



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
**POSGRADO EN CIENCIAS E INGENIERÍA DE LA COMPUTACIÓN**

**TRANSFERENCIA DE ESTILO NEURONAL EN PEQUEÑOS CONJUNTOS  
DE IMÁGENES MÉDICAS**

**TESIS**  
**QUE PARA OPTAR POR EL GRADO DE:**  
**MAESTRO EN CIENCIAS E INGENIERÍA DE LA COMPUTACIÓN**

**PRESENTA:**  
**ALAN MAURICIO CAMARGO HERNÁNDEZ**

**TUTORA PRINCIPAL**  
**DRA. JIMENA OLVERES MONTIEL**  
**FACULTAD DE INGENIERÍA, UNAM**

**CIUDAD UNIVERSITARIA, CDMX, 2024**



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

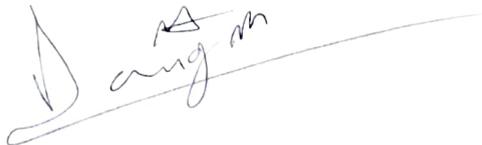
**PROTESTA UNIVERSITARIA DE INTEGRIDAD Y  
HONESTIDAD ACADÉMICA Y PROFESIONAL**

De conformidad con lo dispuesto en los artículos 87, fracción V, del Estatuto General, 68, primer párrafo, del Reglamento General de Estudios Universitarios y 26, fracción I, y 35 del Reglamento General de Exámenes, me comprometo en todo tiempo a honrar a la institución y a cumplir con los principios establecidos en el Código de Ética de la Universidad Nacional Autónoma de México, especialmente con los de integridad y honestidad académica.

De acuerdo con lo anterior, manifiesto que el trabajo escrito titulado Transferencia de estilo neuronal en pequeños conjuntos de imágenes médicas, que presenté para obtener el grado de Maestro en ciencias e ingeniería de la computación, es original, de mi autoría y lo realicé con el rigor metodológico exigido por mi Programa de Posgrado, citando las fuentes, ideas, textos, imágenes, gráficos u otro tipo de obras empleadas para su desarrollo.

En consecuencia acepto que la falta de cumplimiento de las disposiciones reglamentarias y normativas de la Universidad, en particular las ya referidas en el Código de Ética, llevará a la nulidad e los actos de carácter académico administrativo del proceso de titulación/graduación

**Atentamente**

A handwritten signature in blue ink, appearing to read 'Alan Mauricio Camargo Hernández', written over a horizontal line.

Alan Mauricio Camargo Hernández

313123391



# Resumen

Este trabajo presenta un análisis del comportamiento del estilo neuronal, con el objetivo de optimizar su aplicación en tareas de clasificación. También, se implementaron dos estrategias de regularización basadas en la Transferencia de Estilo Neuronal (TEN). La primera estrategia utiliza TEN con control espacial junto con estilos interpolados para el aumento de imágenes en el entrenamiento de modelos de aprendizaje profundo. La segunda estrategia emplea TEN para extraer estilos de imágenes, que, junto a los estilos interpolados, se utilizan para entrenar modelos de aprendizaje máquina. Los resultados obtenidos sugieren que las estrategias propuestas mejoran significativamente el rendimiento de clasificación bajo ciertas circunstancias, especialmente en pequeños conjuntos de datos. A través de un análisis estadístico, se evaluó el impacto de estas estrategias, demostrando que las técnicas basadas en estilos neuronales pueden ofrecer ventajas importantes en la clasificación de imágenes médicas.

# Agradecimientos

Agradezco a mi familia por su apoyo incondicional y constante motivación, que han sido el aliento en cada paso de este proyecto y de mi formación académica. A mis amigos, por su compañía y ánimo durante este camino, haciendo más llevadero cada desafío.

Finalmente, agradezco al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCyT) por el apoyo económico brindado (CVU: 1228000) durante mis estudios de posgrado, que me permitió dedicarme de tiempo completo a mis estudios e investigación. Asimismo, agradezco a los programas UNAM-PAPIIT IN108624 e IT101624 por su apoyo, que ha sido esencial para la realización de este proyecto.

“Tecnología del vuelo compartido: el primer pato que levanta vuelo abre paso al segundo, que despeja el camino al tercero, y la energía del tercero alza al cuarto, que ayuda al quinto, y el impulso del quinto empuja al sexto, que presta viento al séptimo...

Cuando se cansa, el pato que hace punta baja a la cola de la bandada y deja su lugar a otro, que sube al vértice de esa V que los patos dibujan en el aire. Todos se van turnando, atrás y adelante; y ninguno se cree superpato por volar adelante, ni subpato por marchar atrás.”

– E. Galeano (2012), *Los hijos de los días*, "Mayo 1. Día de los trabajadores."

# Índice

Índice de figuras	6
Índice de tablas	7
<b>1. Introducción</b>	<b>8</b>
1.1. Antecedentes	8
1.1.1. Transferencia de Estilo Neuronal (TEN) en imágenes de lesiones de la piel y ultrasonido de mama	9
1.2. Planteamiento del problema	10
1.3. Hipótesis	10
1.4. Objetivo	10
1.4.1. Objetivos específicos	11
1.5. Contribución	11
1.6. Síntesis y estructura de la tesis	12
<b>2. Marco Teórico</b>	<b>13</b>
2.1. Breve historia de la Inteligencia Artificial (IA)	13
2.2. Aprendizaje máquina (AM)	14
2.2.1. Indicadores de desempeño	17
2.2.2. Validación Cruzada y K folds	18
2.3. Análisis de múltiples dimensiones	18
2.4. Introducción al aprendizaje profundo (AP)	20
2.4.1. El perceptrón simple	20
2.4.2. El descenso del gradiente (DG)	21
2.4.3. Redes Neuronales Profundas (RNP)	21
2.4.4. Redes Neuronales Convolucionales (RNC)	22
2.4.5. Transferencia de aprendizaje y fine-tuning	23
2.5. Aumento de datos	24
2.6. Transferencia de Estilo Neuronal (TEN)	25
2.6.1. Noción del estilo en la transferencia de estilo Neuronal	26
2.6.2. Transferencia de Estilo Neuronal (TEN) con control espacial	27
2.6.3. Otras técnicas TEN	28
2.7. Otros métodos de generación de imágenes.	29
<b>3. Metodología</b>	<b>31</b>
3.1. Bases de datos utilizadas	32

3.1.1. Breast cancer	32
3.1.2. Base de datos HAM 10K	33
3.1.3. Conjuntos de estudio	34
3.2. Experimentos	37
3.2.1. Análisis del estilo neuronal	38
3.2.1.1. Análisis 1: estilos con y sin control espacial	39
3.2.1.2. Análisis 2: estilos interpolados	40
3.2.1.3. Interpolación de estilos	40
3.2.2. Experimentos de clasificación	41
3.2.2.1. Experimento de aprendizaje máquina AM	41
3.2.2.2. Experimento de aprendizaje profundo AP	42
3.2.2.3. Comparación de modelo - conjunto	43
3.3. Detalles Técnicos	44
<b>4. Resultados y análisis</b>	<b>46</b>
4.1. Análisis de estilos neuronales	46
4.1.1. Análisis 1: estilos con y sin control espacial.	46
4.1.2. Análisis 2: interpolación de estilos.	47
4.2. Ejemplos de imágenes generadas con TEN	48
4.3. Experimentos de clasificación	50
4.3.1. Resultados de experimentos AM	50
4.3.2. Resultados de experimentos AP	54
4.3.3. Resultados por modelo (AP y AM)	57
<b>5. Conclusiones</b>	<b>60</b>
5.1. Conclusiones por experimento	60
5.1.1. Análisis del estilo neuronal	60
5.1.2. Experimentos de aprendizaje máquina	60
5.1.3. Experimentos de aprendizaje profundo	61
5.1.4. Comparación de mejores Modelos-Conjunto	62
5.2. Conclusiones generales	62
5.3. Trabajo futuro	63
<b>A. Apéndice</b>	<b>65</b>
A. 1. Experimento de aprendizaje profundo sobre solo 3 diagnósticos del conjunto HAM 10K	65
A. 1. 1. Hipótesis	65
A. 1. 2. Metodología	65
A. 1. 3. Resultados de exactitud para HAM 10 K con solo 3 diagnósticos	65
A. 1. 4. Discusión y conclusiones	66
<b>6. Bibliografía</b>	<b>67</b>

# Índice de figuras

- Figura 1. Definición de la matriz de confusión, basada en [30].
- Figura 2. Proceso de transferencia de estilo neuronal TEN, basada en [49].
- Figura 3. Obtención de una matriz Gram.
- Figura 4. Obtención del estilo de una imagen.
- Figura 5. Ejemplo de transferencia de estilo con control espacial en imágenes de lesiones de la piel.
- Figura 6. Obtención de una matriz Gram con control espacial.
- Figura 7. Clasificación: métodos de transferencia de estilo.
- Figura 8. Ejemplos de imágenes de cáncer de mama.
- Figura 9. Descripción de la metodología.
- Figura 10. Descripción del análisis del estilo.
- Figura 11. Proceso de análisis estilo con y sin control espacial.
- Figura 12. Proceso de análisis de estilos interpolados.
- Figura 13. Ejemplo práctico de interpolación de estilos elemento a elemento.
- Figura 14. Experimento de aprendizaje máquina con estilos.
- Figura 15. Experimento de aprendizaje profundo.
- Figura 16. Análisis 1. Ultrasonido de cáncer de mama.
- Figura 17. Análisis 1. Lesiones de la piel.
- Figura 18. Análisis 2, Estilos interpolados de ultrasonido de cáncer de mama.
- Figura 19. Análisis 2, Estilos interpolados de lesiones de la piel.
- Figura 20. Ejemplos de imágenes generadas con TEN y control espacial para ultrasonido de cáncer de mama.
- Figura 21. Ejemplos de imágenes generadas con TEN para lesiones de la piel.
- Figura 22. Ejemplos de imágenes generadas con estilos interpolados para ultrasonido de cáncer de mama.
- Figura 23. Ejemplos de imágenes generadas con estilos interpolados para lesiones de la piel.
- Figura 24. Resultados de la base de datos *Breast Cancer Ultrasound* experimento aprendizaje máquina.
- Figura 25. Resultados de la base de datos *HAM 10K* experimento aprendizaje máquina.
- Figura 26. Resultados de la base de datos *Breast Cancer Ultrasound* experimento aprendizaje profundo.
- Figura 27. Resultados de la base de datos *HAM 10K* experimento aprendizaje profundo.
- Figura 28. Mejores resultados usando los conjuntos de estudio por cada modelo en el subconjunto de prueba de ultrasonido de cáncer de mama.
- Figura 29. Mejores resultados usando los conjuntos de estudio por cada modelo en el subconjunto de prueba de lesiones de la piel.

# Índice de tablas

Tabla 1. Resumen de la base *Breast cancer*.

Tabla 2. *Breast Ultrasound Images Dataset* después del preprocesamiento.

Tabla 3. Resumen de *HAM 10K*.

Tabla 4. *HAM 10K* después del preprocesamiento.

Tabla 5. Resumen de conjuntos de estudio.

Tabla 6. Subconjunto de prueba para *Breast Ultrasound Images Dataset*.

Tabla 7. Subconjunto de prueba *HAM 10K*.

Tabla 8. Prueba *post-hoc* de Nemenyi para el Experimento AM con imágenes de ultrasonido de cáncer de mama.

