se vuelve limitada o repetitiva. A continuación, se presentan otros tipos de GAN en los cuales nos basaremos para crear el método propuesto.

2.2.2. cGAN

Una cGAN es una variante de las redes tipo GAN propuesta por Isola et al. (2017). La cGAN introduce el concepto de generación condicional, que permite la generación de muestras condicionadas a información de entrada. Al igual que la GAN se compone de 2 redes neuronales. Sin embargo, a diferencia de una GAN regular, el generador también recibe información adicional, una condición. Esta condición puede ser de cualquier forma, como etiquetas de clase, descripciones en texto, imágenes, o combinaciones de datos. La condición sirve como guía del generador para generar imágenes que concuerde con las propiedades deseadas o características especificadas.

Las GAN son modelos generativos que aprenden un mapeo de un vector de ruido aleatorio z a una imagen de salida y, $G : z \rightarrow y$, en contraste, las GAN condicionales aprenden un mapeo de una imagen observada x y un vector de ruido aleatorio z, a y, $G : \{x, z\} \rightarrow y$. La definición formal de este nuevo juego minimax con respecto a la condición empleada toma la forma de:

$$V(x, y) = \arg \min_G \max_D \mathbb{E}_y[\log D(y)] + \mathbb{E}_{x,z}[\log(1 - D(G(x, z)))]$$

2.2.3. WGAN

La WGAN (Wasserstein GAN, Arjovsky et al. (2017)) introduce una modificación en la forma en la que las GAN son entrenadas, enfocándose en la estabilidad y aborda problemas comunes asociados con el entrenamiento estándar de una GAN. Esta sección explicará los fundamentos conceptuales detrás de una WGAN y por qué representa un desarrollo significativo en el campo de los modelos generativos.

La idea principal detrás de una WGAN es reemplazar la pérdida tradicional de una GAN por la distancia Wasserstein, también conocida como *Earth Mover*. Esta métrica de distancia mide el costo mínimo de transportar una masa para transformar una distribución (el conjunto de datos sintéticos) en otra (el conjunto de datos real).

Matemáticamente, se puede formular como:

$$W(\mathbb{P}_r, \mathbb{P}_g) = \inf_{\lambda \in \prod(\mathbb{P}_r, \mathbb{P}_g)} \mathop{\mathbb{E}}_{(x,y)\sim\gamma} [\|x - y\|]$$

donde p y q son los datos y distribuciones del modelo, respectivamente, y $\prod(p, q)$ representa el conjunto de todas las posibles distribuciones conjuntas $\lambda(x, y)$ cuyos marginales son p y q. La distancia Wasserstein genera un gradiente más suave y relevante en todas partes.

Penalidad del gradiente

En la implementación inicial de la WGAN, el enfoque para cumplir con la restricción de 1-Lipschitz, necesaria para calcular la distancia Wasserstein, consistía en utilizar el recorte de pesos. Sin embargo, esta técnica presentaba diversas complicaciones, como el subuso de la capacidad de modelado y la potencial explosión de gradientes. Estos desafíos condujeron al desarrollo del método de penalidad de gradiente, propuesto como un mecanismo alternativo y más eficiente para garantizar el cumplimiento de la mencionada restricción 1-Lipschitz.

La restricción de 1-Lipschitz es una condición matemática que garantiza que la función discriminadora sea K-Lipschitz continua, es decir, que su gradiente esté acotado por un valor constante K, en el caso de 1-Lipschitz, que sea K=1. Formalmente, una función f es K-Lipschitz si para todos los puntos x_1 y x_2 en su dominio, se cumple que:

$$|f(x_1) - f(x_2)| \leq K \cdot ||x_1 - x_2||$$

Esta restricción es crucial en la WGAN porque asegura que la función discriminadora no varíe abruptamente, lo que es necesario para que la distancia Wasserstein sea un valor válido. El método de penalidad de gradiente, introducido posteriormente, impone esta restricción de una manera más suave y efectiva, penalizando las desviaciones del gradiente de la función discriminadora respecto a su valor ideal, en lugar

de forzar un recorte rígido de los pesos que tenía el método WGAN sin penalidad. La función de pérdida modificada con la penalidad de gradiente para el discriminador está dada por:

$$L_d = \underbrace{\mathop{\mathbb{E}}_{\tilde{x}\sim\mathbb{P}_g}[D(\tilde{x})] - \mathop{\mathbb{E}}_{x\sim\mathbb{P}_r}[D(x)]}_{\text{Pérdida original de la WGAN}} + \underbrace{\lambda \mathop{\mathbb{E}}_{\hat{x}\sim\mathbb{P}_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2]}_{\text{Penalidad del gradiente}}$$

Donde $\hat{x}$ son puntos de muestreo uniforme a través de líneas rectas entre pares de puntos muestreados de la distribución de datos $\mathbb{P}_r$ y la distribución generadora $\mathbb{P}_g$. λ es el coeficiente de penalidad.

Para el generador la pérdida toma la forma de:

$$L_{adv} = -\mathop{\mathbb{E}}_{\tilde{x}\sim\mathbb{P}_g}[D(\tilde{x})]$$

Wasserstein GAN con Penalidad de gradiente representa un avance robusto en los métodos GAN, abordando problemas clave en la GAN original y ofrece un enfoque más estable para generar imágenes sintéticas de alta calidad. La base teórica es esencial para el entendimiento en su aplicación para la síntesis de imágenes médicas, particularmente para tareas complejas y sensibles, tales como el aumento de imágenes médicas retinales.

2.3. Transferencia de estilo

La transferencia de estilo es una técnica derivada del campo de la visión por computadora, donde elementos de estilo de una imagen son transferidos al contenido de otra. Inicialmente, fue popularizada en aplicaciones artísticas, como transformar fotografías para que se parezcan a pinturas de artistas famosos, la transferencia de estilo tiene un potencial significativo en imágenes médicas. Esta sección pretende profundizar en los mecanismos de la transferencia de estilo, en su adaptación para

aplicaciones médicas y la integración en la WGAN-GP para mejorar conjuntos de datos de imágenes retinales.

La transferencia de estilo tiene como objetivo separar y recombinar el contenido de estilo de imágenes. En el contexto de las redes neuronales, esto es típicamente obtenido usando una Red Neuronal Convolucional (CNN, *Convolutional Neural Network*) que ha sido preentrenada en un conjunto de datos grande. La red aprende a representar contenido en capas superiores mientras captura el estilo en las capas más profundas. Generalmente este proceso involucra:

- Representación de contenido: El contenido de una imagen, tales como la estructura anatómica en una imagen retinal, es capturado en las capas profundas de una CNN, las cuales tienden a codificar características de más alto nivel.
- Representación de estilo: El estilo, definido por las texturas y colores, es capturado por las correlaciones entre características de las capas más superficiales, típicamente se usan matrices de Gram para medir estas correlaciones.

En este proceso, tres tipos de funciones de pérdida tienen un rol crucial: pérdida de contenido, pérdida de estilo y pérdida de variación total. Cada una tiene un propósito específico en guiar la red neuronal hacia generar una imagen de salida que visualmente se parezca al estilo de una imagen mientras conserva el contenido de otra:

- Pérdida de contenido: Esta función de pérdida se asegura de que los contenidos de las características de alto nivel de la imagen objetivo sean preservados en la imagen sintética. Esto es típicamente calculando las diferencias entre los mapas de características de la representación de contenido de una red neuronal preentrenada en capas profundas, las cuales capturan las características más abstractas de una imagen. El objetivo es minimizar las diferencias entre las representaciones de características de la imagen generada y la imagen de

contenido, asegurando que los elementos esenciales y la estructura sean mantenidos. Matemáticamente, se tiene, la diferencia entre una imagen real y una imagen sintetizada medida en el espacio características de una red perceptual, para una capa específica λ y una función de extracción de características F_V^λ, define una función de pérdida como:

$$L_{contenido} = \sum_{i=1}^{N} \left\| F_V^\lambda(x_i) - F_V^\lambda(\hat{x}_i) \right\|$$

- Pérdida de estilo: Se enfoca en capturar y transferir el estilo (texturas y patrones visuales) de la imagen de estilo de referencia a la imagen generada. La función de pérdida es comúnmente calculada utilizando la matriz de Gram, la cual mide las correlaciones entre diferentes mapas de características en las capas superiores de una CNN. Al minimizar la distancia Manhattan entre las matrices Gram de la referencia de estilo y la imagen generada, la red neuronal ajusta la imagen generada a los patrones de estilo y textura de la referencia.

$$L_{estilo} = \sum_{j=1}^{M} ||G_j(x_i) - G_j(\hat{x}_i)||^2$$

- Pérdida de variación total: Esta pérdida es usada para promover la suavidad espacial en la imagen generada, reducir el ruido y alentar elementos visuales coherentes. Trabaja al penalizar las irregularidades en las intensidades de pixel entre pixeles vecinos, así suavizando las transiciones y mejorando la calidad visual general de la imagen. Formalmente:

$$L_{tv} = \sum_{\omega,h} \left[\left\| \hat{x}_i^{(\omega,h+1)} - \hat{x}_i^{(\omega,h)} \right\| + \left\| \hat{x}_i^{(\omega+1,h)} - \hat{x}_i^{(\omega,h)} \right\| \right]$$

En total, la pérdida para el generador tiene la forma de:

$$L_{generador} = L_{adv} + L_{contenido} + L_{estilo} + L_{tv}$$

2.4. Extracción de lesiones

La detección y extracción de lesiones en imágenes retinales son fundamentales para el diagnóstico y monitoreo de enfermedades como la retinopatía diabética. La identificación precisa y la caracterización de lesiones, tales como microaneurismas, hemorragias y exudados, son cruciales debido a su importancia en el diagnóstico clínico. La integración de la extracción de lesiones en modelos generativos, como una WGAN-GP, ofrece una línea de investigación prometedora para crear conjuntos de datos mejorados para el entrenamiento de algoritmos de *machine learning*.

2.4.1. Técnicas tradicionales y avanzadas

Los métodos tradicionales para extracción de lesiones involucran técnicas como umbralización para destacar características específicas, operaciones morfológicas para enfatizar la forma, y varios métodos de filtrados para mejorar el contraste en las imágenes, y con esto hacer las lesiones más detectables.

Las técnicas más avanzadas utilizan métodos *deep learning*, particularmente redes convolucionales, que se han utilizado por su habilidad de aprender patrones complejos en los datos. Estos métodos superan significativamente las técnicas tradicionales de visión en precisión y son más robustos.

Específicamente, en Niu et al. (2022) utilizan un método innovador en donde usan descriptores patológicos que codifican la información basada en las lesiones derivadas de activaciones neuronales con una red que detecta retinopatía diabética. Estos descriptores incluyen coordenadas espaciales y categorías, permitiendo el control preciso sobre las lesiones en las imágenes generadas.

2.5. Algoritmos de redimensionamiento de imágenes

El redimensionamiento de imágenes es una operación crucial en el procesamiento de imágenes, especialmente cuando se preparan datos para modelos de aprendizaje automático. Diferentes algoritmos de redimensionamiento ofrecen diversos balances entre calidad de imagen y eficiencia computacional. A continuación, se describen los algoritmos de redimensionamiento utilizados en los experimentos.

2.5.1. *Nearest Neighbor* (Vecino Más Cercano)

El algoritmo de Vecino Más Cercano (o *Nearest Neighbor*) es uno de los métodos más simples para el redimensionamiento de imágenes. Este algoritmo selecciona el píxel más cercano en la imagen original para asignar su valor al píxel correspondiente en la imagen redimensionada.

- **Ventajas**:
  - **Simplicidad**: Es fácil de implementar y rápido de calcular.
  - **Velocidad**: Muy eficiente en términos computacionales.
- **Limitaciones**:
  - **Calidad de Imagen**: Puede producir imágenes de baja calidad con artefactos visibles, especialmente cuando se reduce el tamaño de la imagen. Se pueden notar bordes escalonados y una falta de suavidad.

2.5.2. *Bilinear Interpolation* (Interpolación Bilineal)

La interpolación bilineal toma un enfoque más avanzado que el método de Vecino Más Cercano, considerando los cuatro píxeles más cercanos en la imagen

original y calculando un valor promedio ponderado para el nuevo píxel.

- **Ventajas**:
  - **Calidad Mejorada**: Produce imágenes más suaves y con menos artefactos visibles comparado con el Vecino Más Cercano.
  - **Suavidad**: Los bordes son más suaves y menos propensos a escalonarse.
- **Limitaciones**:
  - **Eficiencia**: Es más costoso computacionalmente que el método de Vecino Más Cercano.

2.5.3. *Bicubic Interpolation* (Interpolación Bicúbica)

La interpolación bicúbica va un paso más allá que la bilineal, utilizando 16 píxeles (un bloque de 4x4) para calcular el valor de cada nuevo píxel, proporcionando un resultado aún más suave.

- **Ventajas**:
  - **Alta Calidad**: Ofrece una calidad de imagen superior, con transiciones más suaves y menos artefactos.
  - **Suavidad**: Los detalles finos y los bordes se representan mejor que con la interpolación bilineal.
- **Limitaciones**:
  - **Eficiencia Computacional**: Es más lento que los métodos de Vecino Más Cercano y Bilineal debido a su complejidad adicional.

2.5.4. *Lanczos Resampling* (remuestreo de Lanczos)

El remuestreo de Lanczos es un método avanzado que utiliza una función seno cardinal para calcular los valores de los nuevos píxeles. Considera un área mayor alrededor del píxel original (normalmente una ventana de 3x3 o 4x4 píxeles).

- **Ventajas**:
  - **Muy Alta Calidad**: Produce imágenes con muy alta calidad, preservando detalles finos y bordes nítidos.
  - **Fidelidad**: Ideal para aplicaciones donde la calidad de la imagen es crítica.
- **Limitaciones**:
  - **Computacionalmente Intenso**: Es uno de los métodos más costosos en términos de tiempo de cálculo.
  - **Artefactos**: Puede introducir artefactos de *ringing* (anillos) alrededor de los bordes debido a la función seno cardinal.

2.5.5. Mitchell-Netravali

El algoritmo Mitchell-Netravali, también conocido como Mitchell-*Bicubic*, es un método de interpolación bicúbica que utiliza una fórmula específica diseñada para producir imágenes suaves y de alta calidad, evitando algunos de los artefactos comunes en otros métodos bicúbicos.

- **Ventajas**:
  - **Equilibrio**: Ofrece un buen equilibrio entre suavidad y nitidez, produciendo imágenes de alta calidad.

    - **Menos Artefactos**: Minimiza los artefactos como el *ringing* y el *aliasing* que pueden ser comunes en otros métodos de interpolación.

- **Limitaciones**:

    - **Eficiencia Computacional**: Más lento que el método de Vecino Más Cercano y la interpolación bilineal, aunque comparable con la bicúbica estándar.

La elección del algoritmo de redimensionamiento depende del equilibrio deseado entre la calidad de imagen y la eficiencia computacional. En aplicaciones donde la velocidad es crítica, como en el procesamiento en tiempo real, el método de Vecino Más Cercano puede ser adecuado. Por otro lado, en aplicaciones donde la calidad de imagen es primordial, como en la visualización médica o la fotografía profesional, se prefieren los métodos de bicúbica o lanczos, a pesar de su mayor costo computacional. En este trabajo, se comparan empíricamente todos los métodos anteriormente descritos en la sección de experimentos para determinar el mejor algoritmo para nuestro caso.

2.6. Métricas de evaluación

La evaluación eficaz de imágenes sintéticas involucra una combinación de métricas cualitativas y cuantitativas. Estas métricas ayudan a validar el realismo, la precisión y la aplicabilidad de las imágenes generadas, asegurando que sean adecuadas para entrenar modelos de *machine learning* y asistir en aplicaciones específicas, como en nuestro caso, el diagnóstico médico. Esta sección discute varias métricas de evaluación comúnmente utilizadas para medir la calidad de las imágenes sintéticas, enfocándose en la FID, SSIM y MSE, así como en las metodologías para análisis cualitativo y cuantitativo.

2.6.1. Métricas cuantitativas

Fretchet Inception Distance

Las métricas cuantitativas buscan medir objetivamente la calidad de las imágenes generadas mediante criterios numéricos. Entre ellas, destaca la métrica de FID (Heusel et al. (2017)), que evalúa la similitud estadística entre las distribuciones de las imágenes reales y generadas. Una puntuación baja en FID indica una mayor semejanza y, por lo tanto, una generación de imágenes más realista.

La métrica FID mide la similitud entre dos conjuntos de imágenes, típicamente entre imágenes reales e imágenes sintetizadas por un modelo generativo como la redes tipo GAN. Esta métrica calcula la distancia entre vectores de características extraídas de las capas intermedias de una red Inception preentrenada. Al comparar estas distribuciones de características, FID proporciona una medida del realismo de las imágenes generadas, donde valores más bajos indican que las imágenes generadas están más cerca de las reales en términos de su distribución de características.

El cálculo de FID implica la generación de un gran número de imágenes mediante la GAN y la extracción de características de estas imágenes a través del modelo Inception. Con estas características, se calcula la media y la matriz de covarianza para ambas distribuciones (imágenes reales y generadas). La distancia de Frechet entre estas dos distribuciones se determina mediante la fórmula de distancia de Frechet en el espacio de características.

Una puntuación baja en FID indica que las distribuciones de las imágenes reales y generadas son muy similares en términos de estadísticas de alto orden, lo que sugiere una generación de imágenes de alta calidad. Por otro lado, una puntuación alta indica una mayor disparidad entre las distribuciones, señalando que las imágenes generadas pueden carecer de autenticidad o variabilidad en comparación con las imágenes reales.

Índice de similitud estructural SSIM

La métrica de SSIM (Wang et al. (2004)) es una herramienta utilizada para medir la similitud entre dos imágenes. A diferencia de otras métricas que simplemente evalúan las diferencias de píxel a píxel, el SSIM está diseñado para imitar la percepción humana de la calidad de la imagen. Evalúa cambios en las estructuras importantes de la imagen, teniendo en cuenta factores como la luminancia, el contraste y la estructura.