Tabla 9. Prueba *post-hoc* de Nemenyi para el Experimento AM con imágenes de lesiones de la piel.

Tabla 10. Prueba *post-hoc* de Nemenyi para el Experimento AP con imágenes de ultrasonido de cáncer de mama.

Tabla 11. Prueba *post-hoc* de Nemenyi para el Experimento AP con imágenes de lesiones de la piel.

# 1. Introducción

## 1.1. Antecedentes

En la actualidad, los modelos de aprendizaje máquina (AM) y aprendizaje profundo (AP) han logrado buenos resultados en tareas tales como la clasificación y segmentación de imágenes, gracias a su capacidad de generalizar el conocimiento. Esto significa que los modelos pueden obtener buenos rendimientos al enfrentarse a datos no vistos previamente. Sin embargo, es necesario evitar el sobreentrenamiento, un problema donde los modelos tienen buen desempeño con los datos de entrenamiento, pero fallan al hacer inferencia con datos nuevos que nunca ha visto. Una técnica para mitigar este problema es el aumento de datos, que trata el problema desde el inicio, los datos. Esta estrategia comprende métodos diseñados para ampliar y/o mejorar la calidad de los conjuntos de datos, buscando una mejor generalización del conocimiento. Existen estrategias de aumento de datos basadas en aprendizaje profundo, como la generación de imágenes mediante redes generativas adversarias (RGA, o GAN por sus siglas en inglés) o la Transferencia de Estilo Neuronal (TEN) [1].

Los modelos de aprendizaje máquina y aprendizaje profundo se han utilizado con éxito en imágenes médicas para diagnosticar enfermedades al identificar biomarcadores, como características propias de tumores, y así predecir la malignidad o evaluar la respuesta a tratamientos. Para obtener buenos resultados, es fundamental disponer de un número considerable de imágenes médicas anotadas por expertos, lo que implica tiempo, costos y una posible falta de atención a pacientes. En este contexto, las estrategias de aumento de datos con aprendizaje profundo representan una solución prometedora, al permitir la generación de imágenes médicas sintéticas. Estas imágenes pueden ser útiles en la creación de simulaciones realistas para el entrenamiento de modelos cuando los conjuntos de datos son limitados, P. K. Mall, *et al.*, 2023 [2].

Cuando se habla de datos en modelos de aprendizaje profundo para visión computacional, generalmente se hace referencia a conjuntos de imágenes. En cambio, los modelos de aprendizaje máquina, como las máquinas de soporte vectorial, k vecinos más cercanos (KNN, por sus siglas en inglés), árboles de decisión, Naive Bayes, etc. requieren una extracción de características antes de poder realizar tareas como la clasificación o detección de enfermedades [3].

La generación de datos sintéticos se considera hoy en día una opción viable. Un ejemplo de ello es la plataforma MONAI, diseñada específicamente para generar imágenes médicas sintéticas, esta plataforma utiliza modelos de difusión, transformers autorregresivos y redes GAN para crear imágenes como tomografías computarizadas (CT), resonancias magnéticas (MRI), rayos X, entre otros formatos, W. H. L. Pinaya, *et al.*, 2023 [4].

### 1.1.1. Transferencia de Estilo Neuronal (*TEN*) en imágenes de lesiones de la piel y ultrasonido de mama

La Transferencia de Estilo Neural (*TEN*) es un método basado en AP que permite la manipulación y fusión, del contenido y estilo de diferentes imágenes, produciendo resultados visiblemente buenos. Gatys, *et al.* [5] introdujeron este método “Transferencia de Estilo Neural” en su trabajo de 2015. Propusieron utilizar Redes Neuronales Convolucionales (RNC) para transferir el estilo de una imagen a otra, específicamente la red VGG 19. En este enfoque, el contenido de una imagen se define como el conjunto de mapas de características del último bloque convolucional de la red VGG. Mientras que el estilo se representa mediante un conjunto de las matrices *Gram*, obtenidas a partir de los mapas de características de cada bloque convolucional de la red VGG. Más específicamente, el estilo se compone de 5 matrices, estas matrices tienen dimensiones de  $64 \times 64$ ,  $128 \times 128$ ,  $256 \times 256$ ,  $512 \times 512$  y  $512 \times 512$ .

Posteriormente, Gatys, *et al.* ampliaron el método en 2017 [6] al proponer el control espacial en la transferencia de estilo. Esta técnica permite transferir el estilo de un área específica de una imagen, al contenido de un área específica de otra. Además del control espacial, también sugirieron la posibilidad de controlar aspectos como el color, escala de estilos y la velocidad del método, ampliando aún más las posibilidades de la transferencia de estilo neuronal.

Además de la expresión artística, la transferencia de estilo neuronal ha encontrado aplicaciones en el diagnóstico de enfermedades, utilizando imágenes médicas. Mikołajczyk, *et al.* en 2019 [7], exploraron su uso al utilizar *TEN* como técnica de regularización en el aprendizaje profundo, específicamente en la generación de imágenes para el análisis de lesiones de piel. Su estudio involucró la selección de imágenes de estilo y contenido aleatorias para generar nuevas imágenes utilizando *TEN*, que luego se etiquetan con una red neuronal convolucional previamente entrenada. Estas imágenes generadas se ocuparon para entrenar diferentes arquitecturas de redes neuronales, mientras que las imágenes reales se reservaron para los conjuntos de validación y prueba.

En 2020, Liu, *et al.* [8], propusieron la aplicación de la transferencia de estilo neuronal en imágenes de ultrasonido. Su trabajo se centra en abordar el desafío de la variabilidad en la apariencia de estas imágenes, ya que suelen estar afectadas por el ruido. Específicamente, emplearon *TEN* para la segmentación de imágenes fetales. Este tipo de imágenes puede verse influenciado por factores como el ruido y las diferencias en la configuración del dispositivo o en la propia captura de la imagen. Esto provoca que existan cambios en la apariencia que pueden dificultar la tarea de segmentación al utilizar modelos de aprendizaje profundo. En particular, su investigación se enfocó en desarrollar un modelo innovador, denominado *WaveCT-AIN-D*, que utiliza una metodología avanzada para seleccionar las imágenes de estilo y contenido, haciendo uso de Patrones Binarios Locales (LBP, por sus siglas en inglés).

En el 2021, Liu, *et al.* [9], al igual que en el 2020 [8] retomaron el problema del cambio de apariencia en imágenes de ultrasonido fetal, pero ahora proponiendo un modelo llamado *DINSeg*, donde proponen hacer un cambio de apariencia con *TEN* y luego realizar la segmentación. El principal aporte de *DINSeg* es buscar unir estos 2 pasos en un solo proceso.

Otro trabajo interesante es el de Yutong Li, *et al.* en 2021 [10], donde utilizaron la transferencia de estilo neuronal con control espacial para extraer todos los estilos de las lesiones de imágenes de su conjunto de dermatología, para después tratar de pasar cada estilo del conjunto de entrenamiento al

contenido de una sola imagen definida previamente, de esta forma se buscó hacer la clasificación de las imágenes solo teniendo en cuenta el estilo de las imágenes.

La Transferencia de Estilo Neuronal ha emergido como método innovador en el campo de la inteligencia artificial, con aplicaciones prometedoras en el ámbito médico. En el contexto de las imágenes médicas, se han desarrollado diversas estrategias para implementar la TEN, cada una adaptada a tareas específicas o enfoques particulares. En el caso de las lesiones cutáneas, se distinguen dos estrategias: la primera se centra en el aumento de datos, generando nuevas imágenes para enriquecer los conjuntos de entrenamiento y, de este modo, mejorar la precisión de los modelos de clasificación. La segunda estrategia se enfoca en extraer y transferir el estilo de las lesiones a un contenido previamente definido, permitiendo que los modelos de clasificación se concentren exclusivamente en analizar las características del estilo de las lesiones. Por otro lado, en el campo de las imágenes de ultrasonido, la TEN se utiliza para mitigar las variaciones en la apariencia de las imágenes, causadas por los procesos de obtención de imágenes. Esta diversidad de aplicaciones nos hace pensar que existen múltiples estrategias para mejorar el rendimiento de los modelos de inteligencia artificial mediante la implementación de técnicas de transferencia de estilo, abriendo así un amplio campo de investigación y desarrollo en esta área.

## 1.2. Planteamiento del problema

Se requiere una gran cantidad de datos para modelos de aprendizaje profundo, y para tareas muy específicas se puede presentar el problema de no contar con los volúmenes de imágenes necesarios.

- Las imágenes de ultrasonido presentan un desafío para modelos de aprendizaje profundo debido al ruido inherente que se encuentra en ellas, el cual está directamente relacionado con el proceso de adquisición de las imágenes.
- Las imágenes de lesiones de la piel tienen diferentes formas en la lesión, texturas y artefactos que pueden hacer difícil su diagnóstico.

## 1.3. Hipótesis

Las estrategias basadas en la Transferencia de Estilo Neuronal (TEN), como la generación de imágenes y la extracción de estilos neuronales, considerando la lesión y el diagnóstico, pueden mejorar el rendimiento en la clasificación de diagnósticos de los modelos de aprendizaje automático (AM) y aprendizaje profundo (AP) cuando se entrenan con cantidades limitadas de datos.

## 1.4. Objetivo

Examinar los estilos neuronales, con y sin control espacial, e implementar dos estrategias basadas en transferencia de estilo neuronal para el aumento de datos: (1) aumentar la cantidad de imágenes mediante TEN teniendo en cuenta los diagnósticos y lesiones, y (2) emplear el estilo neuronal como técnica de extracción de características para modelos de aprendizaje máquina. Estas estrategias se aplican a imágenes de ultrasonido de cáncer de mama y lesiones de la piel, para tratar de mejorar la exactitud en tareas de clasificación en conjuntos de datos pequeños.

### 1.4.1. Objetivos específicos

- Analizar las aplicaciones de la transferencia de estilo neuronal en imágenes médicas, específicamente en ultrasonidos de cáncer de mama y lesiones de piel, para identificar enfoques relevantes.
- Establecer un preprocesamiento y análisis de conjuntos de imágenes de ultrasonido y lesiones de la piel que defina subconjuntos adecuados de entrenamiento, validación y prueba.
- Investigar las características del estilo neuronal en imágenes médicas mediante técnicas de reducción de dimensionalidad y análisis de matrices de correlación y distancias euclidianas, utilizando un subconjunto representativo.
- Explorar la eficacia del uso de estilos neuronales e interpolados como entrada para modelos de aprendizaje automático en la clasificación de diagnósticos.
- Desarrollar y evaluar conjuntos de imágenes generados mediante transferencia de estilo neuronal con control espacial y estilos interpolados para incrementar la diversidad de datos en entrenamiento.
- Evaluar la eficiencia de modelos de aprendizaje profundo y aprendizaje automático entrenados con conjuntos generados mediante TEN, para determinar su efectividad en la mejora de la clasificación de imágenes médicas.