El cálculo del SSIM se basa en la comparación de tres componentes:

- Luminancia: Evalúa la similitud en el brillo entre dos imágenes. La luminancia se calcula mediante la media de los valores de los píxeles.
- Contraste: Compara la variación de brillo entre dos imágenes. El contraste se mide usando la desviación estándar de los valores de los píxeles.
- Estructura: Mide la similitud en la disposición de los píxeles en dos imágenes. La estructura se evalúa mediante la correlación entre las desviaciones estándar de los valores de los píxeles.

La fórmula del SSIM combina estos tres componentes en una sola métrica:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

donde:

- μ_x y μ_y son las medias de las imágenes x y y, respectivamente.
- σ_x^2 y σ_y^2 son las varianzas de x y y, respectivamente.
- σ_{xy} es la covarianza entre x y y.
- C1 y C2 son constantes para estabilizar la división en caso de denominadores pequeños.

Un valor de SSIM más cercano a 1 significa que las imágenes comparadas son estructuralmente más similares, lo cual es un indicador de alta calidad en las imágenes generadas. Por el contrario, valores más bajos indican diferencias estructurales significativas, sugiriendo que las imágenes generadas no son fieles a las originales.

Error cuadratico medio (MSE)

La métrica de MSE es una medida ampliamente utilizada para evaluar la calidad de las imágenes generadas. MSE calcula la media de los cuadrados de las diferencias entre los valores de píxel correspondientes en las imágenes generadas y las imágenes reales. Esta métrica proporciona una indicación cuantitativa de cuán diferentes son las imágenes generadas de las imágenes originales. El MSE se define matemáticamente como:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2$$

donde:

- N es el número total de pixeles en la imagen.
- x_i representa el valor del pixel en la posición i de la imagen original.
- y_i representa el valor del pixel en la misma posición i de la imagen generada.

El MSE proporciona una única puntuación numérica que indica la media de los errores al cuadrado entre los valores de los píxeles correspondientes en las dos imágenes. Un valor de MSE más bajo indica que las imágenes generadas están más cerca de las imágenes reales, lo que sugiere una mejor calidad de las imágenes generadas. Por el contrario, un valor de MSE más alto indica mayores diferencias entre las imágenes generadas y las reales, lo que sugiere una menor calidad de las imágenes generadas.

2.6.2. Métrica cualitativa

La evaluación cualitativa juega un papel fundamental en la validación de la utilidad clínica y el realismo de las imágenes sintéticas. Esta evaluación se centra en aspectos perceptuales de la calidad de la imagen, como la nitidez, la coherencia visual y la presencia de detalles realistas. La efectividad y la aplicabilidad de las imágenes generadas se determinan mediante el juicio de expertos en el área, quienes examinan meticulosamente estas características para asegurar que las imágenes sintéticas cumplan con los estándares requeridos para su uso práctico.

El análisis realizado por expertos es crucial en la evaluación de imágenes sintéticas para determinar su realismo y precisión. Los especialistas revisan cuidadosamente estas imágenes para confirmar que representan con exactitud las características observadas en imágenes reales. Por ejemplo, en el caso de las imágenes de la retina, las características clave incluyen los vasos sanguíneos, lesiones y el disco óptico. Además, los expertos evalúan la relevancia diagnóstica de las imágenes para asegurar que contengan los detalles patológicos necesarios para un diagnóstico preciso y que no omitan ni distorsionen características esenciales de la imagen.

2.7. Prueba estadística

En el análisis de datos, las pruebas estadísticas juegan un papel crucial para validar la significancia de los resultados obtenidos, especialmente cuando se comparan dos o más grupos. Las pruebas estadísticas permiten determinar si las diferencias observadas en los datos son atribuibles al azar o si reflejan diferencias genuinas entre los grupos en estudio. En el contexto de la generación de imágenes médicas y los modelos de *deep learning*, las pruebas estadísticas son esenciales para evaluar el rendimiento de diferentes métodos, comparando métricas como la precisión, MSE, y SSIM.

Una de las pruebas estadísticas más utilizadas para comparar dos grupos dependientes es la Prueba de Wilcoxon para rangos con signo, también conocido como la Prueba de Wilcoxon. Este método es no paramétrico, lo que significa que no asume que los datos sigan una distribución normal, lo que lo hace ideal para aplicaciones donde los datos no satisfacen los supuestos de normalidad.

2.7.1. Prueba de Wilcoxon

La prueba de Wilcoxon fue propuesta por Frank Wilcoxon en 1945 y se utiliza comúnmente para evaluar si hay diferencias significativas entre dos muestras relacionadas, o entre dos mediciones repetidas de un mismo conjunto de sujetos. A diferencia de pruebas paramétricas como el t-test, la prueba de Wilcoxon compara los rangos de las diferencias en lugar de los valores numéricos directamente, lo que le otorga robustez ante la presencia de valores atípicos y distribuciones no normales. El procedimiento de la prueba de Wilcoxon consiste en lo siguiente:

- Se calculan las diferencias entre las observaciones emparejadas.
- Se ordenan las diferencias absolutas y se les asignan rangos.
- Se asignan signos positivos o negativos a los rangos según el signo de la diferencia.
- Se suman los rangos positivos y negativos, y se calcula la estadística de Wilcoxon.

La estadística resultante se compara con un valor crítico (*p-value*) para determinar si la diferencia es significativa. El test es ampliamente utilizado en estudios donde se desea evaluar el rendimiento de algoritmos antes y después de aplicar una mejora, o al comparar dos modelos bajo las mismas condiciones.

En el presente trabajo, el método de Wilcoxon fue utilizado para comparar las métricas de evaluación de calidad de imágenes generadas por diferentes configuraciones del método propuesto. Debido a que las métricas como FID, MSE y SSIM no necesariamente siguen una distribución normal, la prueba de Wilcoxon proporcionó un marco adecuado para evaluar si las diferencias observadas entre los métodos eran estadísticamente significativas.

CAPÍTULO 3

TRABAJO RELACIONADO

Este capítulo abordará estudios relevantes que han aplicado redes tipo GAN para mejorar, generar o manipular imágenes médicas. Se destacarán investigaciones que demuestran el uso de redes tipo GAN para aumentar conjuntos de datos médicos, mejorar la resolución de imágenes o generar imágenes médicas sintéticas para entrenamiento.

3.1. GAN para aumento de imágenes médicas

En el campo de la medicina, las redes tipo GAN han emergido como una herramienta poderosa para el aumento y mejora de conjuntos de datos de imágenes médicas, facilitando avances significativos en la clasificación y diagnóstico de enfermedades. Un ejemplo notable es el estudio de Frid-Adar et al. (2018), que implementó una Red Convolucional Generativa Adversarial Profunda (DCGAN,*Deep Convolutional Generative Adversarial Network*) para enriquecer una base de datos de lesiones hepáticas. Originalmente compuesta por tres clases con un total de 183 imágenes, esta técnica demostró ser especialmente efectiva para clasificar tales lesiones, evidenciando cómo las GAN pueden contribuir significativamente a la precisión diagnóstica en la medicina.

En otro enfoque, los autores en Wu et al. (2020) aplicaron una GAN contextual para la síntesis de lesiones en imágenes de mamografías inicialmente sanas. Este método permitió la inserción precisa de lesiones, tales como masas y calcificaciones, en imágenes sin alteraciones patológicas previas, equilibrando así bases de datos desbalanceadas y mejorando la clasificación de cáncer de mama. Este enfoque destaca la capacidad de las GAN para generar datos sintéticos que reflejan diversas manifestaciones de enfermedades, permitiendo entrenamientos específicos para distintos tipos de lesiones y mejorando notablemente la capacidad de los sistemas de clasificación.

Por otro lado, en Lee et al. (2019) emplearon una GAN condicionada con etiquetas BIRADS para crear imágenes que incorporan diferentes tipos de lesiones basadas en condiciones específicas. A diferencia de las técnicas previas, este estudio utilizó descripciones embebidas para especificar el tipo de BIRADS deseado, lo que permitió generar imágenes ajustadas a características particulares de las lesiones. Utilizando una red de embebido[1] para las descripciones, un generador basado en U-net (Ronneberger et al. (2015)) y un discriminador CNN, se logró producir un conjunto de 1088 imágenes de alta especificidad, demostrando la versatilidad y precisión de las GAN en la generación de imágenes médicas sintéticas.

Además, en Toda et al. (2021) aprovecharon una InfoGAN para expandir una base de datos de imágenes de tomografías por computadora (CT, *Computer Tomography*) de cáncer de pulmón. A partir de solo 66 pacientes, y mediante un preprocesamiento que incluyó la selección de cortes transversales de imágenes tridimensionales, obtuvieron 7644 imágenes. Este enriquecimiento del conjunto de datos llevó a una mejora del 20 % en la clasificación de imágenes, subrayando la importancia

[1] Una red de embebido es una técnica de aprendizaje automático que transforma datos categóricos o textuales en vectores numéricos de menor dimensión, preservando las relaciones semánticas entre los datos. En este contexto, las descripciones de las etiquetas BIRADS fueron convertidas en representaciones vectoriales a través de una red de embebido, permitiendo que la GAN generara imágenes médicas sintéticas con un alto grado de especificidad y detalle.

de las GAN para la generación de nuevas imágenes y para el fortalecimiento de los modelos de clasificación.

En el estudio realizado por Kossen et al. (2021), se llevó a cabo una evaluación de la capacidad de tres redes tipo GAN para sintetizar y segmentar imágenes de resonancia magnética del cerebro, enfocándose específicamente en la segmentación de vasos cerebrales. Este análisis se realizó utilizando una base de datos compuesta por imágenes de 121 pacientes que habían sido diagnosticados con diversas afecciones cerebrovasculares. Estos pacientes formaban parte de dos estudios distintos, lo que proporcionó una variedad significativa de datos para el entrenamiento y evaluación de los modelos de GAN propuestos. Las tres arquitecturas de GAN evaluadas en el estudio fueron DCGAN, WGAN-GP y WGAN-GP con Normalización Espectral (WGAN-GP-SN). Estas arquitecturas fueron seleccionadas por su potencial para generar imágenes sintéticas de alta calidad y por su capacidad para ser entrenadas de manera eficiente en conjuntos de datos médicos complejos. Para cuantificar la calidad de las imágenes sintéticas generadas por cada modelo de GAN, los investigadores emplearon la FID como métrica principal. De acuerdo con los resultados obtenidos, la arquitectura WGAN-GP-SN demostró ser la más efectiva, logrando un valor de FID de 37.01. Este resultado indica que las imágenes generadas por la WGAN-GP-SN mantienen una gran fidelidad visual con respecto al conjunto de imágenes reales, destacándose sobre las otras arquitecturas evaluadas. La efectividad de la WGAN-GP-SN se atribuye a la incorporación de la normalización espectral, que contribuye a estabilizar el proceso de entrenamiento de la red y a mejorar la calidad de las imágenes generadas.

El trabajo realizado por Wang et al. (2019) emplea una técnica basada en WGAN para abordar el problema de desbalance de datos en la clasificación de imágenes de tomografías computarizadas de nódulos pulmonares. WGAN se utiliza específicamente para sintetizar muestras en clases minoritarias, mejorando así el equilibrio en el conjunto de datos para la clasificación fina de características se-

mánticas de los nódulos pulmonares. Para la tarea de clasificación, se utilizó una CNN convencional como clasificador. La eficacia de la técnica basada en WGAN se comparó no solo con métodos tradicionales de aumento de datos, sino también con otras variantes de GAN, como la GAN original y DCGAN. El conjunto de datos empleado en este estudio incluye 2632 tomografías de pulmones, anotados por radiólogos con calificaciones en 9 atributos semánticos. Sin embargo, para este estudio específico, solo se consideraron 7 de estos atributos debido a la distribución desequilibrada en algunos de ellos. Este desbalance pronunciado en la distribución de las clases ocasionó que el conjunto de datos fuera ideal para evaluar la propuesta de WGAN. Para evaluar la eficacia de sus experimentos, los autores utilizaron tres métricas principales: F1-score, G-mean extendido y Distancia Absoluta. Estas métricas fueron seleccionadas para ofrecer una evaluación de la clasificación de imágenes, enfocándose tanto en la precisión como en la sensibilidad del modelo frente a clases desbalanceadas. Los resultados obtenidos demostraron que, en comparación con los métodos tradicionales y otras variantes de GAN, la técnica basada en WGAN presentó una mejora significativa en la clasificación de clases minoritarias, evidenciando su capacidad para generar muestras sintéticas más representativas y útiles para el entrenamiento de modelos de clasificación en contextos de desbalance de clases.

3.1.1. Discusión

Estos trabajos ilustran la capacidad transformadora de las GAN en el ámbito médico, desde el aumento de bases de datos hasta la mejora en la clasificación de enfermedades. La aplicación de GAN, particularmente las versiones condicionadas y contextualizadas, abre nuevas vías para la investigación médica, permitiendo una simulación detallada y realista de patologías que pueden ser cruciales para el diagnóstico oportuno y la formación médica, a continuación veremos más ejemplos pero utilizando imágenes retinales.

3.2. GAN para imágenes retinales

Las redes tipo GAN han sido utilizadas para segmentar y aumentar datos con imágenes de retina con diferentes propuestas. Un hito significativo en este campo fue introducido por Zhao et al. (2018), quienes innovaron con la primera aplicación de la transferencia de estilo mediante redes tipo GAN condicionadas por segmentación de imágenes retinales. Este enfoque pionero mejora la segmentación de datos y enriquece el conjunto de datos con imágenes sintéticas que replican fielmente el contenido y el estilo de imágenes reales. Utilizando una red VGG16, lograron extraer características distintivas de contenido y estilo de las imágenes, integrando luego esta información para calcular una pérdida diferencial que se añade a la función de pérdida del generador. Este generador, basado en la arquitectura U-net (Ronneberger et al. (2015)) , incorpora un vector de ruido en sus capas intermedias (o "cuello de botella"), lo cual permite la generación de imágenes variadas a partir de una misma anotación de segmentación.

Este enfoque fue validado a través de la aplicación en cuatro conjuntos de datos significativos en el campo: STARE con 20 imágenes (Hoover et al. (2000)), DRIVE con 30 imágenes (Staal et al. (2004)), HRF con 45 imágenes (Köhler et al. (2013)) y NeuB1 (De et al. (2016)), demostrando su eficacia con una métrica de Similitud Estructural (SSIM) de 0.8980. Esta métrica refleja la capacidad del modelo para generar imágenes que sean visualmente similares a las reales, y que también mantengan la coherencia estructural necesaria para aplicaciones médicas.

Los autores en Iqbal and Ali (2018) utilizan una pix2pix (Isola et al. (2017)) modificada para generar imágenes retinales e imágenes segmentadas, también hacen uso de la transferencia de estilo para mejorar los colores y la imagen en general generada. Con 2 entrenamientos por separado para cada objetivo. A pesar de la limitación en la cantidad de datos, los autores demostraron que es posible entrenar redes neuronales profundas para producir resultados de alta calidad, esto gracias a

su enfoque con transferencia de estilo. Enfatizan que un buen *framework* de discriminador es clave para entrenar exitosamente una red GAN. Para evaluar la efectividad de su método, utilizaron métricas estándar, como el F1-score y el Área Bajo la Curva (AUC, *Area Under the Curve*), que proporcionan una medida cuantitativa de la precisión de la segmentación y la calidad general de las imágenes generadas.

Una interesante propuesta por Zhao et al. (2019) utiliza una R-sGAN, la cual es una GAN con características similares a una red recurrente, y con una idea similar a la de una LSTM convolucional. Para segmentar imágenes retinales con mayor precisión, y generar imágenes sintéticas realistas, se emplea la transferencia de estilo para mejorar sus imágenes y agregan una nueva función de pérdida llamada Pérdida de Diversidad, en ella se calcula el producto punto de la diferencia entre vectores de ruido e imágenes. Con esta pérdida esperan obtener resultados más diversos que con otras propuestas. Utilizan 4 bases de datos con pocas imágenes, DRIVE, STARE, IOSTAR (Abbasi-Sureshjani et al. (2015)) y HRF. IOSTAR contiene 24 imágenes de 1024x1024 pixeles.

Los autores de Niu et al. (2022) crearon una red explicable para la detección de DR y generación de imágenes de retina, llamada PathoGAN para extraer la información de la enfermedad de las imágenes, y así poder generar esa misma enfermedad en otra imagen sintética. La forma en la que introduce la condición se explica a continuación. La estrategia consiste en tomar una imagen en la cual se tiene una enfermedad (en este caso retinopatía diabética) y genera un descriptor patológico con ayuda de una red entrenada para detectar retinopatía, utiliza las activaciones de la red para saber en donde se encuentran las lesiones. Así mismo utilizan transferencia de estilo con una VGG16 para ayudar al generador que a crear imágenes más realistas. Utilizan 3 bases de datos de retinopatía diabética, IDRiD (Porwal et al. (2018)), *Retinal-Lesions* (Wei et al. (2020)), FGADR (Zhou et al. (2021)), cada una con distinto tamaño. Utilizaron como evaluación la métrica FID y MSE, logran 20.34, 22.28, 80.13 en FID respetivamente y 0.0086, 0.0149 y 0.0107 en MSE,

respectivamente.

En el estudio realizado por Zhou et al. (2022), se explora el uso de una GAN condicional para la síntesis de lesiones de retinopatía diabética en imágenes de la retina. Esta GAN condicional se distingue por su capacidad para integrar múltiples condiciones en su proceso de generación, tales como una máscara de lesiones, el disco óptico y la segmentación retinal. Este enfoque se enriquece con el uso de módulos de atención de espacio multiescala y de canales, junto con un sistema encoder-decoder de dos etapas diseñadas para generar imágenes de alta resolución (1280x1280). Además, se incorpora un módulo que permite ajustar el grado de severidad de las lesiones generadas, ofreciendo un control fino sobre el resultado final. El generador, basado en la arquitectura U-Net, procesa estas entradas para producir imágenes sintéticas de la retina. La investigación se centra en dos bases de datos principales: EyePACS (Emma Dugas (2015)) y FGADR. Inicialmente, se utiliza un modelo previamente entrenado con FGADR para segmentar las imágenes de EyePACS, y posteriormente, esta base de datos se emplea para el entrenamiento. EyePACS, siendo la base de datos pública más grande de retinopatía diabética, consta de 35,126 imágenes de entrenamiento y 53,576 imágenes de prueba, cada una etiquetada solo con el grado de retinopatía. Para evaluar la efectividad de su GAN, los autores adoptan un enfoque cualitativo, presentando imágenes sintéticas a tres expertos y solicitando su evaluación en una escala de 1 a 10 para determinar la fidelidad de las imágenes en términos de venas, texturas y colores. En promedio, el método alcanzó una puntuación de 7.97 en fidelidad y un 65.8 % en precisión, según la percepción de los expertos sobre si una imagen es real o sintética. Adicionalmente, se realizó otro experimento para determinar si los expertos podían evaluar el nivel de retinopatía diabética en las imágenes, obteniendo un 85.3 % de éxito. El estudio también incluyó una evaluación mediante la métrica FID, logrando un promedio de 4.24 a través de las cinco clases evaluadas. Finalmente, se buscó balancear la base de datos de EyePACS para verificar la capacidad de generalización de la GAN. Se experimentó con

varios clasificadores neuronales y distintas maneras de combinar imágenes sintéticas con reales. Utilizando una arquitectura VGG16, el mejor resultado fue una precisión del 87.72 % utilizando imágenes reales para entrenamiento y sintéticas para pruebas. Resultados similares se obtuvieron con ResNet-50 (89.45 %), InceptionV3 (88.37 %), Zoom-in (89.66 %) y AFN (90.46 %).

3.2.1. Discusión

Estos estudios resaltan el creciente interés en el campo de la segmentación y generación de imágenes de la retina, cada uno de ellos aplicando técnicas de transferencia de estilo para mejorar la síntesis de las imágenes. La identificación y manipulación de lesiones representa también un ámbito significativo de investigación, donde cada investigador adopta enfoques distintos. El trabajo presentado en esta tesis se distingue principalmente por su enfoque en la extracción de lesiones utilizando técnicas aplicadas por Niu et al. (2022) y abordar un enfoque WGAN combinando técnicas como transferencia de estilo con la pérdida Wasserstein, para mejorar la estabilidad del entrenamiento y conseguir una mejor síntesis de imagen. La integración de estos elementos contribuye a superar algunos de los desafíos que se discuten en la literatura, como lo son la disponibilidad de imágenes, y la necesidad de una precisa representación de patologías.

A continuación se presenta la Tabla 3.1, en la cual se compara el proyecto realizado con diversos trabajos relevantes en la generación de imágenes médicas.

Autor	Método Utilizado	Diferencia con el Método Propuesto
Frid-Adar et al. (2018)	DCGAN para enriquecer la base de datos de lesiones hepáticas	Utiliza DCGAN en lugar de WGAN-GP; enfocado en lesiones hepáticas, no retinales. No incorpora técnicas avanzadas de extracción de lesiones ni transferencia de estilo.
Wu et al. (2020)	GAN contextual para la síntesis de lesiones en imágenes de mamografías	Inserta lesiones en imágenes sanas; enfoque en mamografías, no en imágenes retinales. No emplea WGAN-GP ni técnicas de transferencia de estilo.
Lee et al. (2019)	GAN condicionada con etiquetas BIRADS	Usa descripciones embebidas y etiquetas BIRADS; enfoque en imágenes mamográficas. No utiliza WGAN-GP ni métodos de extracción de lesiones avanzados.
Toda et al. (2021)	InfoGAN para expandir una base de datos de imágenes de CT de cáncer de pulmón	Utiliza InfoGAN; enfoque en imágenes de CT y cáncer de pulmón, no en imágenes retinales. Carece de técnicas específicas para la extracción y transferencia de estilo de lesiones.
Kossen et al. (2021)	DCGAN, WGAN-GP y WGAN-GP-SN para sintetizar y segmentar imágenes de resonancia magnética del cerebro	Comparación de diferentes arquitecturas de GAN; enfoque en segmentación cerebral. No se centra en la generación de imágenes retinales ni en la transferencia de estilo.
Zhao et al. (2018)	GAN condicionada por segmentación retinal con transferencia de estilo	Utiliza VGG19 para extraer características de contenido y estilo; enfoque en segmentación retinal. No implementa WGAN-GP ni técnicas de extracción de lesiones avanzadas.
Iqbal y Ali (2018)	Pix2pix modificada para generar imágenes retinales y segmentadas	Uso de transferencia de estilo y entrenamientos separados; enfoque en imágenes retinales. No emplea WGAN-GP ni incorpora técnicas avanzadas de extracción de lesiones.
Zhao et al. (2019)	R-sGAN con características de red recurrente y pérdida de diversidad	Utiliza transferencia de estilo y pérdida de diversidad; enfoque en imágenes retinales. No utiliza WGAN-GP y puede ser menos efectivo en la generación de imágenes de alta calidad.
Niu et al. (2022)	PathoGAN para detección de DR y generación de imágenes de retina	Utiliza descripciones patológicas y transferencia de estilo con VGG19; enfoque en retinopatía diabética. No implementa WGAN-GP y puede tener limitaciones en la generación de imágenes sintéticas diversas.
Zhou et al. (2022)	GAN condicional para la síntesis de lesiones de retinopatía diabética en imágenes de la retina	Integra múltiples condiciones en el proceso de generación, tales como una máscara de lesiones, el disco óptico y la segmentación retinal. Utiliza módulos de atención multiescala. No emplea técnicas avanzadas de extracción de lesiones ni WGAN-GP.

Tabla 3.1: Comparación de métodos para la generación de imágenes médicas

CAPÍTULO 4

MÉTODO PROPUESTO

Como se mencionó en la revisión de la literatura, hay una necesidad considerable de bases de datos equilibradas y de imágenes sintéticas de alta calidad. El método propuesto utiliza una WGAN con penalización de gradiente debido a su alta estabilidad durante el entrenamiento, tal como se describe en el trabajo de Gulrajani et al. (2017). Además, se emplea la transferencia de estilo, una técnica ampliamente utilizada para imágenes de retina, que ha demostrado obtener buenos resultados.

El uso de la transferencia de estilo junto con la pérdida de Wasserstein es un enfoque novedoso en la literatura para la generación de imágenes de retina. Con esta combinación, se espera mejorar los resultados existentes y generalizarla a otros ámbitos. En las siguientes secciones se describirán los métodos que se utilizarán y, al final, se explicará cómo se integran todos estos elementos para crear el método propuesto.

4.1. WGAN-GP

La elección de la WGAN fue un paso importante en el desarrollo de esta investigación, ya que existen varias formas de plantear una GAN para generar imágenes. La WGAN tienen una estabilidad destacable durante su entrenamiento, lo cual es

crucial al trabajar con imágenes médicas como las retinas, donde la precisión y la calidad de las imágenes generadas son fundamentales. Además, la penalidad de gradiente propuesta por Gulrajani et al. (2017) ofrece una solución efectiva a los problemas comunes de las GAN, como el modo colapso, permitiendo generar variedad en las imágenes sintéticas sin perder coherencia ni características clínicas relevantes.

La WGAN-GP tiene como principal diferencia a una GAN normal, el uso de la pérdida Wasserstein y la penalidad del gradiente, primero veamos cómo está diseñada la arquitectura utilizada, después se detallará como se utilizan las pérdidas.

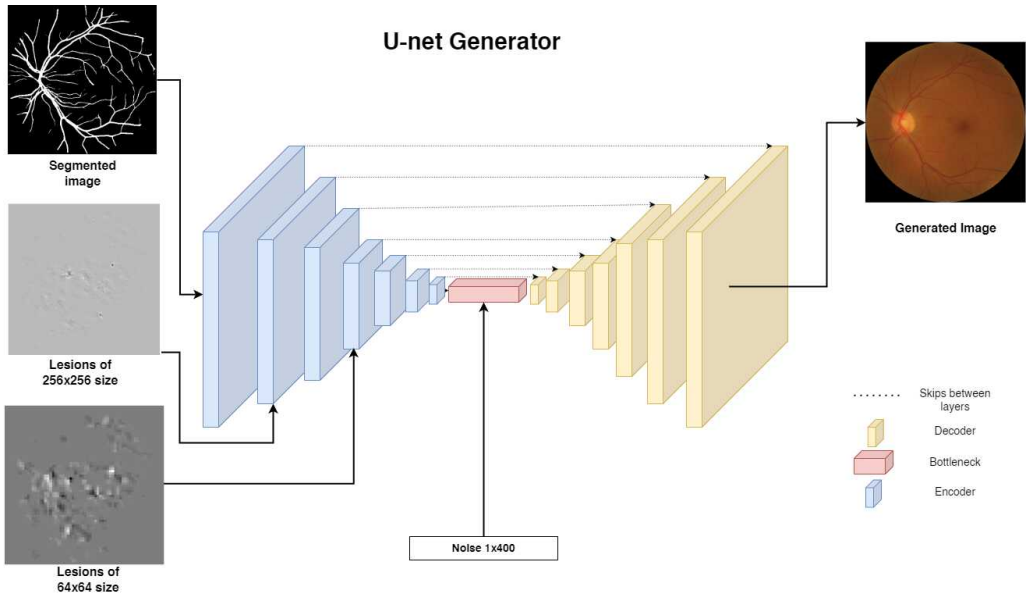


Figura 4.1: Esquema del Generador del método propuesto. Cada bloque *encoder* incluye una capa convolucional, normalización del *batch* y *Leaky* ReLU. Cada bloque *decoder* consta de una capa de redimensionamiento, una capa convolucional, una capa de normalización del *batch*, un *dropout* condicional y una activación ReLU. Fuente: Elaboración propia.

La WGAN-GP en la Figura 4.1 presenta un generador con una arquitectura U-Net que consta de 7 capas de codificación (*encoder*) y 7 capas de decodificación (*decoder*). Cada capa de codificación incluye una capa convolucional 2D con *strides* de (2,2), un tamaño de *kernel* de (4,4) y diferentes tamaños de filtro: 64 para la

primera capa, 128 para la segunda, 256 para la tercera y 512 para las últimas cuatro. Además, el *encoder* contiene una capa de normalización del *batch* (común en este tipo de redes) que se utiliza para proporcionar estabilidad al entrenamiento. Finalmente, se incluye una capa *Leaky* ReLU con un valor de alfa de 0.2, lo que permite números negativos en las activaciones. Esto es importante en las GAN, ya que ayuda a prevenir neuronas muertas, mejorar el gradiente y estabilizar el entrenamiento.

Existen varias formas de diseñar el *decoder*; la más común es utilizar una capa convolucional transpuesta con diferentes números de filtros, un tamaño de *kernel* de (4,4), *strides* de (2,2) y *padding "same"*. Sin embargo, esta configuración tiende a crear artefactos en las imágenes sintéticas, por lo que se optó por utilizar una capa de reescalado, que aumenta el tamaño de la imagen. Se probaron diferentes métodos de reescalado, como el bilineal, bicúbico, Mitchell-*Bicubic* y lanczos 5. Después de esta capa, se añade una capa convolucional 2D con un tamaño de *kernel* de (3,3) y *strides* de (1,1), manteniendo el mismo tamaño de imagen en la entrada y la salida. Esta configuración mejora el entrenamiento y la calidad de las imágenes obtenidas.

Utilizando el método de extracción de lesiones que se detallará más adelante, se insertan en las capas 2 y 4 del generador las características de las lesiones de las imágenes a sintetizar.

El discriminador es una red neuronal convolucional de 6 capas convolucionales. Después de cada capa convolucional, excepto la primera y la última, se incluye una capa de normalización del *batch* y una capa *Leaky* ReLU con un alfa de 0.2. Este discriminador toma como entrada dos imágenes: una imagen segmentada y otra que es la imagen objetivo, que puede ser real o sintética. Su función es aprender a diferenciar entre una imagen sintética y una real, dada una imagen segmentada.

Normalmente, al final del discriminador se utilizaría una capa de activación sigmoide, pero en el caso de la WGAN-GP, no se emplea porque se necesitan los valores reales de esta red para la pérdida de Wasserstein. El método WGAN requiere

que el discriminador se entrene varias veces más con diferentes conjuntos de datos que el generador. En este caso, entrenamos el discriminador cinco veces más que el generador.

4.2. Transferencia de estilo

Aplicamos el método de transferencia de estilo de la siguiente manera (ver Figura 4.2). Utilizamos dos funciones de pérdida para condicionar al generador a utilizar el estilo de la imagen de entrada. La primera, denominada "función de pérdida retinal", calcula el error absoluto medio de las características obtenidas con una VGG-19 (Simonyan and Zisserman (2015)) preentrenada con ImageNet (Russakovsky et al. (2015)), comparando las imágenes reales y las imágenes sintéticas. La segunda, llamada "función de pérdida de severidad", calcula el error cuadrático medio del resultado de una red preentrenada para clasificar las imágenes de retina, nuevamente comparando las imágenes reales y las sintéticas.

Esta red clasifica en cinco clases dependiendo del nivel de lesiones, de cero a cinco, siendo cero sin lesiones o sano, y cinco con una alta cantidad de lesiones. La red fue obtenida de un desafío organizado por *Kaggle* para clasificar imágenes de retinopatía diabética, en el cual ganó el segundo lugar, y los autores la hicieron pública. El desafío de *Kaggle*, denominado *Diabetic Retinopathy Detection*, invitó a los participantes a desarrollar algoritmos que pudieran clasificar el grado de severidad de la retinopatía diabética a partir de imágenes retinales. Los participantes utilizaron diversas técnicas de aprendizaje profundo y procesamiento de imágenes para abordar el problema. La red, que obtuvo el segundo lugar, se destacó por su capacidad para diferenciar con precisión entre los distintos niveles de severidad, proporcionando una herramienta útil para la detección temprana y el manejo de esta condición ocular. Utilizamos esta red para poder encontrar las lesiones de las imágenes, esto nos permitirá extrapolarlas a imágenes nuevas y así generar imágenes sintéticas con

lesiones

Además, la combinación de estas dos funciones de pérdida permite que el generador no solo se enfoque en replicar los patrones y texturas generales de las imágenes de retina, sino también en capturar y reproducir las características específicas que indican la presencia y la gravedad de las lesiones. De esta manera, las imágenes sintéticas se ven realistas y también son útiles para aplicaciones médicas, como la formación de modelos de diagnóstico automático y la investigación sobre la progresión de la retinopatía diabética.

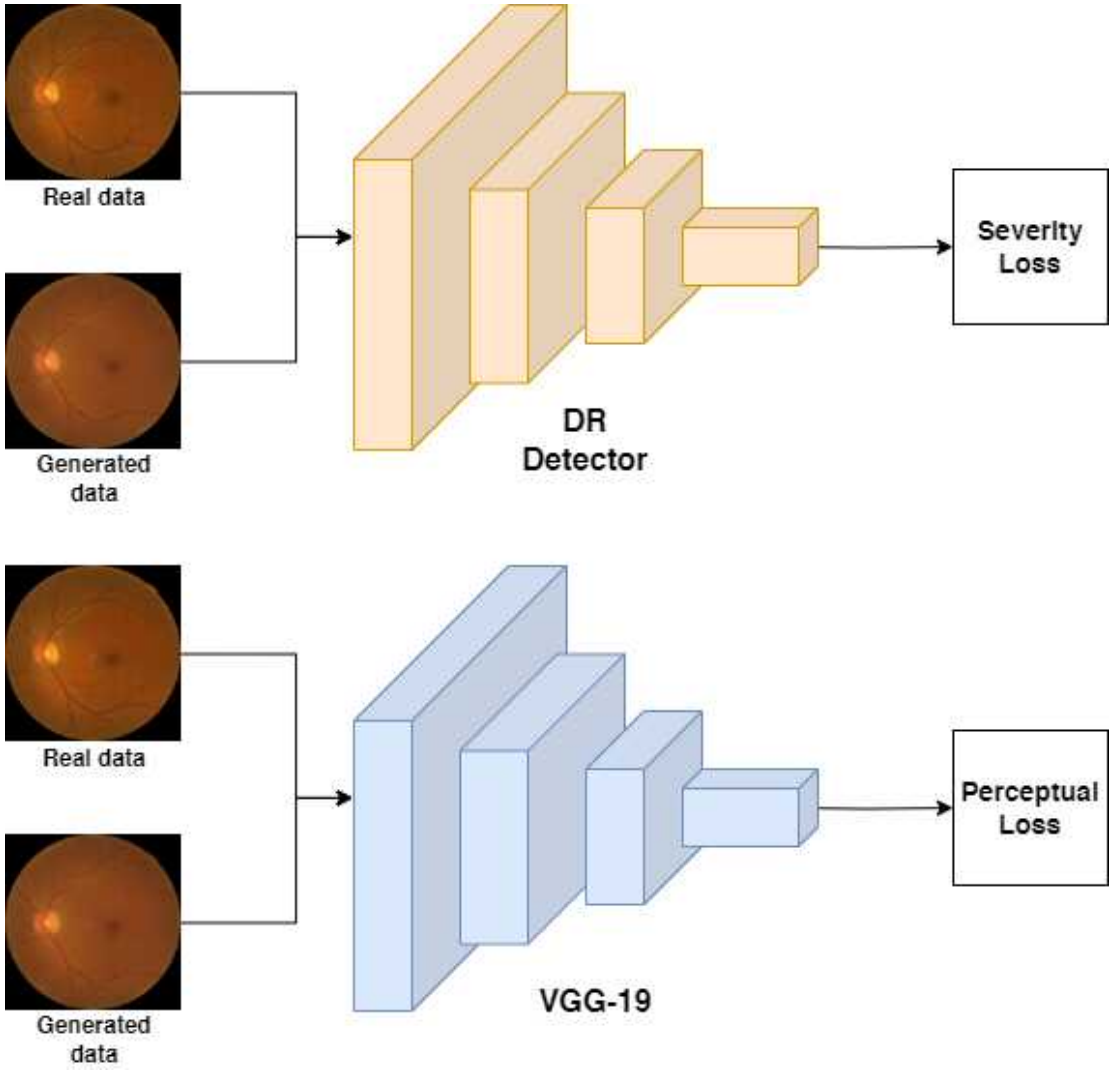


Figura 4.2: Esquema de la transferencia de estilo utilizada. Fuente: Elaboración propia.