## 1.5. Contribución

Las principales contribuciones de este trabajo son:

1. Análisis comparativo de TEN entre diagnósticos:
  - Se investiga el uso del estilo neuronal en imágenes médicas, con y sin control espacial sobre las lesiones, para evaluar su capacidad de discriminación entre diferentes diagnósticos.
2. Aplicación de transferencia de estilo neuronal (TEN) con control espacial sobre las lesiones en imágenes médicas:
  - Se propone el uso de TEN con control espacial sobre las lesiones de las imágenes de ultrasonido y lesiones de la piel.
  - Se investiga la eficacia de esta técnica para mitigar el problema de contar con conjuntos de datos limitados en imágenes médicas.
3. Desarrollo de una nueva sub estrategias de aumento de datos:
  - Se propone una técnica de interpolación entre estilos de imágenes con el mismo diagnóstico, aumentando la diversidad de estilos para el aumento de datos.
4. Una nueva estrategia en extracción de características:
  - Se estudia el potencial del estilo neuronal como método para la extracción de características en imágenes médicas junto con el impacto que estas pueden tener los modelos de aprendizaje máquina.

## 1.6. Síntesis y estructura de la tesis

Durante el desarrollo de este trabajo, se realizó un análisis del comportamiento del estilo neuronal con el objetivo de determinar cómo aprovecharlo mejor en tareas de clasificación. A partir de estos resultados, se implementaron dos estrategias de regularización basadas en la transferencia de estilo neuronal. La primera estrategia emplea TEN (Transferencia de Estilo Neuronal) y una propuesta de estilos interpolados para aumentar los datos de entrenamiento y entrenar modelos de aprendizaje profundo. En la segunda, se utiliza TEN para extraer los estilos de las imágenes y, junto con los estilos interpolados, se entrena a los modelos de aprendizaje máquina.

En el segundo capítulo, se describe el marco teórico, el cual trata de ofrecer una visión de la inteligencia artificial (IA), algunos fundamentos esenciales del aprendizaje máquina (AM), técnicas de reducción de dimensionalidad, el aprendizaje profundo (AP), y el método de transferencia de aprendizaje neuronal (TEN). En el mismo capítulo, se profundiza sobre el aumento de datos. En el tercer capítulo, se explica la metodología que se emplea para desarrollar el trabajo, detallando las bases de datos (de ultrasonido y lesiones de la piel), los conjuntos de estudio (conjuntos que se utilizan para entrenar y validar los modelos), los subconjuntos de prueba para las dos bases de datos, las plataformas y bibliotecas utilizadas, y una explicación detallada de cada experimento. En el capítulo de resultados y análisis, se detallan los resultados obtenidos. En el capítulo siguiente, se presentan las conclusiones tanto para cada experimento como de forma general, y por último se proponen varios caminos para futuros trabajos.

## 2. Marco Teórico

Este capítulo establece el contexto teórico que sustenta las técnicas y metodologías empleadas en los experimentos. Se exploran temas clave como la inteligencia artificial (IA), el aprendizaje automático (AM) y el aprendizaje profundo (AP). Además, se abordan las técnicas de reducción de dimensionalidad, que permiten simplificar la representación de datos tratando de no perder información relevante, con el objetivo de observar la distribución de los datos de alta dimensionalidad. También se aborda el aumento de datos, que es una estrategia que incrementa la cantidad disponibles mediante transformaciones de los datos, con el fin de ayudar a mejorar la capacidad de generalización de los modelos. Por último, se presenta la Transferencia de Estilo Neuronal (TEN) como una estrategia avanzada para la generación y manipulación de imágenes.

### 2.1. Breve historia de la Inteligencia Artificial (IA)

La inteligencia artificial (IA) tomó un lugar en la academia formalmente en la década de 1950, con la famosa conferencia de *Dartmouth* en 1956. Sin embargo, la noción de inteligencia artificial ya existía en la ciencia ficción, un ejemplo de esto es "*I, Robot*" de Isaac Asimov (1950), o incluso su origen puede observarse en inventos antiguos como el autómatas de Filón de Bizancio (siglo III a.C.). Tras avances significativos durante la segunda mitad del siglo XX, la IA empezó a tener varias ramas que son los cimientos de los descubrimientos de hoy en día [11], [2].

Desde la segunda mitad del siglo XX hasta la actualidad, hubo avances significativos en la IA, como la red *LeNet* desarrollada por Yann LeCun y sus colaboradores en 1998 [12], y demás arquitecturas de redes neuronales. Este corto periodo de tiempo, fue el inicio de la revolución de la IA, impulsada por varios trabajos, como el desarrollo e investigación de arquitecturas de redes neuronales y la capacidad de cómputo. Un ejemplo de esto, son las Redes Neuronales Convolucionales (RNC o CNN, por sus siglas en inglés) que surgieron como modelos prometedores para el procesamiento de imágenes y la visión por computadora. Años después, las Redes Generativas Adversarias (GAN) propuestas por Ian Goodfellow y sus colegas en 2014 [13], abrieron la posibilidad de generar datos sintéticos, en especial, imágenes. Más recientemente, los modelos de difusión han surgido como una alternativa para la generación de imágenes de muy alta calidad. Este aumento en la creación de arquitecturas y modelos ha traído avances sin precedentes en múltiples tareas, marcando el inicio de lo que podemos considerar una era dorada de la inteligencia artificial [14].

La IA ha tratado de resolver problemas de la medicina desde sus inicios formales como MYCIN [15] o CASNET [16], por lo tanto, ha experimentado una evolución en las tareas que trata de resolver, y en sus capacidades para hacerlo. En los primeros pasos, la IA se centró en el desarrollo de modelos de búsqueda de información y sistemas expertos. Mientras que en los últimos años, las aplicaciones de la

IA en medicina se han enfocado en el desarrollo del aprendizaje profundo. Esta técnica avanzada ha permitido realizar la segmentación de imágenes médicas con alta precisión [17], proporcionar diagnósticos asistidos por computadora [18] y ofrecer apoyo a los médicos en la toma de decisiones clínicas [19]. El potencial del aprendizaje profundo ha hecho posible utilizar diferentes datos médicos, desde imágenes de diferentes formatos hasta registros clínicos y datos genómicos, haciendo de este modo posible que se abarquen diversas especialidades médicas [14].

## 2.2. Aprendizaje máquina (AM)

El aprendizaje máquina (AM o ML, por sus siglas en inglés), también llamado aprendizaje automático (AA), tiene varias definiciones que dependen del autor; sin embargo, una de las definiciones populares es la de Tom Mitchell, en el libro *Machine Learning* de 1997 [20]:

“Se dice que un programa de computadora aprende de la experiencia  $E$  con respecto a alguna clase de tareas  $T$ , y la medida de desempeño  $P$ , si su desempeño en las tareas en  $T$ , medido por  $P$ , mejora con experiencia  $E$ .” [20].

El aprendizaje máquina amplía las posibilidades de realizar clasificaciones basadas en el uso de la experiencia para adaptar los modelos al entorno. Uno de los tipos más comunes de AM es el aprendizaje supervisado, que son algoritmos que pueden detectar patrones en datos para así predecir datos futuros o para realizar clasificación [21]. Algunos algoritmos o métodos de aprendizaje máquina supervisados son:

- Naive Bayes, es un método de aprendizaje supervisado basado en el teorema de Bayes, donde se establece que la probabilidad de una hipótesis  $H$  está dada por las observaciones o la evidencia  $E$ , es decir,  $P(H|E)$ . Se dice que es *naive* o ingenuo porque el método supone que todas las características o evidencias tienen independencia entre sí, lo que raramente ocurre, pero simplifica los cálculos [22]. De manera más específica, el clasificador de Naive Bayes busca maximizar la probabilidad a posteriori  $P(C|X)$  teniendo un conjunto de  $X$ , donde:
  - $C$  representa a una de las posibles clases en un conjunto discreto  $\{C_1, C_2, \dots, C_n\}$ .
  - $X \subset \mathbb{R}^d$  es un conjunto de vectores de características, donde  $d$  es el número de características.

Utilizando la fórmula del teorema de Bayes tenemos la ecuación 1.

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (1)$$

Donde  $P(C|X)$  son las posibilidades a posteriori,  $P(X)$  las posibilidades a priori y  $P(X|C)$  es la probabilidad de que ocurra una instancia de  $X$  dado que se tiene la clase  $C$  [22].

La complejidad computacional de Naive Bayes se divide en 2 partes, una para el entrenamiento y otra para inferencia:

- Entrenamiento: la complejidad del entrenamiento en Naive Bayes es  $O(n \cdot d)$ , donde  $n$  es el número de instancias de entrenamiento y  $d$  es el número de características. Se comienza calculando la probabilidad a priori de cada clase. Esto se logra contando las instancias que pertenecen a cada clase y dividiendo entre el total de instancias, lo cual tiene una complejidad de  $O(n)$ . Después de obtener las probabilidades a priori, el algoritmo calcula la probabilidad condicional de cada característica dada una clase. Para cada clase, se calcula la probabilidad de cada valor de característica, lo que implica recorrer el conjunto de entrenamiento para cada combinación de característica y clase, resultando en una complejidad de  $O(n \cdot d)$  en total [23].
- Inferencia: durante la inferencia, la complejidad es  $O(d \cdot c)$ , donde  $c$  es el número de clases. Para clasificar una instancia nueva, Naive Bayes calcula la probabilidad posterior de cada clase multiplicando las probabilidades de cada característica [24].
- K vecinos más cercanos (K-NN, por sus siglas en inglés), se basa en la idea de que las instancias similares en su clasificación se encuentran cerca unas de otras en el espacio de características de dimensión  $\mathbb{R}^d$ , donde  $d$  es el número de características. Los vecinos se seleccionan con base en una métrica de distancia, generalmente la distancia euclidiana [25].
  - $x \in \mathbb{R}^d$ , vector de características de cada instancia.
  - $K \in \mathbb{N}$ , número de vecinos a considerar para clasificar una instancia.
  - $C \in \{C_1, C_2, \dots, C_n\}$ , conjunto de posibles clases.

La complejidad computacional de K vecinos tiene complejidad de:

- Entrenamiento: el entrenamiento en KNN es trivial en términos de complejidad, ya que no implica ningún proceso de ajuste de parámetros o aprendizaje sobre los datos. La complejidad aquí es solo  $O(1)$  porque el modelo simplemente almacena el conjunto de datos sin realizar operaciones de entrenamiento.
- Inferencia: se tiene una complejidad de  $O(d \cdot n)$ , ya que se calcula la distancia entre la instancia a clasificar y cada datos etiquetado, de  $d$  es el número de características que se necesitan comparar para obtener la distancia.
- Árbol de decisión, es un modelo de aprendizaje supervisado no paramétrico, es decir, que su estructura no está predefinida. De forma más precisa, clasifica instancias mediante una serie de decisiones basadas en los valores de las características del conjunto de entradas, la entrada  $x \in \mathbb{R}^d$ . En este árbol:
  - Cada nodo interno representa una prueba o pregunta sobre una característica de la entrada  $x$ .
  - Cada rama representa el resultado de la prueba.
  - Cada hoja representa una etiqueta de clase  $C \in \{C_1, C_2, \dots, C_n\}$ .

Un árbol de decisión  $T$  se puede ver como una función que busca una predicción o clasificación de los datos de entrada  $X$ . Este método realiza una partición del espacio de características en regiones  $R_i$ , donde el número de regiones depende de la cantidad de clases. Cada región  $R_i$  tiene asignada una predicción  $y_i$ , correspondiente a una de las clases  $C$  (véase la ecuación 2).

$$T(x) = y_i, si \quad x \in \mathbb{R}_i \quad (2)$$

Para poder construir un árbol de decisión, se ocupa un algoritmo que selecciona en cada nivel del árbol un atributo o característica y un umbral que maximiza una medida de ganancia de la información, típicamente basada en la entropía o el índice Gini [26].

La complejidad computacional de los árboles de decisión se divide entre las fases de entrenamiento e inferencia:

- Entrenamiento: la complejidad promedio es  $O(n \cdot d \cdot \log(n))$ , donde  $n$ , es el número de instancias y  $d$ , el número de características, ya que el algoritmo divide recursivamente el conjunto de datos buscando la mejor característica y punto de división en cada nodo. Esto es así ya que para construir el árbol de decisión se necesita comparar dos los datos, lo que tiene una complejidad  $O(n \cdot \log(n))$ , al tratar de como a su vez se consideran todos los atributos por lo que se tiene  $O(n \cdot d \cdot \log(n))$ . En el peor de los casos, la complejidad puede alcanzar  $O(n^2 \cdot d)$  si el árbol no está balanceado [27].
- Inferencia: durante la inferencia, la complejidad es rápida,  $O(\log(n))$  por instancia, ya que solo es necesario recorrer el árbol desde la raíz hasta una hoja.
- Bosques aleatorios, es un método de aprendizaje que combina múltiples árboles de decisión para obtener el resultado. Cada árbol  $T_j$  se entrena con una muestra aleatoria del conjunto de los datos y un subconjunto características  $X \subset \mathbb{R}^d$ . Donde  $j \in \{1, 2, \dots, m\}$ ,  $m$  es el número de árboles en el bosque. Este enfoque busca reducir la varianza y mejorar la generalización del modelo en comparación con un solo árbol de decisión [26], [28].

La complejidad computacional de los bosques aleatorios se divide entre las fases de entrenamiento e inferencia:

- Entrenamiento: al construir  $T$  árboles de decisión, la complejidad promedio para el entrenamiento de un bosque aleatorio es  $O(T \cdot n \cdot d \cdot \log(n))$ . En el peor de los casos, se tiene  $O(n^2 \cdot d)$ .
- Inferencia: se tiene  $O(T \cdot \log(n))$ , ya que necesitamos obtener la clasificación para todos los árboles en el bosque.
- AdaBoost (*Adaptive Boosting*, en inglés), es un método de aprendizaje de conjunto que combina varios clasificadores débiles para formar un clasificador más robusto. Por ejemplo, supongamos que aplicamos un algoritmo de árbol de decisiones en un conjunto de datos:
  - $X \subset \mathbb{R}^d$ , es un conjunto de vectores de características, donde  $d$  es el número de características.

- $C \in \{C_1, C_2, \dots, C_n\}$ , Conjunto de clases.

Este algoritmo podría alcanzar una precisión del 80%. Luego, aplicamos otros métodos, como Naive Bayes, que produce una precisión del 75%, y finalmente regresión lineal, con una precisión del 70%. En lugar de usar estos modelos por separado, AdaBoost combina sus resultados. Lo hace asignando más peso a las instancias mal clasificadas y ajustando los modelos en consecuencia. De este modo, se busca que las predicciones finales se mejoren, para tratar obtener una mayor precisión que cualquier modelo individual [26], [29].

La complejidad computacional de AdaBoost se divide entre las fases de entrenamiento e inferencia:

- Entrenamiento: la complejidad de AdaBoost depende de la cantidad y la complejidad de cada clasificador débil que se utiliza por lo que  $O(\sum_1^k O(\text{Clasificador entrenamiento}_k))$ , donde  $k$  es el número de clasificadores, por ejemplo, si los clasificadores fueran árboles de decisión la complejidad sería  $O(k \cdot n \cdot d \cdot \log(n))$ .
- Inferencia: se tiene  $O(\sum_1^k O(\text{Clasificador inferencia}_k))$ , en el ejemplo de escoger solo árboles de decisión  $O(k \cdot \log(n))$ .

### 2.2.1. Indicadores de desempeño

Existen varios indicadores de desempeño para medir la eficiencia de métodos de AM, la mayoría se pueden definir a partir de la matriz de confusión, que es una tabla que organiza las predicciones de un clasificador binario en cuatro categorías (véase Figura 1), estas regiones son:

- Verdaderos Positivos (VP): instancias donde el modelo predice correctamente la clase positiva.
- Verdaderos Negativos (VN): instancias donde el modelo predice correctamente la clase negativa.
- Falsos Positivos (FP): instancias donde el modelo predice incorrectamente la clase positiva.
- Falsos Negativos (FN): instancias donde el modelo predice incorrectamente la clase negativa.

	Predicción positiva	Predicción negativa
Clase real positiva	Verdaderos positivos (VP)	Falsos Negativos (FN)
Clase real negativa	Falsos positivos (FP)	Verdaderos Negativos (VN)

**Figura 1.** Definición de la matriz de confusión, basada en [30].

La métrica que comúnmente se utiliza en modelos AM (por lo tanto, en AP) es la exactitud (*accuracy* en inglés). Esta métrica se define como las predicciones correctas entre el total de predicciones, de forma más rigurosa véase la ecuación 3 [31]:

$$\text{Exactitud} = \frac{V_p + V_n}{V_p + V_n + F_p + F_n} \quad (3)$$

Otros Indicadores de desempeño que se obtienen a partir de la matriz de confusión y pueden ser relevantes para evaluar a los modelos de AM [31] son:

- Sensibilidad (*recall*, en inglés), que mide cuantos los verdaderos positivos predice correctamente el modelo.
- La especificidad (*specificity*, en inglés), que mide cuántos verdaderos negativos predice correctamente el modelo [31].

### 2.2.2. Validación Cruzada y *K folds*

Validación cruzada, es una técnica para realizar validación en modelos de AM y AP cuando se tienen pequeños conjuntos de datos de manera que los resultados sean más confiables. Esta técnica se basa en dividir el conjunto de datos de entrenamiento y validación varias veces, de forma que el modelo se entrene varias veces con diferentes subconjuntos de entrenamiento y validación. De esta forma se trata de evitar el sobreajuste por tener pocos datos. *K folds*, es una estrategia de validación cruzada en la que se divide el conjunto de datos en *K* partes. El modelo se entrena *K* veces, usando cada vez *K - 1* partes como conjunto de entrenamiento y 1 parte como conjunto de validación [32].

## 2.3. Análisis de múltiples dimensiones

Cuando los datos de una tarea de interés están en un espacio de alta dimensión, es fundamental saber cómo se comportan tanto de manera individual como en conjunto. Surge la necesidad de responder preguntas como: ¿cuáles son las distancias entre los datos?, o ¿cómo se

agrupan o dispersan en el espacio dimensional?. Para responder esto, se pueden proponer diversas técnicas que permiten observar y medir las distancias o relaciones entre los datos.

- La distancia euclidiana, es una medida fundamental para cuantificar la separación entre dos puntos en un espacio geométrico. Para datos que tienen  $n$  dimensiones, la expresión matemática es la ecuación 4:

$$d = \sqrt{(X_p + Y_n)^2 + (X_p + Y_n)^2 + \dots + (X_p + Y_n)^2} \quad (4)$$

Donde  $d$  es la distancia euclidiana,  $X$  e  $Y$  son los dos datos o puntos que tienen dominio  $\mathbb{R}^n$ , donde  $n$  es la cantidad de dimensiones [33].

De esta forma, para analizar múltiples datos a la vez podemos construir una matriz de distancias euclidianas, donde cada elemento  $d_{ij}$  de la matriz representa la distancia euclidiana calculada entre el dato  $i$  y el dato  $j$ .

- La correlación, es una medida estadística que mide la relación lineal entre la magnitud y dirección de dos variables. Es decir, para dos variables aleatorias,  $X$  e  $Y \in \mathbb{R}^n$ , el coeficiente de correlación  $\rho$  se puede definir como se muestra en la ecuación 5:

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma X * \sigma Y} \quad (5)$$

Donde  $Cov(X, Y)$  es la covarianza entre  $X$  e  $Y$ , mientras que  $\sigma X$  y  $\sigma Y$  son las desviaciones estándar de  $X$  e  $Y$ , respectivamente. Esta expresión matemática define a la correlación de Pearson, que es la correlación más popular. El valor de esta correlación oscila entre -1 y 1, donde -1 expresa una correlación negativa máxima, el 0 una ausencia de correlación y el 1 una correlación positiva máxima [34].

- Matriz de correlación, es una matriz simétrica que contienen los coeficientes de correlaciones entre múltiples variables, es decir, para un conjunto  $m$  de variables aleatorias  $\{X_1, X_2, \dots, X_m\}$ , los elementos de la matriz  $R$  se definen como  $\rho(X_i, X_j)$  donde  $i, j$  expresan las posiciones de los renglones y las columnas de la matriz, es decir, las variables aleatorias [35].

La reducción de dimensionalidad, es el proceso de reducir la alta dimensionalidad de los datos a un subespacio de menor dimensión, tratando de conservar la parte más relevante de la información. Esto puede ayudar a comprender mejor la distribución o esencia de los datos y facilitar su análisis. Además, permite reducir el tamaño de los datos de entrada en los modelos estadísticos de predicción, lo que puede hacer que la predicción sea más rápida y precisa. La razón por la que estas técnicas suelen ser efectivas es porque, en muchos casos, los datos pueden describirse en términos de un pequeño número de factores latentes que capturan la mayor parte de su variabilidad. Un ejemplo de esto es el modelado de rostros, donde factores latentes como la iluminación, la posición de la cara, de los ojos, nariz y boca describen las características principales [36], [37]. Algunas técnicas de reducción de dimensionalidad son:

- El Análisis de Componentes Principales (ACP o PCA, por sus siglas en inglés) es una técnica de reducción de dimensionalidad que transforma un conjunto de características en otro conjunto de nuevas características llamadas componentes principales. De manera más

rigurosa, para una matriz de datos  $X \in \mathbb{R}^{m \times n}$ , donde  $m$  es el número de muestras y  $n$  el número de características, ACP encuentra una matriz de proyección  $W \in \mathbb{R}^{m \times k}$  que proyecta  $X$  a un espacio de  $k$  dimensiones. Esto permite representar los datos en un espacio de menor dimensión, donde los componentes principales seleccionados buscan describir de mejor manera la estructura del conjunto de datos [38], [39].

- Escalado Multidimensional (MDS, por sus siglas en inglés), es una técnica que busca representar las similitudes, o en su caso disimilitudes, entre pares de muestras con distancias en un espacio de baja dimensión. Para un conjunto de  $n$  muestras con una matriz de disimilitudes  $D \in \mathbb{R}^{n \times n}$ , MDS busca una configuración de puntos  $Y \in \mathbb{R}^{n \times k}$  en un espacio de baja dimensión que minimiza una función llamada estrés [39].
- Mapeo isométrico (ISOMAP, por sus siglas en inglés), es un método de reducción de dimensionalidad no lineal que busca conservar las distancias geodésicas entre los datos en un espacio de alta dimensión [39]. Conjunto de muestras  $X \in \mathbb{R}^{m \times n}$  y matriz de distancias geodésicas  $D_g \in \mathbb{R}^{m \times m}$ . El método puede definirse en 3 pasos:
  1. Construir un grafo de vecinos más cercanos.
  2. Calcular las distancias geodésicas entre todos los pares de datos.
  3. Realizar el *MDS* clásico a la matriz de distancias geodésicas [39].

## 2.4. Introducción al aprendizaje profundo (AP)

Los métodos de aprendizaje profundo (AP) son un subconjunto de los métodos de aprendizaje máquina, que se basan en utilizar modelos complejos de redes neuronales artificiales. Estos métodos buscan extraer características y predecir salidas a partir de los datos de entrada, dentro de un solo modelo [40].