4.3. Extracción de características.

Para extraer características relevantes (véase la Figura 4.3), se emplea la red previamente mencionada del desafío de *Kaggle*. Esta red ha demostrado ser eficaz en la clasificación de imágenes de retinopatía diabética, permitiendo así una extracción precisa de características. Para maximizar el rendimiento, hemos generado dos subredes a partir de esta estructura: una que abarca hasta la primera capa convolucional y otra que se extiende hasta la tercera capa convolucional.

El procedimiento para obtener las activaciones es meticuloso. Primero, una imagen se introduce en la red; luego, a partir de la salida de la red, se obtienen los gradientes necesarios para derivar las activaciones. Estas activaciones, que se generan en diferentes tamaños, nos brindan una visión detallada de las características presentes tanto en las capas superficiales como en las profundas de la red. Las capas profundas suelen capturar patrones más complejos y abstractos, mientras que las capas superficiales detectan características más básicas y generales. En particular, estas activaciones son cruciales para identificar las zonas donde se localizan las lesiones características de la retinopatía diabética.

Durante la fase de entrenamiento del modelo, se calcula la presencia de lesiones en cada imagen introducida en la red. Posteriormente, estas características se concatenan en las capas internas del generador, específicamente en la segunda y cuarta capa convolucional. La elección de estas capas no es arbitraria; se debe a que sus dimensiones son adecuadas para permitir una concatenación efectiva con los datos que fluyen a través de la red. Esta integración de datos asegura que la información relevante se mantenga intacta y se utilice de manera óptima en las etapas posteriores del procesamiento.

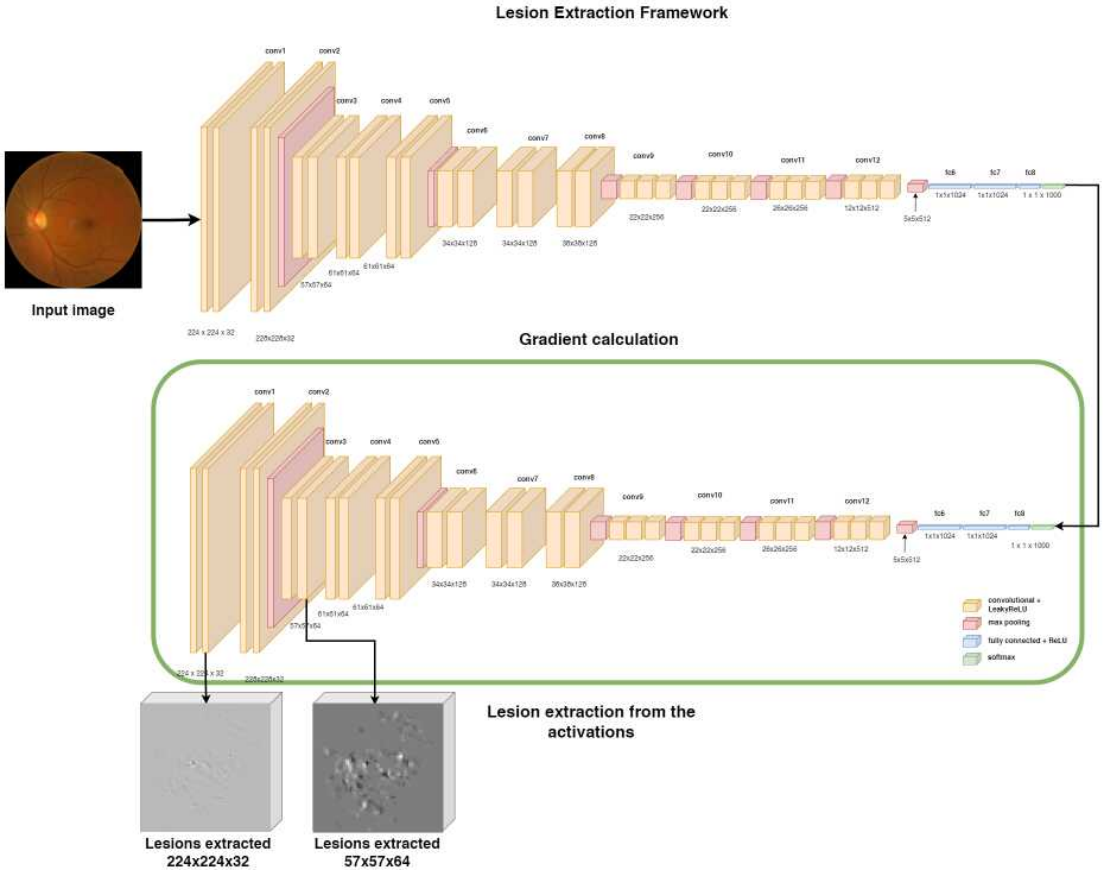


Figura 4.3: Esquema de la extracción de características. Los datos de salida son de diferentes capas en la red de detección de DR con tamaños diferentes, la primera con 224x224x32 y la segunda con 57x57x64. Fuente: Elaboración propia.

4.4. Entrenamiento

El entrenamiento de la red completa se lleva a cabo de la siguiente manera: se seleccionan 5 imágenes reales junto con sus segmentaciones correspondientes. De cada imagen real, se extraen las lesiones. Se utilizan 5 imágenes porque es necesario entrenar al discriminador 5 veces más que al generador, utilizando diferentes datos en cada iteración.

Las lesiones y las imágenes segmentadas se introducen en el generador para producir imágenes sintéticas. Con estas imágenes sintéticas, se calcula la pérdida de

Wasserstein y se aplica la penalidad del gradiente. Este proceso ajusta los pesos del discriminador y se repite 5 veces.

Una vez finalizado este ciclo, se genera una nueva imagen sintética que, junto con la imagen real, se utiliza para calcular las pérdidas de Wasserstein, de severidad y de retina. La suma de estas tres pérdidas se emplea para ajustar los pesos del generador. Este proceso de entrenamiento se repite hasta que se hayan procesado todas las imágenes de la base de datos, completándose después de un determinado número de épocas.

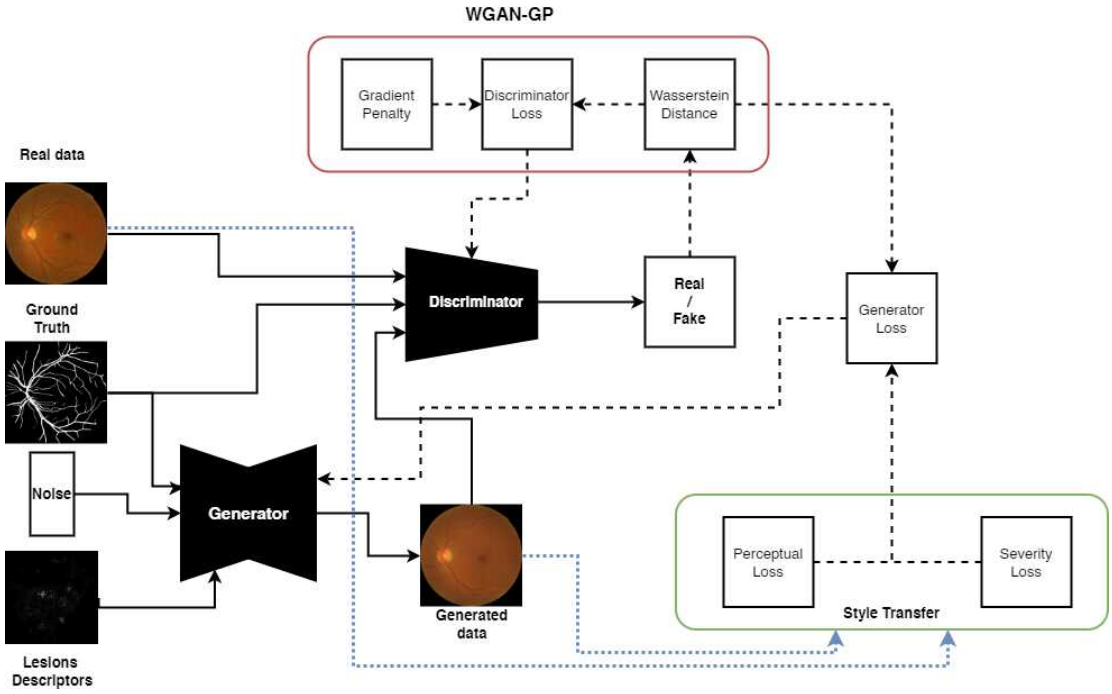


Figura 4.4: Método propuesto y su esquema de entrenamiento, las líneas sólidas indican el flujo de datos, las punteadas el flujo de la pérdida y la azul el flujo de datos para la transferencia de estilo. Fuente: Elaboración propia.

CAPÍTULO 5

EXPERIMENTOS Y RESULTADOS

En esta sección se presentan los experimentos realizados para validar el método propuesto, el cual utiliza una Wasserstein GAN con penalidad de gradiente, combinada con transferencia de estilo y extracción de lesiones, para generar imágenes retinales sintéticas. Estos experimentos fueron diseñados específicamente para evaluar la calidad de las imágenes generadas.

5.1. Bases de datos

Los experimentos conducidos se llevaron a cabo con 5 bases de datos, *Kaggle* (Emma Dugas (2015)), IDRiD (Porwal et al. (2018)), *Retinal-Lesions* (Wei et al. (2020)), FGADR (Zhou et al. (2021)) y RFMiD (Pachade et al. (2020)), descritos en la Tabla 5.1.

5.1.1. Base de datos de *Kaggle*

La base de datos *Kaggle Diabetic Retinopathy Detection* (Emma Dugas (2015)) fue creada para una competencia en *Kaggle* con el objetivo de desarrollar modelos que identifiquen rasgos de retinopatía diabética en imágenes de retina. Esta base

Base de datos	Descripción	Resolución	Número de imágenes
Kaggle	Imágenes clasificadas en 5 categorías	1444x1444, 2184x3456	1379 entrenamiento, 2069 prueba
IDRiD	Imágenes con segmentación en lesiones	4288x2848	54 entrenamiento, 27 prueba
Retinal-Lesions	Imágenes con etiquetado en lesiones y niveles de severidad	896x896	337 entrenamiento, 1256 prueba
FGADR	Imágenes con segmentación de lesiones y niveles de severidad	1280x1280	500 entrenamiento, 1342 prueba
RFMiD	Imágenes con diferentes tipos de enfermedades, etiquetada por enfermedad	2048x1536, 512x512	348 entrenamiento, 174 prueba

Tabla 5.1: Resumen de las bases de datos utilizadas.

de datos fue proporcionada por la plataforma *EyePACS* y patrocinada por la fundación *California Healthcare*. Es una de las bases de datos más utilizadas para la investigación de la retinopatía diabética, con 88,702 imágenes RGB de alta resolución capturadas bajo diversas condiciones.

El uso principal de este conjunto de datos es clasificar la severidad de la retinopatía diabética en cinco categorías: sin Retinopatía Diabética, Leve, Moderado, Severo y Proliferativo Retinopatía Diabética. Nosotros utilizamos este conjunto de datos con solo 700 imágenes de cada clase, seleccionándolas aleatoriamente, sumando un total de 3,500 imágenes, las cuales dividimos en 1,400 para entrenamiento y

2,100 para prueba. Optamos por usar menos imágenes debido al tiempo de entrenamiento requerido; este conjunto específico toma alrededor de 2 días para entrenar 100 épocas. Como se detallará más adelante, se realizaron varios experimentos con diferentes parámetros.

5.1.2. Base de datos IDRiD

La Base de datos de Retinopatia Diabetica de la India (IDRiD, *Indian Diabetic Retinopathy Image Dataset*) (Porwal et al. (2018)) es una base de datos pública creada por el Indian Institute of Technology Delhi. Esta base de datos incluye imágenes retinianas de alta resolución con anotaciones a nivel de píxel para las lesiones. El uso principal de la base de datos IDRiD es facilitar la investigación en la segmentación y clasificación de lesiones de retinopatía diabética.

5.1.3. Base de datos *Retinal-lesions*

Esta base de datos (Wei et al. (2020)) fue creada para proporcionar anotaciones detalladas de lesiones y niveles de severidad en imágenes retinianas. Es comúnmente utilizada para entrenar modelos que detecten y clasifiquen diversos tipos de lesiones asociadas con la retinopatía diabética. Contiene 1,842 imágenes seleccionadas de la base de datos de *Kaggle*, las cuales han sido reetiquetadas por un panel de 45 oftalmólogos expertos en los 5 niveles de retinopatía diabética y en ocho clases de lesiones.

5.1.4. Base de datos FGADR

La Base de datos de Retinopatia Diabetica con Anotaciones de Grano-Fino (FGADR, *Fine-Grained Annotated Diabetic Retinopathy*) fue desarrollada por Zhou

et al. (2021) con el objetivo de proporcionar recursos extensivos para el desarrollo y evaluación de modelos de aprendizaje automático en el campo de la retinopatía diabética. Esta base de datos contiene 2,842 imágenes, divididas en dos subconjuntos: 1,842 imágenes con anotaciones a nivel de píxel de las lesiones y 1,000 imágenes con etiquetas del grado de retinopatía diabética, evaluadas por 6 oftalmólogos. Dado que requeríamos una mayor cantidad de lesiones para verificar la capacidad de transferir estas lesiones a las imágenes sintéticas, utilizamos únicamente las 1,842 imágenes con anotaciones. Sin embargo, no empleamos las anotaciones en nuestro proceso.

5.1.5. Base de datos RFMiD

La Base de datos de Imágenes de Multi-enfermedad Retinales (RFMID, *Retinal Fundus Multi-Disease Image Dataset*, Pachade et al. (2020)) es un conjunto de datos exhaustivo que contiene imágenes de retina capturadas para el diagnóstico de múltiples enfermedades oculares. Esta base de datos está diseñada para abordar diversos problemas de salud visual, incluyendo, pero no limitándose a, la retinopatía diabética, la degeneración macular relacionada con la edad, el glaucoma y otras patologías de retina. RFMiD es especialmente valiosa debido a la variedad y complejidad de las imágenes que alberga, cada una de las cuales está etiquetada con precisión según el tipo de enfermedad presente. Las imágenes de fondo de ojo en RFMiD no solo proporcionan una representación visual de las patologías oculares, sino que también incluyen detalles clínicos esenciales, como la presencia de microaneurismas, exudados duros, hemorragias y neovascularización. Este nivel de detalle es crucial para entrenar y evaluar modelos de diagnóstico automatizado que buscan mejorar la precisión en la detección y clasificación de enfermedades oculares. El contexto de trabajo de RFMiD se centra en el diagnóstico asistido por inteligencia artificial, proporcionando un recurso robusto para investigadores y desarrolladores en la construcción de modelos capaces de identificar múltiples enfermedades desde

una única imagen de fondo de ojo.

5.2. Preprocesamiento

Dado que cada base de datos tiene diferentes tamaños de imágenes, se tuvieron que escalar a 512x512 píxeles para mantener un tamaño uniforme sin perder demasiada información y para no agotar los recursos computacionales. Una vez reescaladas, se extrajeron sus máscaras y se utilizó una red preentrenada con el dataset DRIVE (Staal et al. (2004)), llamada Spacial Attention U-net (SA-Unet) (Guo et al. (2020)), para segmentar las imágenes retinianas. Con esta red, obtuvimos las segmentaciones de todas las imágenes de las bases de datos.

5.3. Configuración del método propuesto

Se experimentó con diferentes configuraciones. Al principio, se probó con una cGAN con transferencia de estilo, pero a pesar de varias configuraciones, incluyendo diferentes tasas de aprendizaje para el generador y el discriminador, así como distintos pesos para las pérdidas de severidad, contenido y adversarial, el modelo cGAN no conseguía FID menores que el estado del arte, y mostraba mucha inestabilidad. Por esta razón, se cambió a la WGAN-GP con transferencia de estilo y se probaron diferentes configuraciones, modificando los hiperparámetros para observar los cambios en la generación de imágenes. Los hiperparámetros ajustados incluyeron el algoritmo de interpolación utilizado por la capa de redimensionamiento en el generador, que expande la imagen, y los pesos de las funciones de pérdida.

El cambio a la arquitectura WGAN-GP nos trajo mejoras sustanciales en la eficiencia. La configuración inicial de la WGAN-GP con transferencia de estilo mostró resultados prometedores al reducir la métrica del FID en comparación con la cGAN

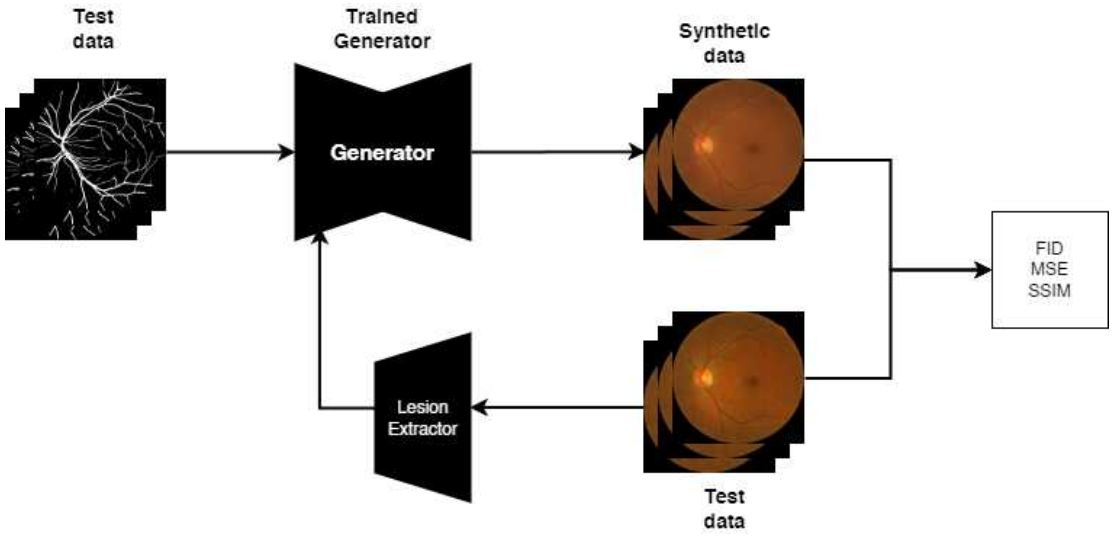


Figura 5.1: Esquema de la evaluación aplicada. Fuente: Elaboración propia.

y el estado del arte. Se probaron varios métodos de interpolación para la capa de redimensionamiento del generador, tales como Mitchell-*Bicubic*, bilineal, bilineal con antialias, cúbico, *nearest neighbor* y lanczos 5. Cada configuración fue entrenada durante 400 épocas.