### 2.4.1. El perceptrón simple

El perceptrón simple, es un método que trata de simular el supuesto comportamiento de una sola neurona para realizar la tarea de clasificación, este método trata de modelar una función  $f: \mathbb{R}^n \rightarrow \{-1, +1\}$ . De forma más clara se define la ecuación 6:

$$f(x) = \text{func}(w \cdot x + b) \quad (6)$$

Dónde  $x \in \mathbb{R}^n$  es un vector de los datos de entrada,  $w \in \mathbb{R}^n$  es un vector de pesos (un peso para cada uno de los datos  $x$ ),  $b \in \mathbb{R}$  es un sesgo, y la *func* que es una función que tiene como salida 1 o -1.

La forma en que el perceptrón aprende es ajustando iterativamente los pesos y el sesgo para tratar de minimizar errores de predicción dado un conjunto de entrenamiento [40].

La complejidad en el perceptrón simple se divide en dos partes:

- Entrenamiento: la complejidad se puede definir como  $O(I \cdot N \cdot d)$ , cada iteración  $I$  se tiene una complejidad de  $O(N \cdot d)$ ,  $N$  es la cantidad de datos y  $d$  es la dimensión de una instancia.
- Inferencia: el cálculo de la salida para una muestra requiere el producto punto entre los pesos y la entrada, con una complejidad de  $O(d)$ .

### 2.4.2. El descenso del gradiente (DG)

El descenso del gradiente (DG) es un algoritmo o método de optimización iterativo que busca minimizar una función objetivo  $J(\theta): \mathbb{R}^d \rightarrow \mathbb{R}$ , donde  $\theta \in \mathbb{R}^d$ , representa los parámetros del modelo y  $d$  es el número de parámetros. Este método puede explicarse con un ejemplo, imaginemos que estamos en medio de la noche en una montaña, sin poder ver nada a nuestro alrededor, y deseamos llegar al punto más bajo para pedir ayuda. El descenso del gradiente es similar a dar pequeños pasos en la dirección que parece descender más, con la esperanza de alcanzar el lugar más bajo. De forma más rigurosa se define la ecuación 7:

$$\theta(t + 1) = \theta(t) - \eta \nabla J(\theta(t)) \quad (7)$$

Donde  $\theta(t)$  es el valor de los parámetros en la iteración  $t$ ,  $\eta \in \mathbb{R}^+$  es la tasa de aprendizaje (debe ser mayor a 0),  $\nabla J(\theta(t))$  es el gradiente de la función objetivo con respecto a  $\theta$  en el punto  $\theta(t)$ . El método del gradiente descendente actualiza los parámetros de forma iterativa en la dirección opuesta al gradiente, para así tratar de encontrar los mínimos de la función  $J(\theta)$  [40].

La complejidad del descenso del gradiente es de  $O(I \cdot N \cdot d)$ , en cada iteración, el algoritmo debe calcular el gradiente de la función de pérdida respecto a cada parámetro en  $\theta$ . Esto implica pasar por todos los  $N$  datos y realizar operaciones sobre cada uno de los  $d$  parámetros, lo que resulta en una complejidad  $O(N \cdot d)$ .

### 2.4.3. Redes Neuronales Profundas (RNP)

Las Redes Neuronales Profundas (RNP), son métodos de aprendizaje profundo que se componen por múltiples capas de neuronas buscando un procesamiento no lineal de los datos. De forma más rigurosa, una RNP puede verse como un conjunto de funciones anidadas (véase la ecuación 8).

$$f(x) = f_L(f_{L-1}(\dots (f_2(f_1(x))))) \quad (8)$$

Dónde,  $x \in \mathbb{R}^d$  es el vector de los datos de entrada,  $L \in \mathbb{N}$  el número de capas y las  $f_i$  son las funciones de activación para la  $i$ -ésima capa. Cada función  $f_i$  típicamente se define en la ecuación 9:

$$f_i(x) = \sigma(W_i x + b_i) \quad (9)$$

Dónde,  $W_i \in \mathbb{R}^{n_i \times n_{i-1}}$  es la matriz de pesos en la capa  $i$ , donde  $n_i$  es el número de neuronas en la capa  $i$ ,  $b_i \in \mathbb{R}^{n_i}$  son los sesgos para la capa  $i$ ,  $\sigma$  es una función de activación no lineal como *ReLU*, *sigmoid*, *tanh*, etc.

Para modificar los modelos de redes neuronales profundas (RNP) y obtener los resultados deseados, es necesario optimizar los parámetros de la red (pesos y sesgos por capa) de forma que se minimice

una función de pérdida sobre el conjunto de entrenamiento. Esto se realiza típicamente mediante algoritmos de retropropagación y descenso del gradiente [40].

La complejidad para el RNP se divide en dos partes:

- Entrenamiento: la complejidad es de  $O(I \cdot N \cdot L \cdot n^2)$ , por cada iteración  $I$ , la capa  $L$  tiene un número de neuronas  $n$ , tal que para:
  - Propagación hacia adelante (*feedforward*): la complejidad por capa es  $O(n^2)$  si cada neurona se conecta a todas en la capa siguiente. Así, la complejidad total por iteración de la propagación hacia adelante es  $O(L \cdot n^2)$ .
  - Propagación hacia atrás (*backpropagation*): el cálculo del gradiente y la actualización de los pesos también requiere  $O(L \cdot n^2)$  por iteración.
  - Como se entrena para todas los datos  $N$ , se tiene  $O(N \cdot L \cdot n^2)$ .
- Inferencia: solo se realiza una propagación hacia adelante a través de todas las capas  $O(L \cdot n^2)$ .

#### 2.4.4. Redes Neuronales Convolucionales (RNC)

Las redes neuronales convolucionales (RNC), son un tipo de RNP que tratan de extraer características de datos con una estructura de cuadrícula, como una imagen,  $x \in \mathbb{R}^{h \times w \times c}$  donde  $h$  es la altura,  $w$  es el ancho y  $c$  son los canales de la imagen. Utilizan el concepto de convolución junto con filtros o *kernels*  $\in \mathbb{R}^{f_h \times f_w \times c}$  para extraer características importantes de las imágenes, donde  $f_h \times f_w$  es el tamaño del kernel, [41]. Las RNC suelen tener los siguientes tipos de capas neuronales:

1. Capas convolucionales. Estas realizan la operación de convolución de un *kernels* sobre el dato de entrada, y tiene como salida un conjunto de características, también llamado mapas de características. Las capas convolucionales reducen secuencialmente la resolución espacial de las imágenes, y amplían la profundidad de sus mapas de características.
2. Capas completamente conectadas. Estas capas son capas estándar, donde las salidas de una capa son las entradas para todas las neuronas de la siguiente capa [41].

La complejidad para el RNC se divide en dos partes:

- Entrenamiento: en una RNC con  $L$  capas, la complejidad depende en de las capas convolucionales:
  - Para una capa convolucional con  $f$  filtros de tamaño  $k \times k$  aplicada a una entrada de tamaño  $m \times m$ . La complejidad es de  $O(m^2 \cdot k^2 \cdot f)$ .
  - Al añadir las capas completamente conectadas, la complejidad total puede aproximarse a  $O(I \cdot N \cdot (L_{conv} \cdot m^2 \cdot k^2 \cdot f + L_{fc} \cdot n^2))$ , donde  $L_{conv}$  son las capas convolucionales y  $L_{fc}$  son las capas totalmente conectadas.

- Inferencia: se compone de una sola propagación hacia adelante, por lo que una complejidad aproximada es  $O(m^2 \cdot k^2 \cdot f + L_{fc} \cdot n^2)$ .

Algunas de las arquitecturas populares de RCN son:

- VGG19, es una arquitectura de Red Neuronal Convolutiva (RCN) presentada en el artículo “*Very Deep Convolutional Networks for Large-Scale Image Recognition*” [42]. Es una de las variantes de la familia VGG, que se caracteriza por su simplicidad y profundidad [43].
- Las ResNets (como ResNet18 y ResNet101), son un tipo de arquitectura de RCN diseñada específicamente para el reconocimiento de imágenes. Abordan el problema del gradiente que se desvanece (*vanishing gradient*), que dificulta el entrenamiento de redes neuronales muy profundas [43]. Este enfoque fue presentado por primera vez en el artículo “*Deep Residual Learning for Image Recognition*” [44].
- MobileNet v3, es otra red neuronal convolutiva (RCN) presentada en el artículo “*Efficient Convolutional Neural Networks for Mobile Vision Applications*” [45]. Prioriza dos aspectos clave: eficiencia y precisión. Utiliza convoluciones separables en profundidad para reducir el número de parámetros en comparación con modelos de aprendizaje profundo similares, lo que la convierte en una red neuronal profunda ligera.

#### 2.4.5. Transferencia de aprendizaje y *fine-tuning*

La transferencia de aprendizaje es una técnica que consiste en reutilizar un modelo previamente entrenado para una tarea base, como la clasificación de imágenes en el conjunto *ImageNet* [46], como punto de partida para el entrenamiento en una tarea específica. La idea es que el modelo ya ha aprendido a reconocer características generales, como formas y texturas de objetos comunes (carros, personas, animales, etc.), que luego se pueden aplicar para identificar características en otro contexto, como distinguir enfermedades mediante sus patrones de textura o forma [47], [48].

El *fine-tuning*, o ajuste fino, es una técnica de transferencia de aprendizaje que se aplica en modelos de aprendizaje automático. Esta técnica permite adaptar un modelo preentrenado a una nueva tarea, utilizando dos esquemas posibles: uno en el que se entrena todo el modelo, ajustando completamente los pesos, y otro en el que solo se ajusta cierta parte del modelo, mientras los pesos del resto permanecen fijos. Esto permite que el modelo conserve el conocimiento general adquirido en el preentrenamiento y, al mismo tiempo, se adapte mejor a los datos específicos de la tarea objetivo. [47], [48].

## 2.5. Aumento de datos

El aprendizaje profundo ha tenido un auge en la última década ayudando en diferentes tareas, una de ellas es la clasificación y análisis de imágenes. Sin embargo, existen algunas complicaciones, como el sobreajuste, que parece resolverse cuando existen grandes cantidades de imágenes. El sobreajuste (*overfitting* en inglés) es cuando una RNP aprende perfectamente los datos de entrenamiento, pero no puede predecir datos que no ha visto, es decir, no generaliza el conocimiento. No siempre se tiene a la disposición imágenes suficientes para poder entrenar a los métodos de aprendizaje profundo. Por lo que existen varias estrategias para encontrar una solución a este problema, por ejemplo: *dropout*, normalización por lotes, aprendizaje por transferencia, etc. Sin embargo, existe una estrategia que aborda el problema desde los cimientos (los datos de entrenamiento) que es el aumento de datos.

El aumento de datos se refiere a generar nuevos datos o imágenes a partir del conjunto de entrenamiento para sobre muestrear la información del conjunto de entrenamiento. Se puede proponer una clasificación de estrategias de aumento de datos partiendo de dos grandes clases:

Las técnicas basadas en manipulaciones básicas (también conocidas como aumento de datos tradicional) incluyen:

- Transformaciones geométricas, que aplican transformaciones afines o proyectivas a las imágenes.
- Transformaciones del espacio de color, que modifican los canales de color de las imágenes.
- Combinación de imágenes, que fusiona dos imágenes en una sola, por ejemplo, utilizando parches.
- Añadir ruido a la imagen.
- Uso filtros que modifican la imagen.

Por otro lado, las técnicas basadas en aprendizaje profundo incluyen:

- Entrenamiento con ataques adversarios, que modifica las imágenes de manera controlada para desafiar la robustez del modelo.
- Aumento de imágenes mediante redes generativas adversarias (GAN), que crean nuevas imágenes.
- Transferencia de estilo neuronal (TEN), que son las técnicas en las que se centra el presente trabajo.
- Aumento del espacio de características, entre otras.

Una técnica de aprendizaje profundo particularmente interesante es el aumento del espacio de características. Las redes neuronales, al procesar datos, pueden transformar una entrada de alta dimensionalidad en una salida de menor dimensionalidad, como es el caso de la clasificación de imágenes en distintas clases. En este proceso, las redes generan representaciones intermedias entre la entrada y la salida, conocidas como características, que en conjunto forman el espacio de

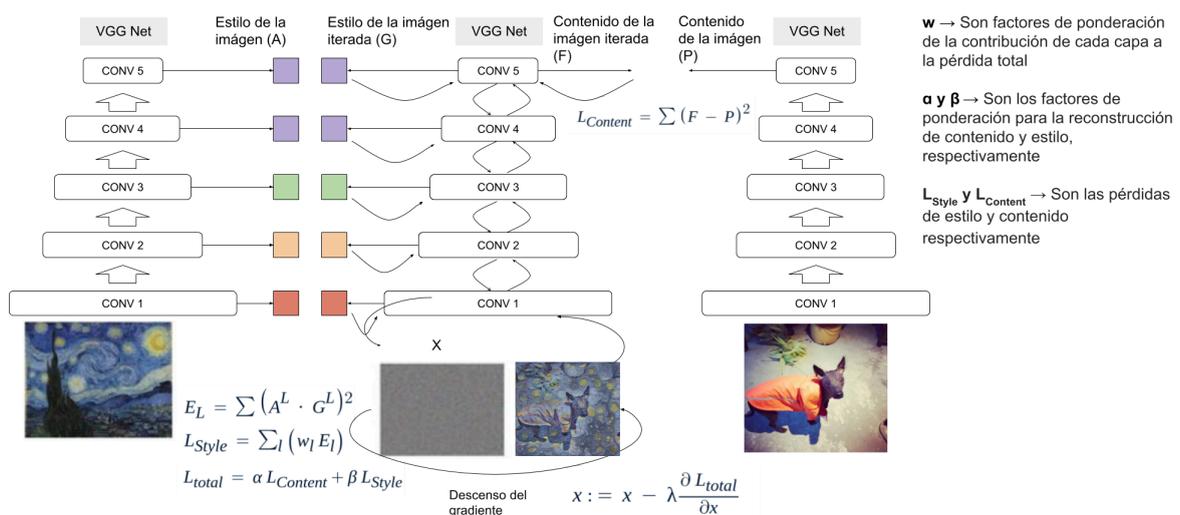
características. Este espacio puede ampliarse mediante la adición de ruido, la interpolación o la extrapolación de estas características, lo que puede contribuir a mejorar la generalización del modelo al aumentar la diversidad de las características con las que se entrena.

Cada una de las estrategias antes mencionadas tiene sus ventajas y desventajas. Cada método de aumento tradicional busca resolver el problema de no contar con suficientes imágenes, al tratar que la red sea coherente ante las distorsiones o los cambios de apariencia de las imágenes. En cambio, las estrategias de aumento de datos con aprendizaje profundo buscan aumentar la cantidad de imágenes, variando la naturaleza de las mismas. Por ello, algunos trabajos proponen usar ambos.

Otra cosa que hay que destacar, es la diferencia del aumento de datos con el preprocesamiento, ya que existen estrategias de aumento de datos que pueden usarse como preprocesamiento. El aumento de datos trata de aliviar los efectos del sobreajuste al contar con pocas imágenes, ya sea al hacer coherente al modelo ante variaciones en las imágenes o al sintetizar nuevas imágenes. Mientras, el preprocesamiento busca reducir el espacio de los datos de entrada a un dominio de las imágenes que esté más acotado [1].

## 2.6. Transferencia de Estilo Neuronal (TEN)

La transferencia de estilo neuronal (TEN) ha emergido como un campo de investigación atractivo en el ámbito de la inteligencia artificial, se basa en transferir el estilo de una imagen  $I_s$  (como una pintura) a otra  $I_c$  (como una fotografía),  $I_s, I_c \in \mathbb{R}^{h \times w \times c}$  donde  $h$  es la altura,  $w$  es el ancho y  $c$  son los canales de la imagen. Esta técnica se basa en utilizar una red neuronal VGG para extraer el estilo de una imagen y el contenido de otra. Después, a partir de una imagen inicial y con el algoritmo del descenso por gradiente, se traslada el estilo al contenido minimizando la pérdida total [4], (véase la Figura 2).



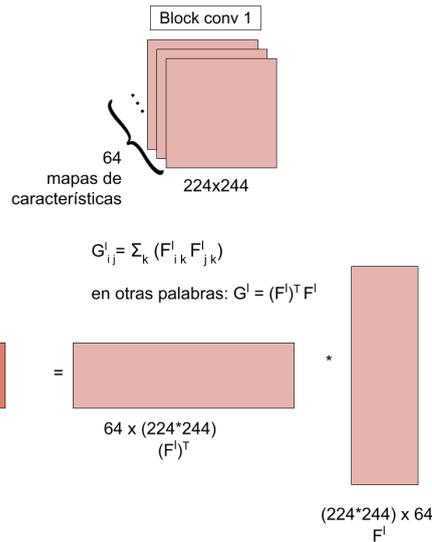
**Figura 2.** Proceso de transferencia de estilo neuronal TEN, basada en [49]. **Imagen de estilo:** La noche estrellada de Vincent Van Gogh. **Imagen de contenido:** fotografía de mi perro "Ikal" comiendo elotes. Imagen inicial  $x$  sobre la que se realizan las iteraciones.

La complejidad para TEN puede verse como 3 pasos:

- Extracción de características, para cada capa convolucional de la VGG 19 implica una complejidad de  $O(H \cdot W \cdot D_l)$ , donde  $H$  y  $W$  es la dimensión de la imagen de la entrada y  $D_l$  es la cantidad de filtros de la la capa  $l$ . Esto porque los  $D_l$  recorren en  $H \cdot W$  pasos a la entrada.
- Cálculo de la matriz *Gram*, es de una complejidad de  $O(H \cdot W \cdot D_l^2)$ , ya que se realiza el producto entre todos los mapas de características en esa capa.
- Descenso del gradiente, es de una complejidad de  $O(I \cdot (H \cdot W \cdot D_l + H \cdot W \cdot D_l^2))$ , ya que el descenso por gradiente hace  $I$  iteraciones, para optimizar la pérdida total que se compone de las sumas de la pérdida de contenido y la pérdida del estilo.
  - La pérdida del contenido tiene una complejidad de  $O(H \cdot W \cdot D_l)$ , ya que se tiene que obtener los mapas de características de cada capa. Si bien, la pérdida se obtiene con el error cuadrático entre la imagen de contenido  $I_c$  y la imagen inicial  $x$ , para big  $O$ , solo es relevante la parte de la extracción de características ya que es la parte que más se tarda.
  - La pérdida del estilo tiene una complejidad de  $O(H \cdot W \cdot D_l^2)$  ya que se obtiene las matrices *Gram* para la imagen de estilo  $I_s$  y la imagen inicial  $x$ , para después obtener el error cuadrático. Sin embargo, para la complejidad big  $O$  solo es relevante la parte de obtener la matriz *Gram*.

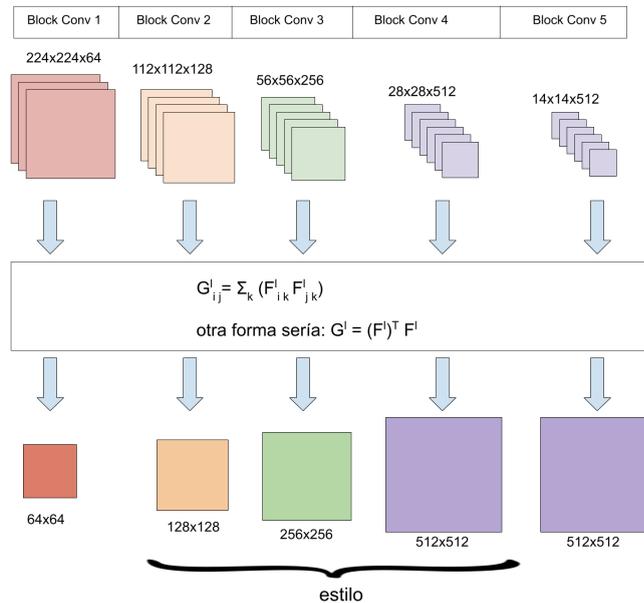
### 2.6.1. Noción del estilo en la transferencia de estilo Neuronal

El estilo se define utilizando la matriz *Gram* de las características extraídas de los bloques convolucionales de una red neuronal convolucional VGG. Esta matriz *Gram* captura las correlaciones entre mapas de características. Esta matriz contiene información sobre cómo se correlacionan entre sí las diferentes características de una capa. Al obtener las matrices *Gram* de las capas de la red, se obtiene una representación completa del estilo de una imagen [4], donde  $F^l \in \mathbb{R}^{H \times W \times D_l}$ ,  $H$  y  $W$  son la altura y ancho de los mapas de características y  $D$  es el número de filtros para la capa  $l$  (véase la Figura 3).



**Figura 3.** Obtención de una matriz *Gram*. Aquí,  $k$  representa la cantidad de filtros en la capa  $l$ .

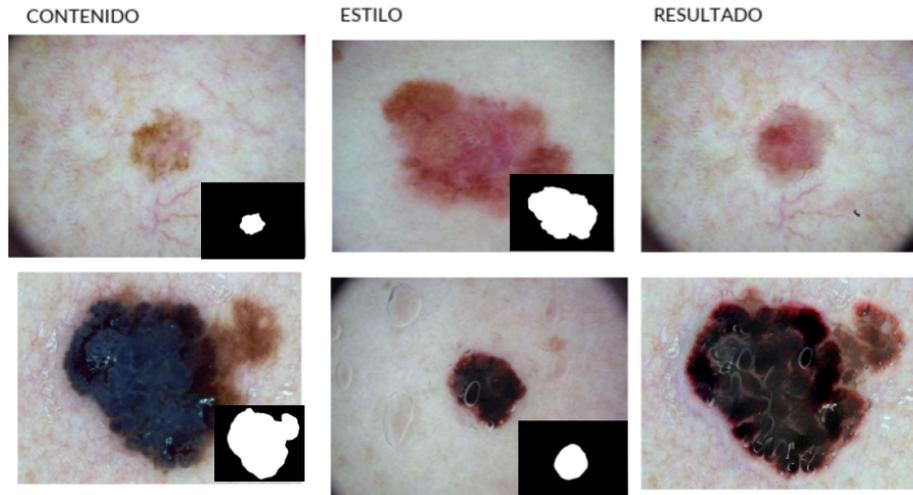
El estilo se define como un conjunto de matrices *Gram* de diferente tamaño, el contenido como las características de las últimas capas del modelo, Gram para una capa o  $G^l \in \mathbb{R}^{D_l \times D_l}$  donde  $D$  es el número de filtros para la capa  $l$  (véase la Figura 4).