Para evaluar la red, como se muestra en la Figura 5.1 se toman las segmentaciones del conjunto de prueba, se extraen las lesiones, y generamos un conjunto de imágenes sintéticas del mismo tamaño que el conjunto de prueba. Con estos dos conjuntos de datos, se calcula el FID; el SSIM y MSE, imagen por imagen y promediando sobre todas las imágenes.

5.4. Resultados

En esta sección se presentan los resultados obtenidos utilizando el método propuesto, así como la comparación con diferentes configuraciones y con un método del estado del arte. Además, se muestran los resultados de la capacidad del método para transferir lesiones entre imágenes, y los resultados de una encuesta a varios

expertos en oftalmología sobre las imágenes sintéticas generadas.

5.4.1. Bases de datos

A continuación se describen los resultados de cada base de datos utilizada.

Retinal-lesions

En la Tabla 5.2 se presentan los resultados de los experimentos usando la base de datos de *Retinal-Lesions*. El conjunto de prueba es de 1256 imágenes, se generaron la misma cantidad de imágenes sintéticas con las imágenes segmentadas correspondientes. En la primera columna se muestran los métodos utilizados y sus variaciones, en este caso tenemos el código original de los autores de PathoGAN (Niu et al. (2022)), que también utiliza imágenes de retinopatía, se implementó su método, y se evaluó. Igualmente, se compara con el método que desarrollamos anteriormente, pero que no logro mejorar contra PathoGAN, finalmente ponemos las variaciones del método propuesto. Dado que se trata de redes neuronales y procesos estocásticos, se realizaron 5 repeticiones para evaluar cada método, reportando la desviación estándar correspondiente. Se muestra que la mejor configuración fue la de WGAN-GP con el algoritmo de *nearest neighbor*, para todas las métricas.

En la Figura 5.2 se hace un comparativo entre todas las configuraciones, excepto cuando se usa la cGAN, que se muestra en la Figura 5.3, calculando el FID cada 10 épocas en el entrenamiento. Se aprecia como se va reduciendo la métrica, y como la mejor configuración se mantiene por debajo de las demás. Además, se observa como la cGAN está muy por encima de todas las configuraciones experimentadas.

Se generaron una muestra de imágenes en la Figura 5.4 en donde la imagen original no contiene lesiones. Las imágenes generadas por WGAN-GP (primera fila y dos primeras imágenes de la izquierda en la segunda fila) muestran una síntesis

exitosa del color general y de los vasos sanguíneos de la retina, con un disco óptico bien formado y claramente visible. Mientras que en la imagen generada por cGAN (tercera imagen de la segunda fila), se observan varias zonas donde el color no es adecuado, así como manchas y un disco óptico mal definido. La imagen generada por PathoGAN (cuarta imagen de la segunda fila) logra verse un poco mejor que la cGAN, con una estructura más consistente y un disco óptico mejor definido. Sin embargo, el detalle que tienen las imágenes generadas por la WGAN-GP es que, con el algoritmo *nearest neighbor*, se logra percibir un patrón de colores que no debería de existir, algo común en las imágenes generadas por redes tipo GAN en general.

Retinal-Lesions			
Método	**MSE↓**	**SSIM↑**	**FID ↓**
PathoGAN	0.012045 ± 0.00002	0.8258 ± 0.00004	24.45 ± 0.15
cGAN	0.00902 ± 0.00001	0.8143 ± 0.00003	30.09 ± 0.12
WGAN-GP w/mitchellbicubic	0.00414 ± 0.00001	0.8736 ± 0.00004	21.37 ± 0.03
WGAN-GP w/bicubic	0.00417 ± 0.00001	0.8724 ± 0.00003	20.62 ± 0.06
WGAN-GP w/nearest	$\mathbf{0.00358} \pm 0.00001$	$\mathbf{0.8759} \pm 0.00005$	$\mathbf{15.87} \pm 0.06$
WGAN-GP w/lanczos5	0.00387 ± 0.00001	0.8747 ± 0.00005	20.55 ± 0.05
WGAN-GP w/bilineal	0.00391 ± 0.00001	0.8739 ± 0.00004	18.51 ± 0.03
WGAN-GP w/bilineal+antialias	0.00441 ± 0.00001	0.8717 ± 0.00002	24.56 ± 0.02

Tabla 5.2: Tabla de resultados de los experimentos con la base de datos *Retinal-lesions*. Métodos basados en WGAN-GP son los propuestos y cGAN.

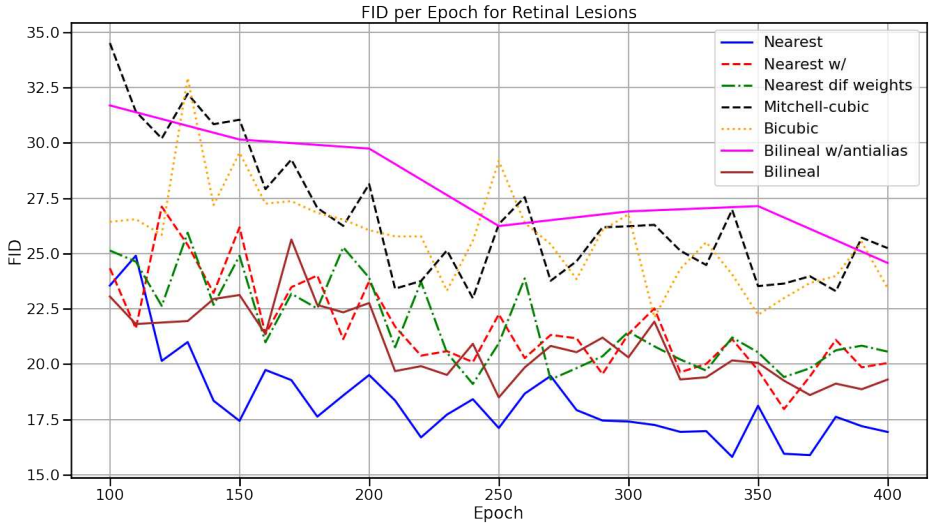


Figura 5.2: Gráfica de la métrica FID en cada configuración por épocas. Se grafican los resultados de los métodos propuestos.

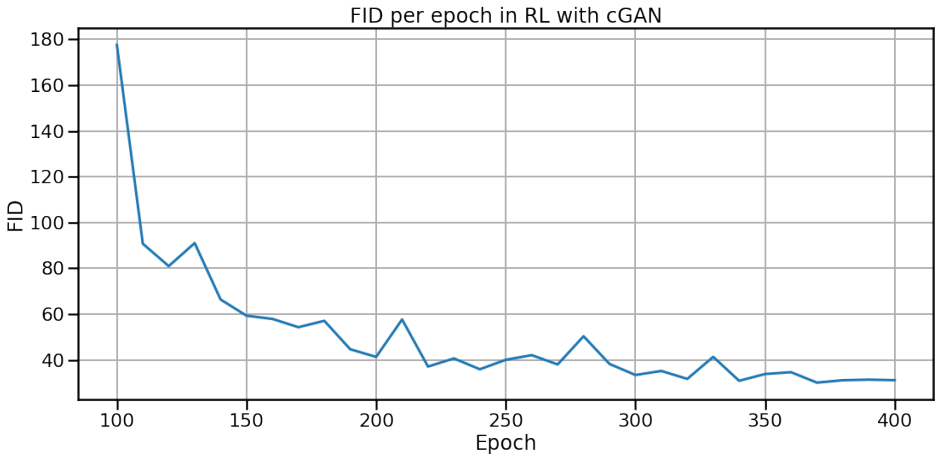


Figura 5.3: Gráfica de la métrica FID en la configuración con cGAN. Comparar con la Figura 5.2

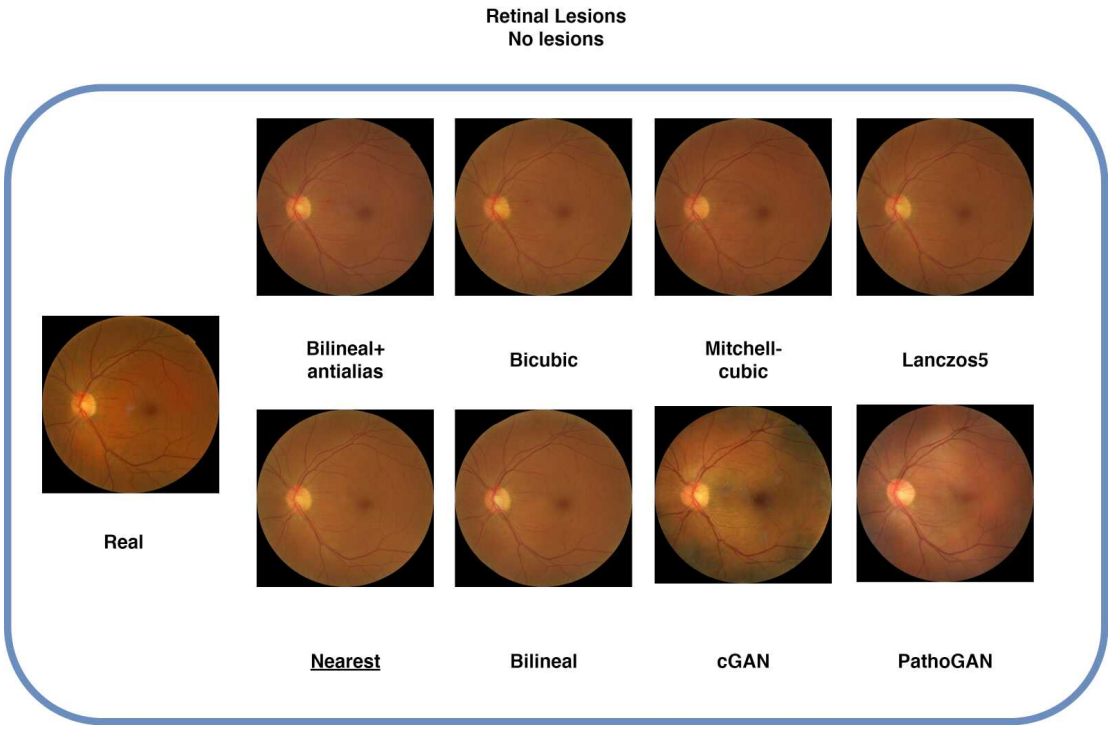


Figura 5.4: Muestra de imágenes de cada configuración, en donde la real no muestra lesiones. La imagen con etiqueta PathoGAN es la implementación de Niu et al. (2022), las demás son utilizando WGAN-GP, pero con diferentes métodos de redimensionamiento, exceptuando cGAN donde cambia la WGAN-GP por una cGAN.

FGADR

En la Tabla 5.3 se muestran las métricas obtenidas con la base de datos FGADR. Esta base de datos presentó desafíos significativos para la implementación de nuestro método, con WGAN-GP obteniendo los resultados en FID de 28.16 con lanczos 5, MSE con 0.544 con *nearest neighbor* y SSIM con 0.793 con bilinear. No hay una configuración que destaque en las tres métricas, debido a la naturaleza de las imágenes, que contienen un tipo de ruido difícil de aprender para una red adversarial, y se puede observar con la SSIM que en todos es menor a 0.6 en comparación con *Retinal-lesions* que obtienen más de 0.8, lo cual implica que las imágenes generadas no son muy parecidas estructuralmente. En la Figura 5.5 se observa que,

cGAN es el mejor entre todas las configuraciones con la métrica FID, esta métrica es de principal interés, ya que fue diseñada para evaluar redes tipo GAN, esto nos dice que este método es mejor para reproducir el ruido que contienen las imágenes. Este ruido presente en las imágenes, se aprecia en la Figura 5.6. A pesar de obtener la mejor métrica de FID, en la sección donde se analizan las lesiones, veremos que no siempre un bajo FID implica que las lesiones se muestren correctamente.

FGADR			
Método	**MSE↓**	**SSIM↑**	**FID ↓**
PathoGAN	0.010445 ± 0.00004	0.5318 ± 0.00004	24.61 ± 0.13
cGAN	0.01106 ± 0.00002	0.5134 ± 0.00005	**21.72** ± 0.10
WGAN-GP w/mitchellbicubic	0.00802 ± 0.00001	0.5692 ± 0.00003	43.05 ± 0.08
WGAN-GP w/bicubic	0.00863 ± 0.00002	0.5649 ± 0.00008	46.31 ± 0.03
WGAN-GP w/nearest	**0.00544** ± 0.00001	0.5692 ± 0.00004	30.61 ± 0.04
WGAN-GP w/lanczos5	0.00582 ± 0.00001	0.5729 ± 0.00006	28.16 ± 0.09
WGAN-GP w/bilineal	0.00569 ± 0.00001	**0.5793** ± 0.00004	28.81 ± 0.08
WGAN-GP w/bilineal+antialias	0.00554 ± 0.00001	0.5779 ± 0.00005	31.10 ± 0.03

Tabla 5.3: Tabla de resultados de los experimentos con la base de datos FGADR. Métodos basados en WGAN-GP y cGAN son los propuestos.

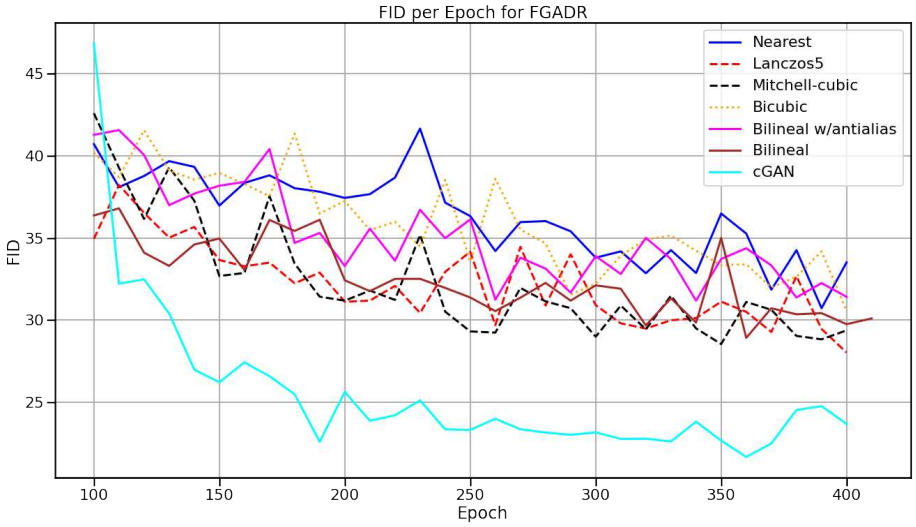


Figura 5.5: Gráfica de la métrica FID en cada configuración por épocas.

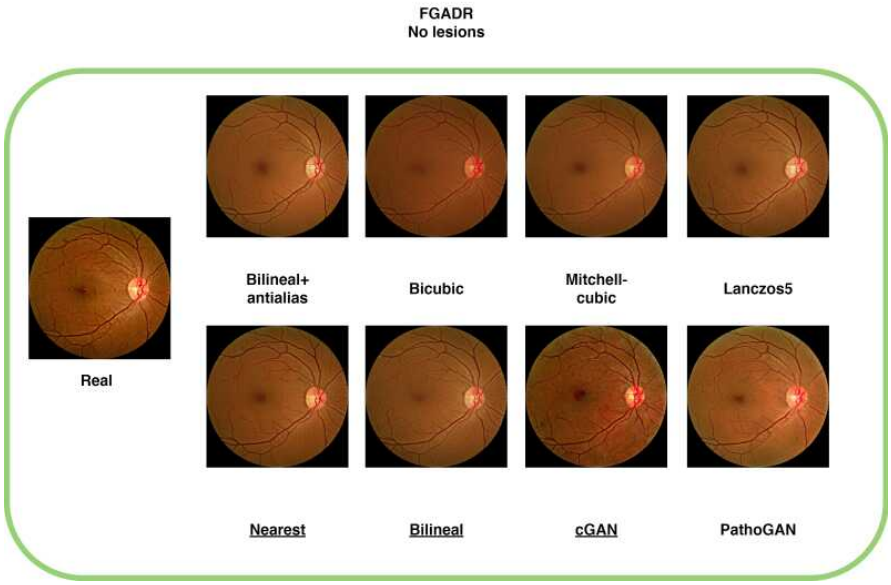


Figura 5.6: Muestra de imágenes de cada configuración, en donde la real no muestra lesiones. Se observa que las imágenes generadas por el método propuesto realiza un suavizado del ruido presente en la imagen original. Por otro lado, la cGAN logra mantener ese ruido, mientras que la PathoGAN también muestra un efecto de suavizado similar.

IDRiD

Para la base de datos IDRiD, la configuración de *nearest neighbor* resultó ser la más óptima, obteniendo un FID de 71.54. La configuración cGAN logró un MSE de 0.00722 y la de Mitchell-*Bicubic* alcanzó un SSIM de 0.6852, como se muestra en la Tabla 5.4. Dado que esta base de datos es muy reducida, se espera que no se capturen todas las características con detalle, resultando en un sobreajuste. Esto se observa en las imágenes de la Figura 5.7, donde se aprecia que las imágenes presentan un buen color y características propias de una imagen retinal, como las venas, el disco óptico y la mácula. En la Figura 5.8, se nota que ningún método se destaca claramente sobre los demás, ya que la mayoría de las configuraciones presentan resultados muy similares.

IDRiD			
Método	**MSE↓**	**SSIM↑**	**FID ↓**
PathoGAN	0.00847 ± 0.0006	0.66024 ± 0.0009	79.88 ± 0.36
cGAN	**0.00722** ± 0.00004	0.6576 ± 0.0004	82.88 ± 1.02
WGAN-GP w/mitchellbicubic	0.00764 ± 0.00002	**0.6852** ± 0.0003	79.99 ± 0.39
WGAN-GP w/bicubic	0.00751 ± 0.00005	0.6762 ± 0.0002	74.43 ± 0.65
WGAN-GP w/nearest	0.00742 ± 0.00006	0.6699 ± 0.0004	**71.54** ± 0.55
WGAN-GP w/lanczos5	0.00727 ± 0.00008	0.6831 ± 0.0005	72.80 ± 0.62
WGAN-GP w/bilineal	0.00729 ± 0.00003	0.6804 ± 0.0005	74.43 ± 0.65
WGAN-GP w/bilineal+antialias	0.00787 ± 0.00007	0.6781 ± 0.0007	72.93 ± 0.70

Tabla 5.4: Tabla de resultados de los experimentos con la base de datos IDRiD. Métodos basados en WGAN-GP y cGAN son los propuestos.

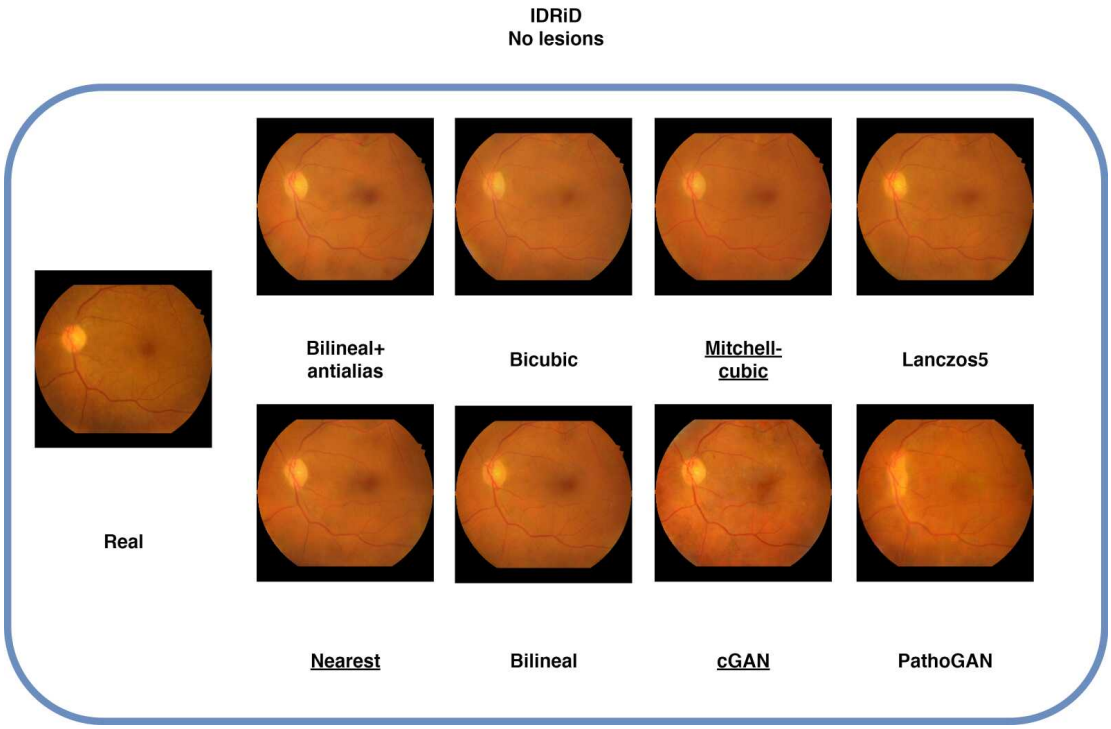


Figura 5.7: Comparación con muestras de imágenes generadas y una real. Las imágenes generadas con el método propuesto presentan colores y texturas que son más similares a la imagen real. En contraste, las imágenes generadas por la cGAN y la PathoGAN muestran variaciones de color en áreas donde la imagen real no las presenta.

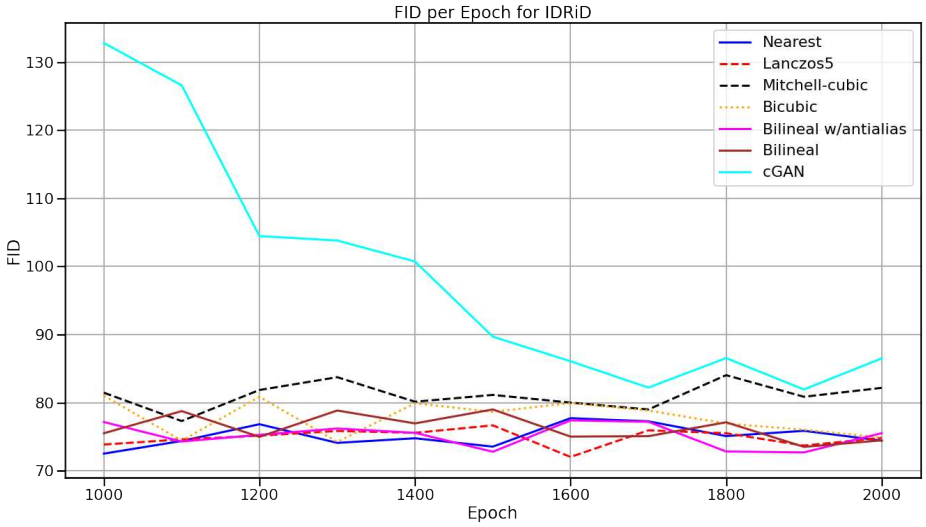


Figura 5.8: Gráfica de la métrica FID en cada configuración por épocas.

Kaggle

En esta base de datos se obtuvo la mejor configuración con FID con el algoritmo *nearest neighbor*, en donde se obtiene un FID de 15.21, y en Bilineal con antialias con MSE de 0.002025, y un SSIM de 0.89. Esta base de datos representó un desafío en términos de tiempo, ya que es bastante grande, cada configuración se llegaba a tardar entre 3 a 9 días, siendo el algoritmo bicúbico el que más tardo con 9 días de entrenamiento por su complejidad. Cabe la pena también mencionar que en la literatura solo existe un artículo que trabaja con cGAN y esta base de datos (Zhou et al. (2022)), pero no se consiguió el código y por esto no se hace la comparación.

En la Figura 5.9 se observa como las configuraciones logran una métrica de FID más baja que las reportadas en las bases de datos anteriormente, se aprecia que las diferentes configuraciones están en constante mejora y no hay un único mejor. Para ilustrar la base de datos se aprecia en la Figura 5.10 una muestra de imágenes obtenidas con cada configuración y como los métodos logran sintetizar correctamente las características fundamentales de una imagen retinal.

Kaggle			
Método	**MSE**	**SSIM**	**FID**
cGAN	0.00902 ± 0.00001	0.8123 ± 0.0006	31.85 ± 0.08
WGAN-GP w/mitchellbicubic	0.00323 ± 0.00006	0.8809 ± 0.0003	$\mathbf{15.21} \pm 0.56$
WGAN-GP w/bicubic	0.00292 ± 0.00008	0.8783 ± 0.0003	16.93 ± 0.68
WGAN-GP w/nearest	0.00350 ± 0.00004	0.8748 ± 0.0005	16.00 ± 0.65
WGAN-GP w/lanczos5	0.00445 ± 0.00008	0.8724 ± 0.0006	19.30 ± 0.55
WGAN-GP w/bilineal	0.00312 ± 0.00005	0.8821 ± 0.0001	16.06 ± 0.45
WGAN-GP w/bilineal+antialias	$\mathbf{0.00275} \pm 0.00006$	$\mathbf{0.8850} \pm 0.0005$	16.85 ± 0.75

Tabla 5.5: Tabla de resultados de los experimentos con la base de datos *Kaggle*

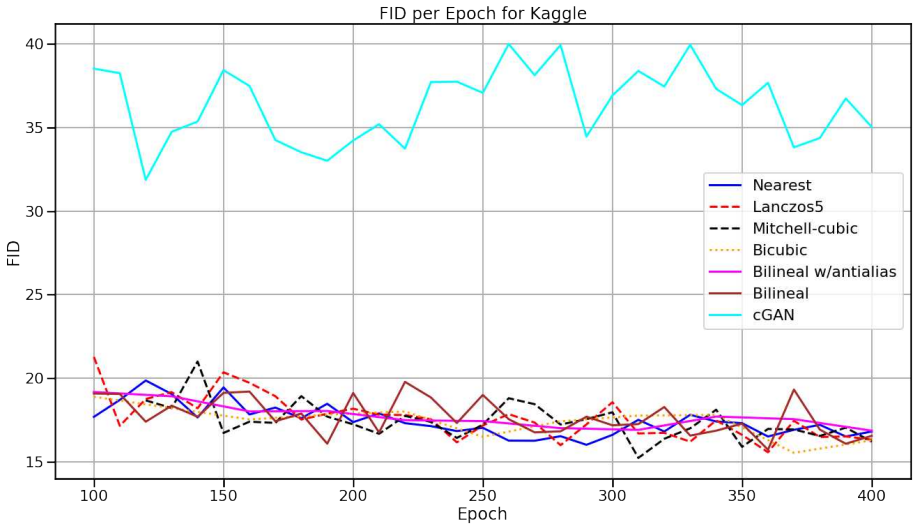


Figura 5.9: Gráfica de la métrica FID en cada configuración por épocas.

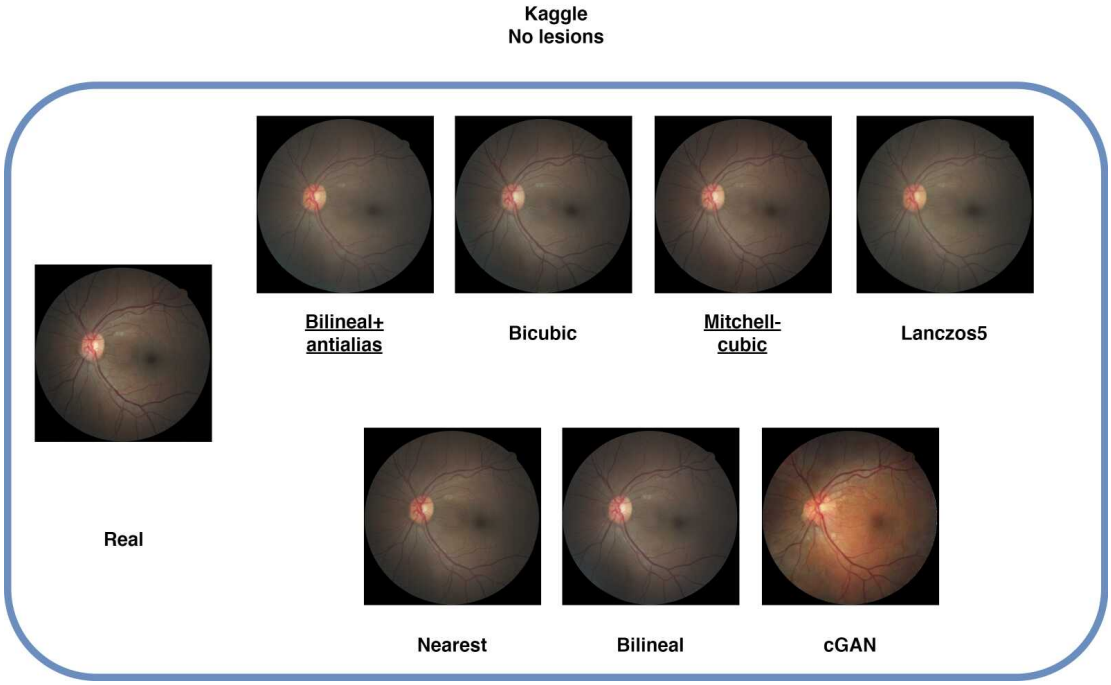


Figura 5.10: Comparación con muestras de imágenes generadas y una real. Se muestra que el método propuesto logra extraer y preservar el color y la textura de la imagen original, mientras que la cGAN presenta tonos diferentes.

RFMiD

También se experimentó con la mejor configuración general de las bases de datos anteriores, que fue el método *nearest neighbor*. Esta base de datos se empleó para evaluar la capacidad del método propuesto de generar lesiones en contextos diferentes a la retinopatía diabética.

En la Figura 5.11 se presentan 3 imágenes reales y 3 imágenes sintéticas de la base de datos, generadas con la configuración implementada. La evaluación se realizó 5 veces, como en las evaluaciones anteriores. Los resultados, mostrados en la Tabla 5.6, indican un valor de la métrica FID de 69.25, comparable a la calidad de las imágenes de IDRiD. Esto sugiere que el método propuesto tiene una calidad aceptable en la generación de imágenes sintéticas. Además, en la Figura 5.12 se observa cómo el entrenamiento mejora gradualmente la métrica FID, indicando que el método sí aprende gradualmente las imágenes. En cuanto a la transferencia de lesiones, la Figura 5.11 muestra que, aunque no fue diseñado específicamente para este contexto, el método logra detectar y transferir adecuadamente las lesiones a las imágenes sintéticas.

RFMiD			
Método	**MSE↓**	**SSIM↑**	**FID ↓**
WGAN-GP w/nearest	0.0044 ± 0.0001	0.82 ± 0.01	69.25 ± 0.02

Tabla 5.6: Tabla de resultados del experimento con la base de datos RFMiD.

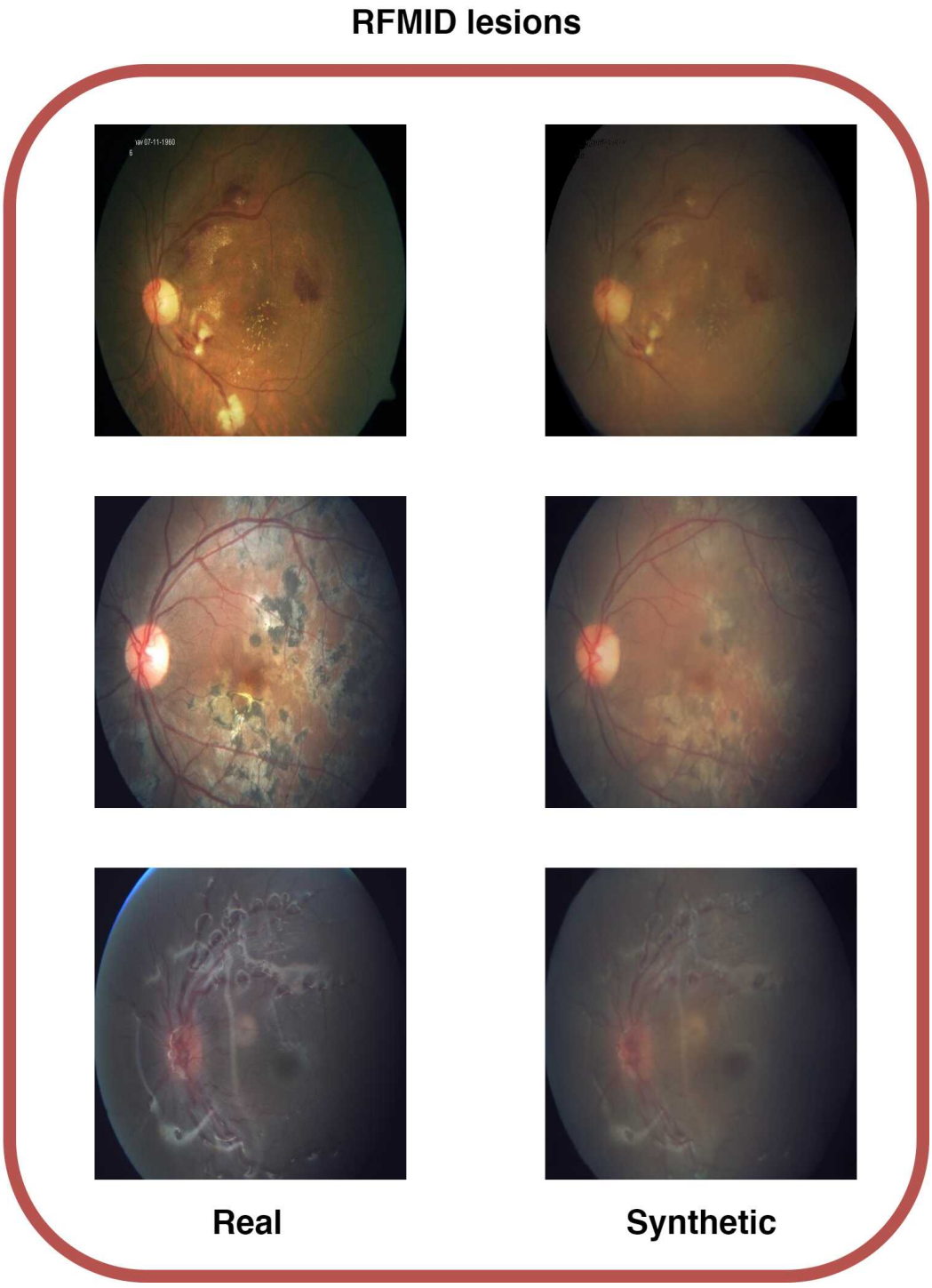


Figura 5.11: Comparación con muestras de imágenes generadas y reales. El método propuesto logra transferir las lesiones de las imágenes originales.

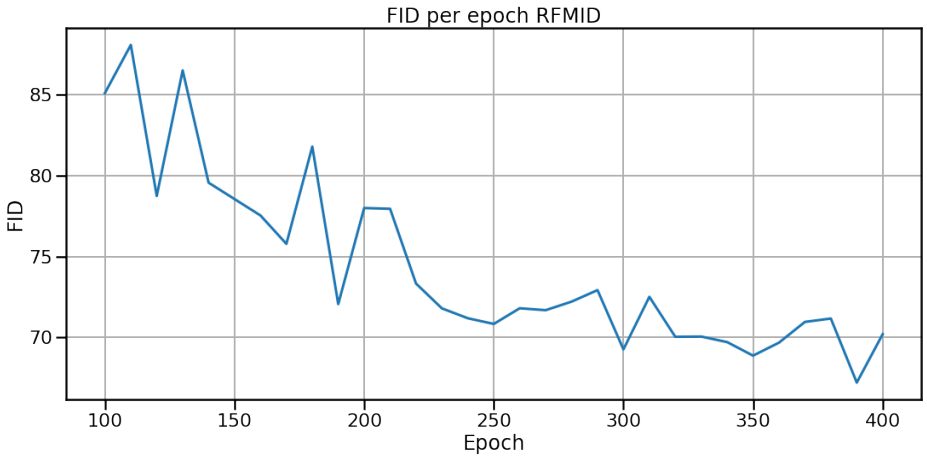


Figura 5.12: Gráfica de la métrica FID por época.