**Figura 4.** Obtención del estilo de una imagen. El estilo es un conjunto de matrices Gram. Aquí,  $k$  representa la cantidad de filtros en la capa  $l$ .

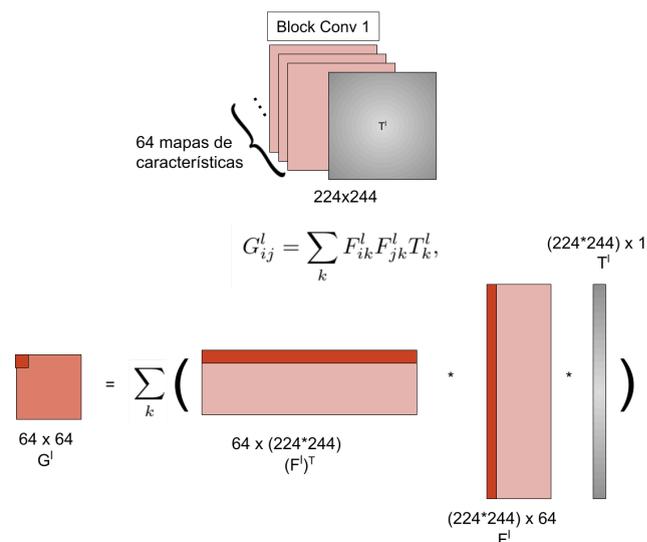
### 2.6.2. Transferencia de Estilo Neuronal (TEN) con control espacial

La transferencia de estilo neuronal con control espacial busca trasladar el estilo de una determinada parte de la imagen de estilo a una parte determinada de la imagen de contenido, utilizando máscaras sobre las zonas de interés. En otras palabras, con esta técnica podemos aislar el estilo de una parte de la imagen para trasladarla a otra parte de la imagen de contenido (véase la Figura 5).



**Figura 5.** Ejemplo de transferencia de estilo con control espacial en imágenes de lesiones de la piel.

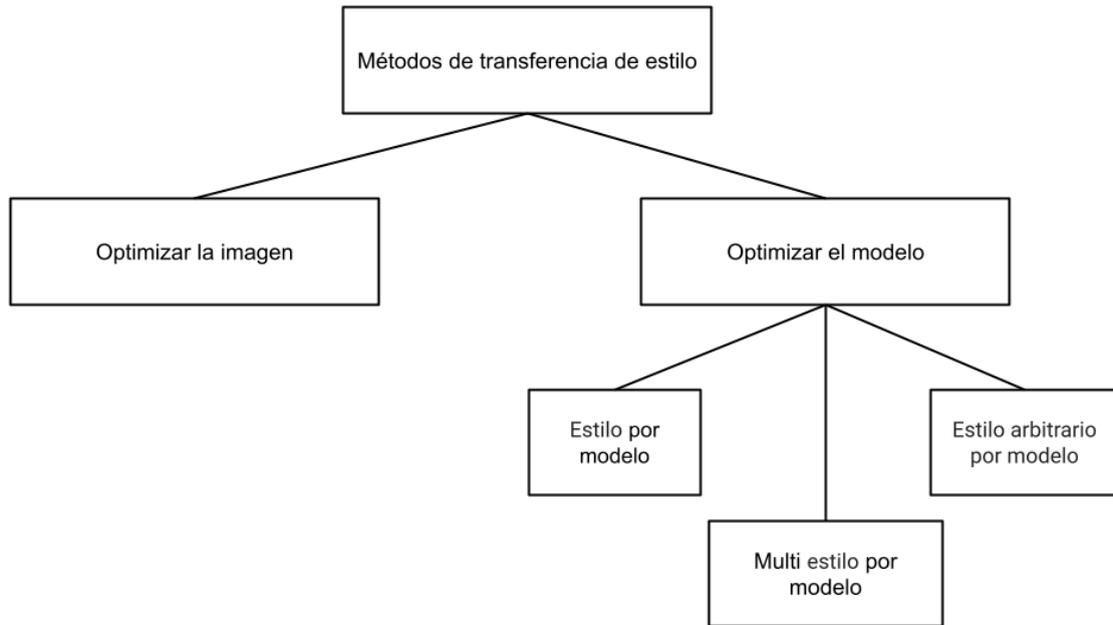
Para realizar un control espacial en la transferencia de estilo (obtener el estilo de una parte específica de la imagen), los mapas de características de la capa convolucional se multiplican por la máscara correspondiente. Esta multiplicación por la máscara asegura que solo se conserven las características de las regiones de interés especificadas por la máscara [50], (véase la Figura 6).



**Figura 6.** Obtención de una matriz *Gram* con control espacial. Aquí,  $k$  representa la cantidad de filtros en la capa  $l$ .

### 2.6.3. Otras técnicas TEN

Después de que Gatys *et al.* propusieron el método de Transferencia de Estilo Neuronal (TEN), surgieron diversos trabajos que intentaron resolver problemas asociados a la técnica o mejorar los resultados obtenidos. Al haber tantos trabajos sobre TEN, en el artículo de *LapStyle* de Tianwei Lin *et al.* [51], se propone una clasificación de métodos de transferencia de estilo. Los basados en optimizar la imagen, por ejemplo TEN con control espacial [6] y *STROTSS* [52], y los basados en optimizar el modelo [51], [53], (véase las Figura 7).



**Figura 7.** Clasificación: métodos de transferencia de estilo.

Un ejemplo de método basado en optimizar la imagen es el de Kolkin *et al.* del 2019 donde presentaron STROTSS, *Style Transfer by Relaxed Optimal Transport and Self-Similarity* [52], este enfoque de TEN trata de mejorar el control sobre el estilo y contenido en la transferencia de estilo, al modificar la función objetivo para obtener resultados más precisos y personalizados.

A su vez, un método basado en optimizar el modelo es *LapStyle* de Tianwei Lin *et al.* del 2021 [51]. La idea detrás de este algoritmo es realizar un borrador con una red llamada *Drafting Network*, y luego enfocarse en los detalles con otra red llamada *Revisión Network* y una pirámide laplaciana de 2 niveles. Este método ocupa la idea original de Gatys *et al.* 2015, y de otros métodos de TEN como *STROTSS*, Kolkin *et al.*, 2019 [5], y *AdaIn*, Xun Huang con Serge Belongie, 2017 [53].

## 2.7. Otros métodos de generación de imágenes.

En los últimos años, la generación de imágenes ha surgido como una posible solución para abordar la limitación de conjuntos de imágenes pequeños en modelos de aprendizaje profundo. Este enfoque ha ganado particular relevancia en aplicaciones médicas, donde la cantidad de datos disponibles para entrenar modelos suele ser poca [4], [54]. Entre los métodos más utilizados se encuentran las Redes Generativas Adversarias (GAN) y, más recientemente, los modelos de difusión.

Las Redes Generativas Adversarias (GAN), introducidas por Ian Goodfellow en 2014 [13], son una arquitectura de aprendizaje profundo que utiliza dos redes neuronales: la generadora y la discriminadora. Ambas redes compiten entre sí; la generadora tiene como objetivo crear imágenes que sean lo más realistas posible, mientras que la discriminadora se entrena para distinguir entre imágenes

reales y generadas. Este proceso de competencia ayuda a que la red generadora mejore sus resultados [54], [55].

Por otro lado, los modelos de difusión, como *Stable Diffusion* en 2021 [56], son un enfoque diferente de generación de imágenes. Estos modelos funcionan añadiendo ruido a las imágenes de manera progresiva y luego invirtiendo ese proceso, eliminando el ruido paso a paso. En otras palabras, los modelos de difusión pueden crear imágenes a partir de ruido puro al aprender a revertir ese proceso de agregar ruido [54], [55].

Tanto las GAN como los modelos de difusión enfrentan ciertos desafíos cuando se utilizan para el aumento de datos. Uno de los principales problemas es que el modelo "memorice" los datos del conjunto de entrenamiento, generando imágenes demasiado similares o incluso idénticas a las originales. Esto reduce la efectividad del aumento de datos, ya que los modelos que utilizan estas imágenes se pueden sobreentrenar [54], [55].

Si bien el objetivo principal de este trabajo no es comparar las GAN, los modelos de difusión y los métodos de transferencia de estilo neuronal (TEN), hay características por las que se eligió este último método. Los métodos TEN, basados en la optimización de imágenes, producen resultados más interpretables, ya que permiten identificar con mayor claridad de dónde provienen los elementos de contenido y estilo en la imagen generada. Esto facilita un análisis más preciso al crear imágenes a partir de diferentes diagnósticos o clases. Además, los métodos TEN ofrecen un mayor control sobre las características espaciales de las imágenes, lo que resulta especialmente útil en aplicaciones médicas, donde modificar solo ciertas áreas específicas de la imagen puede crear variaciones solo en las lesiones.

### 3. Metodología

En este capítulo se presentan los experimentos realizados para analizar los estilos neuronales de imágenes médicas y entrenar modelos de clasificación basados en Transferencia de Estilo Neuronal (TEN). Se utilizaron dos bases de datos de imágenes médicas, utilizando estas, se llevaron a cabo dos experimentos principales:

#### 1. Análisis de estilos neuronales:

- Análisis 1: estilos con y sin control espacial. Se extraen los estilos neuronales utilizando las características extraídas de la red preentrenada VGG 19, junto con matrices *Gram*, tanto para obtener el estilo de toda la imagen, como para obtener solo el estilo de la lesión (con control espacial). Después, se analizan los estilos aplicando técnicas de reducción de dimensionalidad (ACP, ISOMAP, MDS) para visualizar los datos, junto con matrices de correlación y distancias euclidianas.
- Análisis 2: interpolación de estilos. Se realiza una interpolación de estilos con control espacial entre imágenes del mismo diagnóstico para generar nuevos conjuntos de estilos. Estos conjuntos se analizan utilizando las mismas técnicas aplicadas en el análisis 1.

#### 2. Clasificación de diagnósticos utilizando TEN:

- Experimento de aprendizaje máquina. Se evalúa el rendimiento de técnicas de aprendizaje automático utilizando conjuntos de estudio basados en estilos neuronales como técnica de extracción de características. Estos resultados se comparan con aquellos obtenidos usando imágenes aplanadas, para evaluar la efectividad de los estilos neuronales en la mejora del rendimiento en tareas de clasificación.
- Experimento de aprendizaje profundo. Se evalúa el rendimiento de modelos de aprendizaje profundo preentrenados con el conjunto de datos *ImageNet* [46] y posteriormente entrenados con distintos conjuntos de estudio. Estos conjuntos incluyen imágenes originales, imágenes generadas mediante aumento de datos tradicional, imágenes generadas mediante TEN con control espacial, y conjuntos de imágenes generadas mediante TEN con interpolación de estilos. El rendimiento de los modelos se mide utilizando un conjunto de prueba independiente.

A continuación, se describen los datos utilizados, las técnicas aplicadas y los experimentos realizados.