5.4.2. Prueba de Wilcoxon

Se utilizaron pares de resultados generados por la WGAN-GP y la cGAN para evaluar si el método propuesto lograba mejoras significativas en la generación de imágenes sintéticas, comparando las métricas de evaluación entre ambos modelos. La prueba de Wilcoxon permitió verificar que las mejoras en la similitud estructural y la reducción del error de predicción no eran atribuibles al azar, proporcionando evidencia robusta de la superioridad del modelo propuesto. Pero para FID no funcionó debido a que la prueba requiere una cantidad considerable de muestras y solo se tienen 5 pares. En la tabla 5.7 se muestran los resultados de la prueba de Wilcoxon. Se evaluaron los mejores resultados de cada métrica, contra PathoGAN (Niu et al. (2022)) y en *Kaggle* contra cGAN. Para la base de datos FGADR, se observa que todos los *p-value* son menores a 0.05, demostrando que existe una diferencia significativa entre los métodos. Con IDRID aunque hay 2 comparaciones que logran ser menores a 0.05, las demás no lo son, esto puede ser debido a que la evaluación solo usa 27 imágenes, quitándole el poder estadístico a la prueba. Por lo que con IDRID

no es concluyente si es significativa la diferencia. En *Retinal-lesions* los *p-value* son menores a 0.05, con esto concluimos que la diferencia en métodos fue significativa. Finalmente, *Kaggle*, notamos que los *p-value* son de 0, esto puede deberse a que son muchas las cantidades de datos (2096) y el número es tan pequeño que se redondeó a 0, demostrando que los métodos sí tienen una diferencia significativa. Por otro lado, la estadística de Wilcoxon, nos dice cuán consistentes son las diferencias entre dos muestras. En el caso de FGADR a pesar de que en algunos casos el valor es alto, el *p-value* es menor a 0.05, indicando una alta diferencia. Las demás comparaciones tienen estadísticas de Wilcoxon bajas, indicando una diferencia consistente entre las muestras.

5.4.3. Lesiones

También se generaron imágenes con lesiones en las Figuras 5.13, 5.14, donde se muestran las 4 bases de datos, donde la imagen real contiene lesiones fáciles de observar. Se aprecia que el método propuesto captura las lesiones y las sintetiza de manera fiel a la original. Las imágenes generadas por el método, logran mantener la estructura general de la retina, sintetizando las lesiones y características como el color, las venas y el disco óptico sin artefactos ni deformidades. Por otro lado, las imágenes generadas por la cGAN y la PathoGAN no logran transmitir adecuadamente las lesiones, y en general, los colores difieren notablemente del original, con artefactos visibles y un disco óptico menos definido, exceptuando IDRID donde al ser pocas imágenes es fácil sobreajustar el modelo con los mismos tonos. Específicamente, en la Figura 5.13b, a pesar de que cGAN obtuvo la mejor métrica de FID, no logra transferir las lesiones de igual manera que nuestro método mejorado con WGAN-GP. El problema con el método es que las lesiones se notan un poco difuminadas con respecto a la original.

Comparación	MSE (Estadística de Wilcoxon, *p-value*)	SSIM (Estadística de Wilcoxon, *p-value*)
FGADR PathoGAN y WGAN-GP w/nearest	36077.0, 2.55×10^{-187}	22309.0, 7.88×10^{-200}
FGADR PathoGAN y cGAN	358591.0, 9.30×10^{-11}	173708.0, 1.14×10^{-84}
FGADR PathoGAN y WGAN-GP w/Bilineal	63204.0, 7.28×10^{-164}	4133.0, 5.71×10^{-217}
FGADR WGAN-GP w/-Nearest y CGAN	9210.0, 4.13×10^{-212}	301.0, 1.13×10^{-220}
IDRID PathoGAN y WGAN-GP w/mitchellbicubic	166.0, 0.5944	63.0, 0.0017
IDRID PathoGAN vs cGAN	137.0, 0.2198	173.0, 0.7140
IDRID WGAN-GP w/mitchellbicubic vs cGAN	182.0, 0.8779	50.0, 0.0004
Retinal-Lesions PathoGAN vs WGAN-GP w/nearest	7979.0, 9.53×10^{-199}	5446.0, 2.48×10^{-201}
Kaggle WGAN-GP w/bilineal antialias vs cGAN	171.0, 0.0	0.0, 0.0

Tabla 5.7: Resultados de la prueba de Wilcoxon para MSE y SSIM, para las diferentes bases de datos.

5.4.4. Función de pérdida

La pérdida para la mejor configuración que fue WGAN-GP con el algoritmo *nearest neighbor*, se muestra en la Figura 5.15. La gráfica está procesada con el método de media móvil, ya que, al corresponder al discriminador, presenta fluctuaciones prominentes que dificultan la observación detallada del comportamiento de la pérdida. Inicialmente, muestra un valor alto que desciende rápidamente en las primeras épocas, lo cual se debe a que los pesos se inician de manera aleatoria. Por cada época va aprendiendo y, por lo tanto, reduciéndose la pérdida. Después de la fase inicial, la pérdida del discriminador se estabiliza y se mantiene relativamente plano por el resto del entrenamiento, indicando que el discriminador distingue efectivamente entre imágenes reales y sintéticas, y su valor cercano a 0 indica que también está bien balanceado con el generador, previniendo que este se vuelva muy poderoso. Por otro lado, el generador, muestra más fluctuaciones que el discriminador. Se observan picos y caídas en la pérdida, particularmente en las partes medias del entrenamiento. Esta volatilidad se da cuando el generador está intentando mejorar y el discriminador se ajusta conforme a este. Alrededor de la época 150-250 existe un incremento en la pérdida, indicando que al generador le está costando más engañar al generador dentro de esta fase. Después de la época 250 la pérdida cae, indicando una gran mejora en la eficacia del generador.

5.4.5. Evaluación con expertos

Se llevó a cabo una encuesta para evaluar que tan bien pueden las imágenes generadas pasar por imágenes reales. En esta encuesta se incluyeron 50 imágenes seleccionadas aleatoriamente del conjunto de datos *Retinal-lesions* y 50 generadas.

En cada imagen se preguntó si la imagen observada era real o sintética, y además se solicitó a los encuestados que calificaran la imagen mostrada. La encuesta fue aplicada de manera individual a tres expertos en el área de oftalmología, incluyendo dos doctores del área de Ciencias y Tecnologías Biomédicas del INAOE y un oftalmólogo externo con amplia experiencia en la práctica clínica. Cada experto evaluó las imágenes de manera independiente, lo que permitió recoger una diversidad de opiniones basadas en sus respectivos campos de especialización. Las diferencias en las evaluaciones de los expertos podrían atribuirse a sus enfoques únicos y experiencias profesionales, ya que los expertos del INAOE tienen un enfoque más académico y técnico, mientras que el oftalmólogo externo aporta una perspectiva clínica más directa. En la Tabla 5.8 se presenta la precisión de cada experto en distinguir entre imágenes reales y sintéticas. Una menor precisión indica una mayor calidad de las imágenes generadas. También se muestra la fidelidad, que es el promedio de la calificación del 1 al 10 que otorgaron los encuestados a cada imagen sintética. Como muestran los resultados, solo el 56.66 % de la mezcla entre imágenes reales e imágenes generadas por el método WGAN-GP, pudieron ser correctamente identificadas por los expertos. En teoría, la probabilidad de seleccionar una imagen sintética al azar es del 50 %, por lo tanto, haber obtenido este resultado indica que nuestras imágenes sintéticas lograron convencer a los expertos, demostrando así una alta calidad en el color, estructura y textura.

Encuestados	Precisión	Calificación
Experto 1	52 %	8.84
Experto 2	58 %	7.06
Experto 3	60 %	7.84
Promedio	56.66 %	7.91

Tabla 5.8: Tabla de la precisión y calificación dadas por los expertos encuestados.

5.5. Discusión de resultados

En esta sección se analizan en detalle los resultados obtenidos, comparando la eficacia del método propuesto con otras técnicas existentes, y discutiendo las implicaciones de estos hallazgos en el contexto de la clasificación de imágenes médicas.

Los resultados muestran que el método propuesto basado en WGAN-GP con el algoritmo *nearest neighbor* supera significativamente a otros métodos, incluyendo cGAN y PathoGAN, en la mayoría de las bases de datos evaluadas. En particular, la configuración WGAN-GP con *nearest neighbor* logró las mejores métricas de FID, MSE y SSIM en las bases de datos de *Retinal-Lesions*, FGADR y *Kaggle*. Esta superioridad puede atribuirse a la capacidad del WGAN-GP para manejar el problema de *mode collapse* [1], lo cual es una limitación común en otros tipos de GAN.

En la base de datos FGADR, aunque la WGAN-GP con *nearest neighbor* mostró un desempeño notable, se observó que la presencia de ruido en las imágenes impactó negativamente en su rendimiento. Este resultado subraya la necesidad de mejorar las técnicas de preprocesamiento y posiblemente investigar métodos adicionales de regularización para mejorar la robustez del modelo frente a datos ruidosos.

Una de las contribuciones más significativas de este trabajo es la capacidad del método propuesto para transferir lesiones entre imágenes de manera fiel. Las imágenes sintéticas generadas por WGAN-GP mantienen las características fundamentales de las imágenes reales, como el color, la estructura de los vasos sanguíneos y el disco óptico, mientras que las lesiones se replican con alta precisión. Sin embargo, se observó que en algunos casos, las lesiones aparecen ligeramente difuminadas. Esto podría mejorarse mediante la incorporación de técnicas de refinamiento de detalles en el generador.

[1] *Mode collapse* en una GAN ocurre cuando el generador produce un conjunto limitado de resultados repetitivos, ignorando la diversidad completa de los datos de entrenamiento, lo que limita la variedad de las imágenes generadas.

En comparación, los métodos cGAN y PathoGAN mostraron un rendimiento inferior en la transferencia de lesiones, con artefactos visibles y problemas en la replicación de colores. Este hallazgo resalta la importancia de la arquitectura del generador y del discriminador en la calidad de las imágenes generadas.

La evaluación por expertos en oftalmología proporcionó una validación cualitativa de la calidad de las imágenes generadas. Con una precisión promedio de 56.66 % en la distinción entre imágenes reales y sintéticas, los resultados indican que las imágenes generadas por el método propuesto tienen una alta fidelidad y pueden engañar a los expertos. La calificación promedio de las imágenes sintéticas también fue alta, lo que refuerza la idea de que las imágenes generadas no solo son visualmente plausibles, sino también clínicamente relevantes.

Los resultados varían según la base de datos utilizada, lo cual es esperado debido a las diferencias en tamaño, calidad y tipo de imágenes en cada base de datos. Por ejemplo, en la base de datos IDRiD, el sobreajuste fue un problema notable debido a la cantidad reducida de imágenes. Esto sugiere que el método propuesto podría beneficiarse de técnicas de regularización adicionales o de estrategias de aumentación de datos más sofisticadas para mejorar su generalización.

En la base de datos *Kaggle*, el desafío principal fue el tiempo de entrenamiento debido al tamaño de la base de datos. A pesar de esto, el método propuesto demostró ser eficiente, logrando métricas de FID, MSE y SSIM competitivas. Este resultado sugiere que, aunque el tiempo de entrenamiento es una consideración importante, la configuración óptima del modelo puede lograr una síntesis de alta calidad incluso en bases de datos grandes.

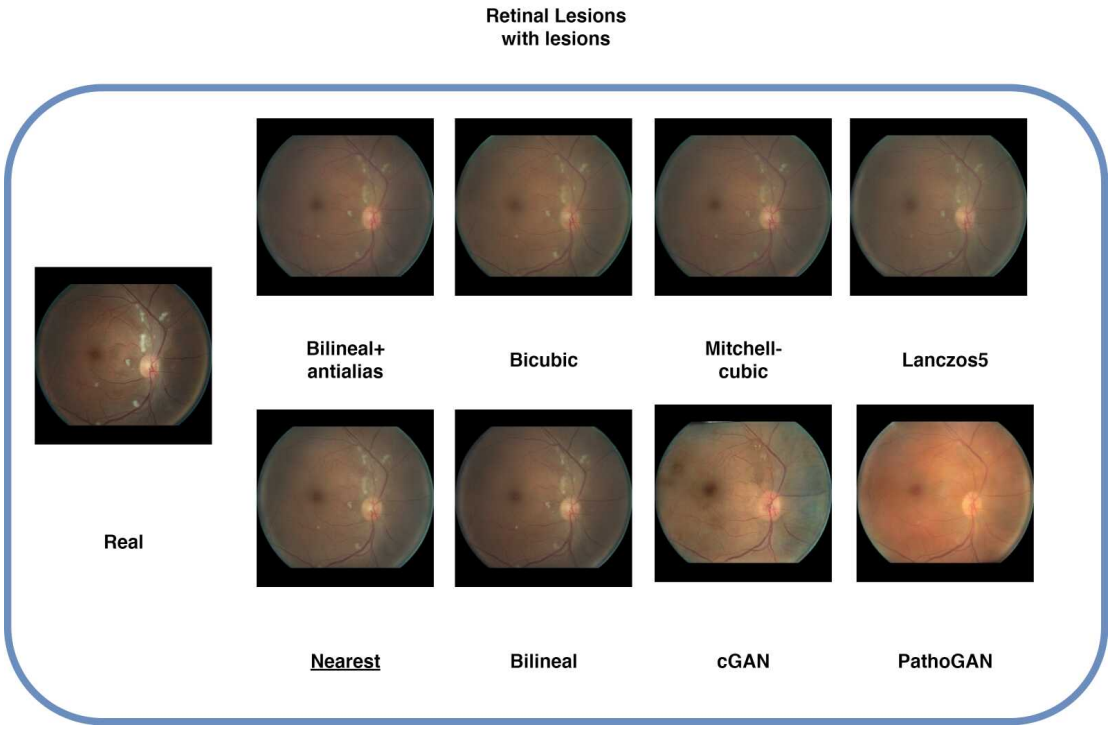
Aunque el método propuesto ha demostrado ser eficaz en varias métricas y contextos, existen algunas limitaciones que deben ser abordadas en trabajos futuros. La difuminación de las lesiones y los artefactos de color son áreas donde se puede mejorar. Además, la capacidad de manejar ruido en las imágenes sigue siendo un

desafío. Futuros trabajos podrían explorar arquitecturas más avanzadas, técnicas de preprocesamiento mejoradas y estrategias de aprendizaje semisupervisado para abordar estos problemas.

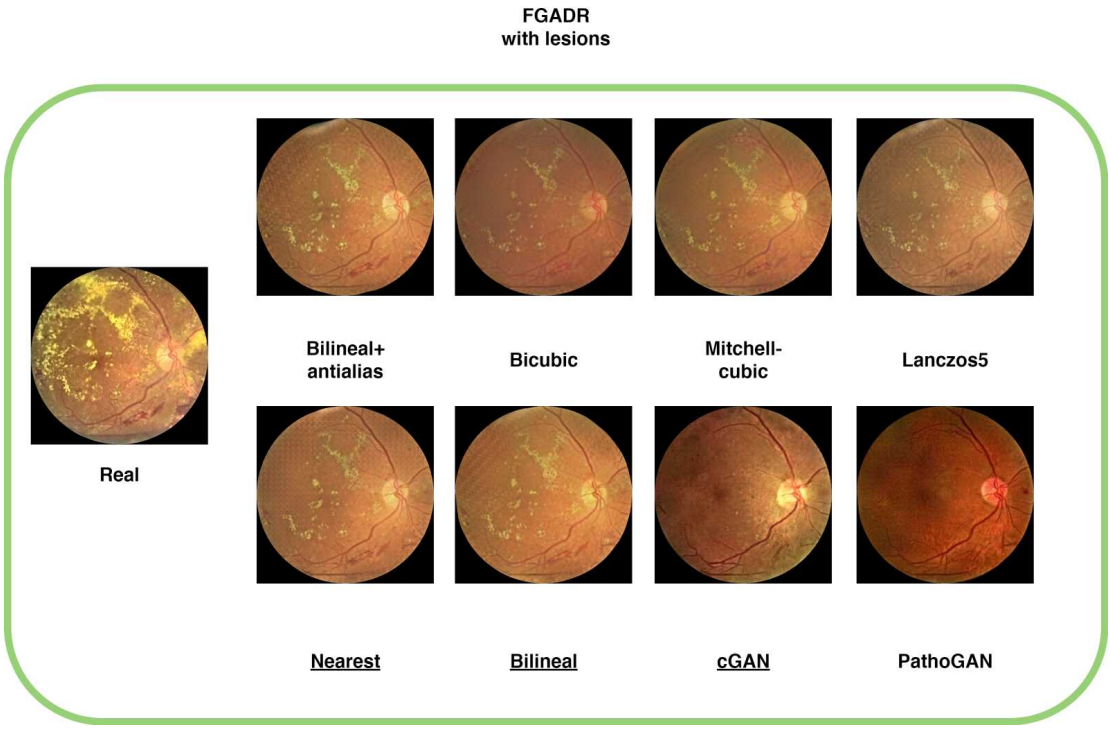
En conclusión, el método propuesto basado en WGAN-GP con *nearest neighbor* ha mostrado un rendimiento superior en la generación de imágenes sintéticas para la clasificación de imágenes médicas. La capacidad de transferir lesiones de manera fiel y la alta calidad de las imágenes generadas validada por expertos sugieren que este enfoque tiene un gran potencial para aplicaciones clínicas y de investigación.

Los hallazgos de esta investigación destacan la efectividad del enfoque propuesto basado en WGAN-GP, para la generación de imágenes médicas sintéticas. Los resultados obtenidos demuestran una mejora en la calidad y diversidad de las imágenes generadas en comparación con otros métodos existentes, esto subraya las ventajas de la metodología implementada, particularmente en la preservación de características clínicas y la reducción del error de predicción en contextos médicos.

Estos hallazgos proporcionan una base sólida para futuras investigaciones y aplicaciones en el campo de la imagen médica sintética. En el capítulo final, se discutirán las implicaciones de estos resultados, así como las oportunidades para la expansión y mejora del enfoque propuesto, abriendo nuevas direcciones para el desarrollo de técnicas avanzadas en este dominio.

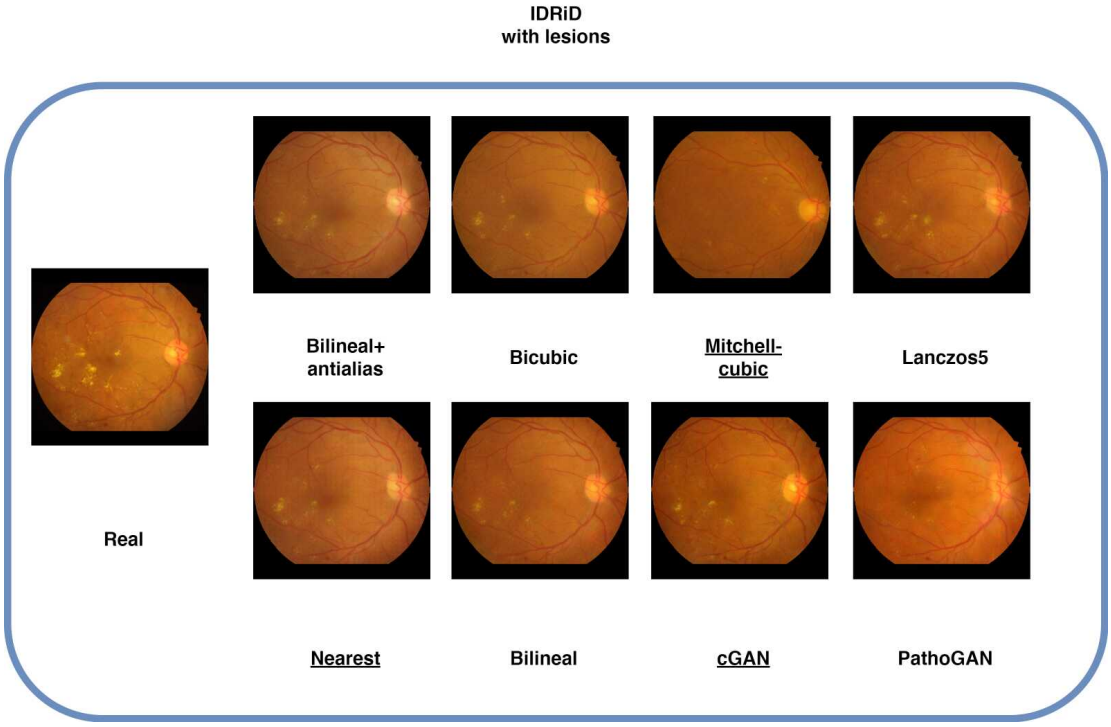


(a) Muestra de imágenes con lesiones, de la base de datos *Retinal-Lesions*.

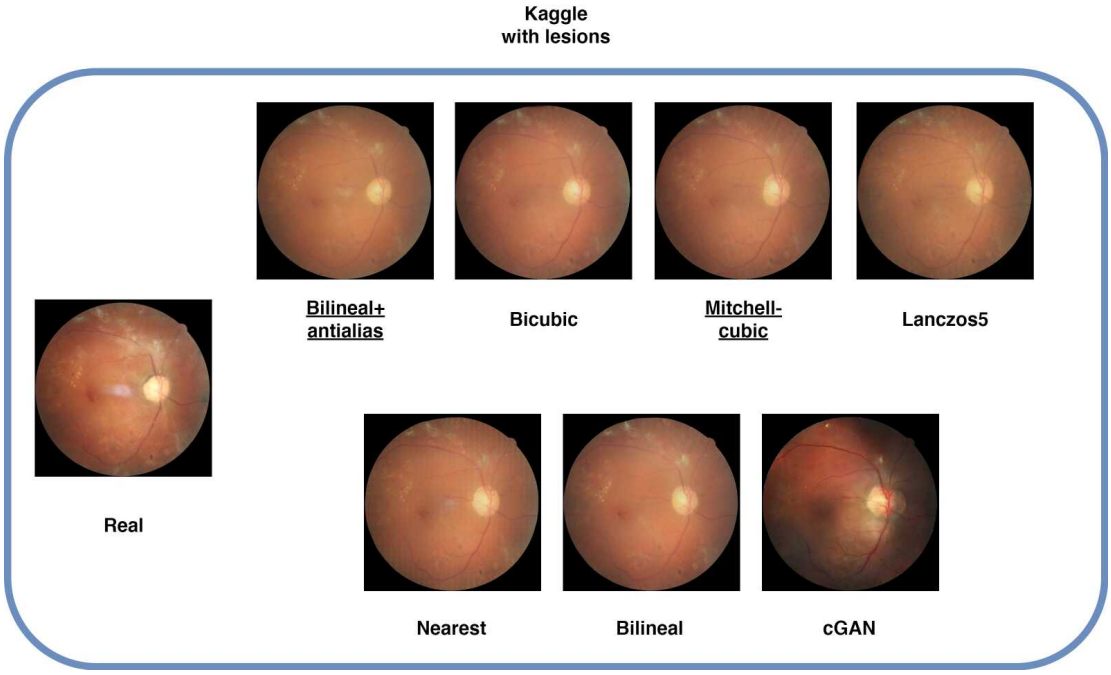


(b) Muestra de imágenes con lesiones, de la base de datos FGADR.

Figura 5.13: Comparaciones de imágenes con lesiones de cada base de datos.



(a) Muestra de imágenes con lesiones, de la base de datos IDRiD.



(b) Muestra de imágenes con lesiones, de la base de datos *Kaggle*.

Figura 5.14: Comparaciones de imágenes con lesiones de cada base de datos.

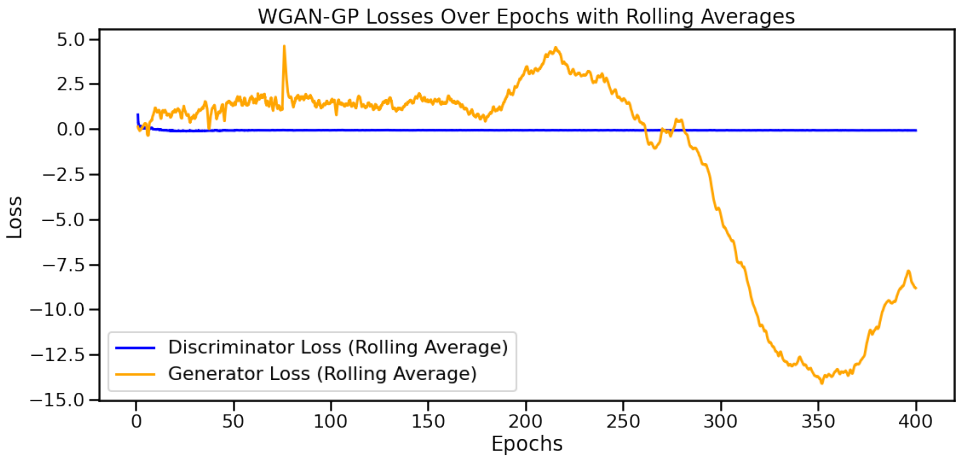


Figura 5.15: Gráfica de la pérdida con el método de media móvil.

CAPÍTULO 6

CONCLUSIONES Y TRABAJO FUTURO

La generación de imágenes médicas sintéticas es un desafío complejo en el ámbito de la síntesis de datos visuales, que involucra consideraciones rigurosas respecto a la calidad visual, la conservación de características clínicas críticas, la variabilidad de las imágenes producidas y su coherencia con las condiciones médicas originales. En este estudio, se planteó como objetivo general "Desarrollar un método basado en redes tipo GAN condicionales para la generación de imágenes médicas sintéticas, con el fin de incrementar la cantidad y diversidad de datos disponibles, mejorando así la precisión y el rendimiento de los modelos de diagnóstico". Para lograr este objetivo, se exploraron diversas estrategias de generación de imágenes, las cuales fueron implementadas y evaluadas utilizando varias configuraciones de redes WGAN-GP.

A través de los experimentos realizados y el análisis detallado de los resultados obtenidos, se alcanzaron conclusiones significativas que se presentan a continuación.

6.1. Conclusiones

En esta tesis se ha propuesto y evaluado un método basado en Wasserstein GAN con penalidad de gradiente (WGAN-GP) y transferencia de estilo para la generación de imágenes sintéticas de retina. Los resultados obtenidos a lo largo de

los experimentos realizados demuestran que el método propuesto es efectivo para la generación de imágenes realistas y de alta calidad, mejorando significativamente sobre otros enfoques como cGAN y PathoGAN. A continuación, se destacan las principales conclusiones del trabajo:

- Eficiencia en la Generación de Imágenes: El método WGAN-GP con *nearest neighbor* demostró ser superior en la mayoría de las métricas evaluadas (FID, MSE, SSIM) en diversas bases de datos, incluyendo *Retinal-Lesions*, FGADR y *Kaggle*. La capacidad de WGAN-GP para manejar problemas comunes de las GAN, como el *mode collapse*, fue fundamental para estos resultados.

- Transferencia de Lesiones: El método propuesto logró replicar de manera fiel las lesiones presentes en las imágenes originales, manteniendo la estructura general de la retina y minimizando artefactos. Esta capacidad es crucial para aplicaciones médicas donde la precisión y el detalle son esenciales.

- Validación Cualitativa: La evaluación por parte de expertos en oftalmología, los cuales 2 son miembros del Área de Ciencias y Tecnologías Biomédicas del INAOE, y un oftalmólogo externo, corroboró la alta calidad de las imágenes generadas, con una precisión del 56.66 % en distinguir entre imágenes reales y sintéticas. Esta validación cualitativa respalda el uso de las imágenes sintéticas generadas por el método propuesto en aplicaciones clínicas y de investigación.

- Robustez ante el Ruido: Aunque la base de datos FGADR presentó un desafío debido a su alto nivel de ruido, el método propuesto mostró resultados competitivos. Sin embargo, estos hallazgos subrayan la necesidad de mejorar las técnicas de preprocesamiento para manejar mejor el ruido en las imágenes.

- Algunas limitaciones del método incluyen la aparición de artefactos de color y la difuminación de ciertas lesiones. Además, el sobreajuste observado en bases de datos pequeñas como IDRiD sugiere la necesidad de técnicas de regularización adicionales. Igualmente, se utilizaron imágenes con una resolución

relativamente baja, por la limitación en procesamiento computacional en el proyecto, esto puede sugerir que usar imágenes con mejor resolución nos dé más detalles en las imágenes sintéticas.

6.2. Trabajo futuro

El presente trabajo sienta las bases para futuras investigaciones y mejoras en la generación de imágenes sintéticas para la clasificación de imágenes médicas. Algunas direcciones prometedoras para el trabajo futuro incluyen:

- Refinamiento de Detalles: Desarrollar técnicas adicionales para el refinamiento de detalles en las imágenes generadas, con el fin de reducir la aparición de artefactos de color y mejorar la definición de las lesiones.
- Manejo del Ruido: Investigar y aplicar técnicas avanzadas de preprocesamiento y regularización para mejorar la robustez del modelo ante imágenes ruidosas, especialmente en bases de datos como FGADR.
- Incremento en la resolución de datos: Emplear imágenes de mayor resolución para capturar detalles más finos, permitiendo la generación de imágenes sintéticas con un nivel de detalle superior. Esto mejorará la precisión de los modelos en la identificación y clasificación de características sutiles, contribuyendo a un análisis más exhaustivo y preciso. Además, el uso de datos de alta resolución puede facilitar la detección de patrones y anomalías que de otro modo pasarían desapercibidos.
- Integración con modelos del estado del arte: Explorar la integración de WGAN-GP con otros modelos generativos y de *deep learning*, como las Redes Adversariales Generativas Condicionales Variacionales (VAE-GAN, *Variational Auto-*

Encoder Generative Adversarial Network), los Transformadores y Difusión Estable, para mejorar la calidad y la funcionalidad de las imágenes generadas.

- Desarrollo de métricas para lesiones: Investigar técnicas avanzadas de evaluación de lesiones para medir con precisión la calidad y fidelidad de las lesiones generadas. Esto permitirá optimizar los algoritmos de generación, garantizando que las lesiones sintéticas sean realistas y representen fielmente las características de las lesiones reales.

- Desarrollo de Herramientas Interactivas: Crear herramientas interactivas para que los profesionales de la salud puedan generar y evaluar imágenes sintéticas en tiempo real, facilitando su adopción en la práctica clínica y en la investigación.

En conclusión, este trabajo ha demostrado el potencial de las WGAN-GP con transferencia de estilo en la generación de imágenes médicas sintéticas, abriendo nuevas oportunidades para mejorar la clasificación de imágenes médicas y avanzar en el diagnóstico y tratamiento de enfermedades. Las futuras investigaciones y desarrollos basados en estos hallazgos pueden contribuir significativamente al campo de la medicina digital y la inteligencia artificial aplicada a la salud.

Bibliografía

Abbasi-Sureshjani, S., Smit-Ockeloen, I., Zhang, J., and Ter Haar Romeny, B. (2015). Biologically-inspired supervised vasculature segmentation in slo retinal fundus images. In Kamel, M. and Campilho, A., editors, *Image Analysis and Recognition*, pages 325–334, Cham. Springer International Publishing.

Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein gan.

De, J., Cheng, L., Zhang, X., Lin, F., Li, H., Ong, K. H., Yu, W., Yu, Y., and Ahmed, S. (2016). A graph-theoretical approach for tracing filamentary structures in neuronal and retinal images. *IEEE Transactions on Medical Imaging*, 35(1):257–272.

Emma Dugas, J. J. W. C. (2015). Diabetic retinopathy detection.

Frid-Adar, M., Klang, E., Amitai, M. M., Goldberger, J., and Greenspan, H. (2018). Synthetic data augmentation using gan for improved liver lesion classification. *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 289–293.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial networks.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C. (2017). Improved training of wasserstein gans. *CoRR*, abs/1704.00028.

Guo, C., Szemenyei, M., Yi, Y., Wang, W., Chen, B., and Fan, C. (2020). Sa-unet: Spatial attention u-net for retinal vessel segmentation.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Klambauer, G., and Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a nash equilibrium. *CoRR*, abs/1706.08500.

Hoover, A., Kouznetsova, V., and Goldbaum, M. (2000). Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical Imaging*, 19(3):203–210.

Iqbal, T. and Ali, H. (2018). Generative adversarial network for medical images (mi-gan). *Journal of Medical Systems*.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976.

Kossen, T., Subramaniam, P., Madai, V. I., Hennemuth, A., Hildebrand, K., Hilbert, A., Sobesky, J., Livne, M., Galinovic, I., Khalil, A. A., Fiebach, J. B., and Frey, D. (2021). Synthesizing anonymized and labeled tof-mra patches for brain vessel segmentation using generative adversarial networks. *Computers in Biology and Medicine*, 131:104254.

Köhler, T., Budai, A., Kraus, M. F., Odstrčilik, J., Michelson, G., and Hornegger, J. (2013). Automatic no-reference quality assessment for retinal fundus images using vessel segmentation. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, pages 95–100.

Lee, H., Kim, S. T., Lee, J.-H., and Ro, Y. M. (2019). Realistic breast mass generation through birads category. In Shen, D., Liu, T., Peters, T. M., Staib, L. H.,

Essert, C., Zhou, S., Yap, P.-T., and Khan, A., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, pages 703–711, Cham. Springer International Publishing.

Niu, Y., Gu, L., Zhao, Y., and Lu, F. (2022). Explainable diabetic retinopathy detection and retinal image generation. *IEEE Journal of Biomedical and Health Informatics*.

Pachade, S., Porwal, P., Thulkar, D., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., Giancardo, L., Quellec, G., and Mériaudeau, F. (2020). Retinal fundus multi-disease image dataset (rfmid).

Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., and Meriaudeau, F. (2018). Indian diabetic retinopathy image dataset (idrid).

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252.

Sampath, V., Maurtua, I., Aguilar Martin, J., and Gutierrez, A. (2021). A survey on generative adversarial networks for imbalance problems in computer vision tasks. *Journal of Big Data*, 2021.

Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition.

Staal, J., Abramoff, M., Niemeijer, M., Viergever, M., and van Ginneken, B. (2004). Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4):501–509.

Toda, R., Teramoto, A., Tsujimoto, M., Toyama, H., Imaizumi, K., Saito, K., and Fujita, H. (2021). Synthetic ct image generation of shape-controlled lung cancer using semi-conditional infogan and its applicability for type classification. *International Journal of Computer Assisted Radiology and Surgery*.

Wang, Q., Zhou, X., Wang, C., Liu, Z., Huang, J., Zhou, Y., Li, C., Zhuang, H., and Cheng, J.-Z. (2019). Wgan-based synthetic minority over-sampling technique: Improving semantic fine-grained classification for lung nodules in ct images. *IEEE Access*, 7:18450–18463.

Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.

Wei, Q., Li, X., Yu, W., Zhang, X., Zhang, Y., Hu, B., Mo, B., Gong, D., Chen, N., Ding, D., and Chen, Y. (2020). Learn to segment retinal lesions and beyond. In *International Conference on Pattern Recognition (ICPR)*.

Wu, E., Wu, K., and Lotter, W. (2020). Synthesizing lesions using contextual gans improves breast cancer classification on mammograms.

Zhao, H., Li, H., Maurer-Stroh, S., and Cheng, L. (2018). Synthesizing retinal and neuronal images with generative adversarial nets. *Medical Image Analysis*, 49:14–26.

Zhao, H., Li, H., Maurer-Stroh, S., Guo, Y., Deng, Q., and Cheng, L. (2019). Supervised segmentation of un-annotated retinal fundus images by synthesis. *IEEE Transactions on Medical Imaging*.

Zhou, Y., Wang, B., He, X., Cui, S., and Shao, L. (2022). Dr-gan: Conditional generative adversarial network for fine-grained lesion synthesis on diabetic retinopathy images. *IEEE Journal of Biomedical and Health Informatics*.

Zhou, Y., Wang, B., Huang, L., Cui, S., and Shao, L. (2021). A benchmark for studying diabetic retinopathy: Segmentation, grading, and transferability. *IEEE Transactions on Medical Imaging*, 40(3):818–828.