### 3.1. Bases de datos utilizadas

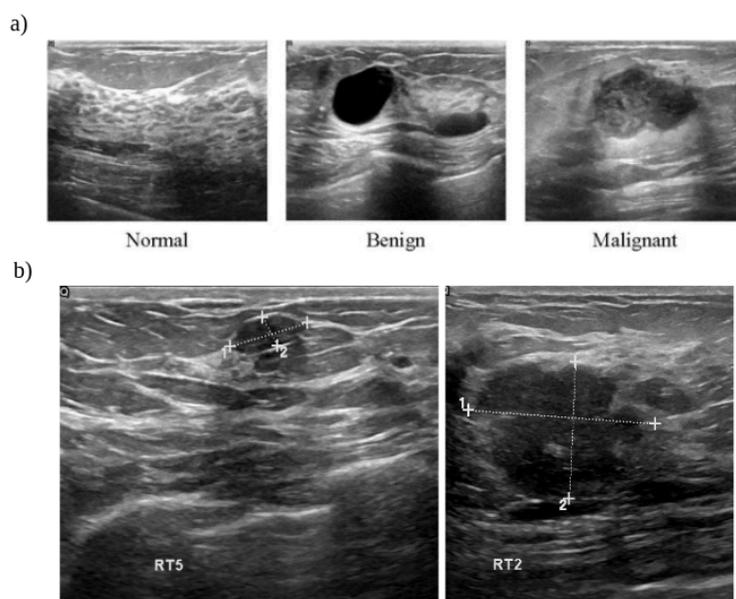
Para llevar a cabo los experimentos propuestos en este trabajo, se emplearon dos bases de datos: una centrada en imágenes de ultrasonido de cáncer de mama y otra a imágenes de lesiones de la piel. Estas bases de datos fueron seleccionadas por su diversidad en diagnósticos evaluados por expertos y por contar con máscaras que identifican la lesión, lo cual es necesario para aplicar las estrategias de TEN propuestas. En ambas bases de datos, se realizaron pasos adicionales de preprocesamiento para asegurar la calidad de las imágenes utilizadas en los experimentos y facilitar la generación de nuevos conjuntos mediante técnicas TEN. A continuación, se detallan las características principales de cada una de las bases de datos, el preprocesamiento realizado, y la creación de los diferentes conjuntos de estudio utilizados en el entrenamiento de los experimentos de clasificación.

#### 3.1.1. Breast cancer

La base de datos *Breast cancer* [57] consiste en imágenes de ultrasonido de mama de mujeres entre 25 y 75 años de edad. Hay un total de 600 pacientes mujeres. Las imágenes están en formato PNG y se clasifican en tres clases: normal, benigno y maligno (véase la Figura 8), cada una con su máscara correspondiente. El número de imágenes en cada clase se muestra en la Tabla 1. El preprocesamiento realizado en [57] incluyó: eliminar imágenes duplicadas, revisar y corregir anotaciones incorrectas, convertir las imágenes a formato PNG, recortar las imágenes y una revisión por radiólogos. Finalmente, buscando las mejores imágenes, se eliminaron las que tuvieran artefactos como anotaciones o texto sobre la lesión, ver Figura 8. Para la cantidad final de imágenes, véase la Tabla 2.

<b>Diagnóstico</b>	<b>No. de imágenes</b>
Benigno	487
Maligno	210
Normal	133

**Tabla 1.** Resumen de la base *Breast cancer*.



**Figura 8.** Ejemplos de imágenes de cáncer de mama. a) Ejemplo de imágenes por diagnóstico del conjunto de datos *Breast Ultrasound Images Dataset*. b) Ejemplo de imágenes con artefactos (anotaciones, letras, etc.).

Diagnóstico	No. de imágenes
Benigno	341
Maligno	182
Normal	133

**Tabla 2.** *Breast Ultrasound Images Dataset* después del preprocesamiento.

### 3.1.2. Base de datos HAM 10K

La base de datos *HAM10K* [58] (*Human Against Machine with 10000 training images*) es una colección de imágenes de lesiones de la piel. Contiene 10,015 imágenes, con sus respectivas máscaras para 7 diagnósticos: (AKIEC) *Queratosis actínica*, (BCC) *Carcinoma de células basales*, (BKL) *Lesiones queratósicas benignas*, (DF) *Dermatofibroma*, (MEL) *Melanoma*, (NV) *Nevos melanocíticos* y (VASC) *Lesiones vasculares* (véase Tabla 3).

<b>Diagnóstico</b>	<b>No. de imágenes</b>
<b>Queratosis actínica (akiec)</b>	327
<b>Carcinoma de células basales (bcc)</b>	514
<b>Lesiones queratósicas benignas (bkl)</b>	1,099
<b>Dermatofibroma (df)</b>	115
<b>Melanoma (mel)</b>	1,113
<b>Nevos melanocíticos (Nv)</b>	6,705
<b>Lesiones vasculares (Vas)</b>	142

**Tabla 3.** Resumen de *HAM 10K*.

Para fines de simular un conjunto de datos pequeño, se descarta el diagnóstico de nevus por su gran cantidad de datos y su semejanza con las lesiones de melanoma. Como preprocesamiento, se eliminan las imágenes con artefactos como burbujas, bello, etc., y se normaliza el color [59] como lo hacen en Yutong Li, *et al.* 2021 [10], para la cantidad final de imágenes véase la Tabla 4.

<b>Diagnóstico</b>	<b>No. de imágenes</b>
Queratosis actínica (akiec)	320
Carcinoma de células basales (bcc)	497
Lesiones queratósicas benignas (bkl)	1,073
Dermatofibroma (df)	112
Melanoma (mel)	1,110
Lesiones vasculares (vasc)	138

**Tabla 4.** *HAM 10K* después del preprocesamiento.

### 3.1.3. Conjuntos de estudio

Para cada una de las bases de datos descritas anteriormente, se definieron los conjuntos de estudio que se muestran en la Tabla 5. El conjunto de imágenes originales (CO), que sirve como base para los demás conjuntos, se obtuvo seleccionando 100 imágenes de forma aleatoria por diagnóstico. A partir del CO, se generaron los conjuntos CO 300 y CO 700 aplicando aumento de datos

tradicional. Los conjuntos CO aplanados y CO aplanados con máscara se obtuvieron al redimensionar las imágenes del conjunto CO a un vector.

Los demás conjuntos de imágenes se generaron a partir de CO y técnicas de Transferencia de Estilo Neuronal (TEN). Estos se dividen en dos categorías principales: aquellos que son de imágenes y aquellos que son de estilos neuronales extraídos.

Conjuntos de imágenes, estos conjuntos incluyen variaciones generadas a partir de las imágenes originales (CO) mediante técnicas de TEN. Son los siguientes:

- CG: Conjunto de imágenes generadas mediante TEN, aplicando distintos estilos a las imágenes del conjunto CO que provienen de imágenes con el mismo diagnóstico.
- CGI: Conjunto de imágenes generadas con estilos interpolados del mismo diagnóstico.
- CO + CG + CGI: Conjunto combinado que integra las imágenes originales (CO) junto con las imágenes generadas mediante TEN (CG y CGI).

Conjuntos de estilos neuronales, estos conjuntos están compuestos por los estilos extraídos de las imágenes, en lugar de las imágenes en sí, y son los siguientes:

- CEO: Conjunto de estilos extraídos de las imágenes originales de CO.
- CEI: Conjunto de estilos interpolados entre estilos del mismo diagnóstico.
- CEOI: Conjunto combinado que agrupa los estilos de CO, CEO, y CGI.

En lo que respecta a los conjuntos para el análisis, el conjunto de análisis (CA) se obtuvo de forma aleatoria e independiente del conjunto de imágenes originales (CO).

<b>Conjunto de estudio</b>	<b>No. de imágenes por diagnóstico</b>	<b>Experimento</b>
<b>Conjunto de análisis de estilos CA</b>	100	Análisis 1 y 2
<b>Conjunto de imágenes originales CO</b>	100	Aprendizaje profundo
<b>Conjunto de imágenes originales con aumento de datos tradicional CO 300</b>	300	Aprendizaje profundo
<b>Conjunto de imágenes originales con aumento de datos tradicional CO 700</b>	700	Aprendizaje profundo
<b>Conjunto de imágenes generadas con TEN CG</b>	300	Aprendizaje profundo
<b>Conjunto de imágenes generadas con estilos interpolados CGI</b>	300	Aprendizaje profundo
<b>CO + CG + CGI</b>	700	Aprendizaje profundo
<b>Conjunto de imágenes originales aplanados sin máscara</b>	100	Aprendizaje máquina
<b>Conjunto de imágenes originales aplanados con máscara</b>	100	Aprendizaje máquina
<b>Conjunto de estilos de las imágenes originales CEO</b>	100	Aprendizaje máquina
<b>Conjunto de estilos interpolados CEI</b>	300	Aprendizaje máquina
<b>CEOI</b>	400	Aprendizaje máquina

**Tabla 5.** Resumen de conjuntos de estudio.

El entrenamiento de los modelos se realiza usando K Folds, por lo que cada conjunto anterior (excepto CA) fue usado para el entrenamiento y validación de los modelos. Para el experimento de aprendizaje profundo AP se realiza un  $K = 5$ , mientras que para el experimento de aprendizaje máquina AM se utiliza una  $K = 7$ . Los subconjuntos de prueba para cada base de datos se describen en las Tablas 6 y 7.

<b>Diagnóstico</b>	<b>No. de imágenes</b>
<b>Benigno</b>	241
<b>Maligno</b>	82
<b>Normal</b>	33
<b>Total</b>	<b>356</b>

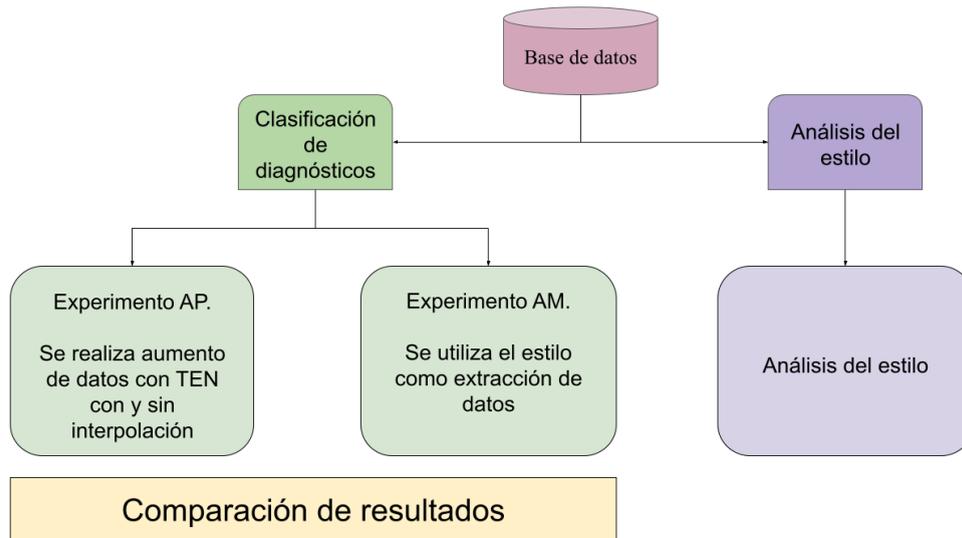
**Tabla 6.** Subconjunto de prueba para *Breast Ultrasound Images Dataset*.

<b>Diagnóstico</b>	<b>No. de imágenes</b>
<b>Queratosis actínica (akiec)</b>	220
<b>Carcinoma de células basales (bcc)</b>	397
<b>Lesiones queratósicas benignas (bkl)</b>	973
<b>Dermatofibroma (df)</b>	12
<b>Melanoma (mel)</b>	1,010
<b>Lesiones vasculares (Vasc)</b>	38
<b>Total</b>	<b>2,650</b>

**Tabla 7.** Subconjunto de prueba *HAM 10K*.

## 3.2. Experimentos

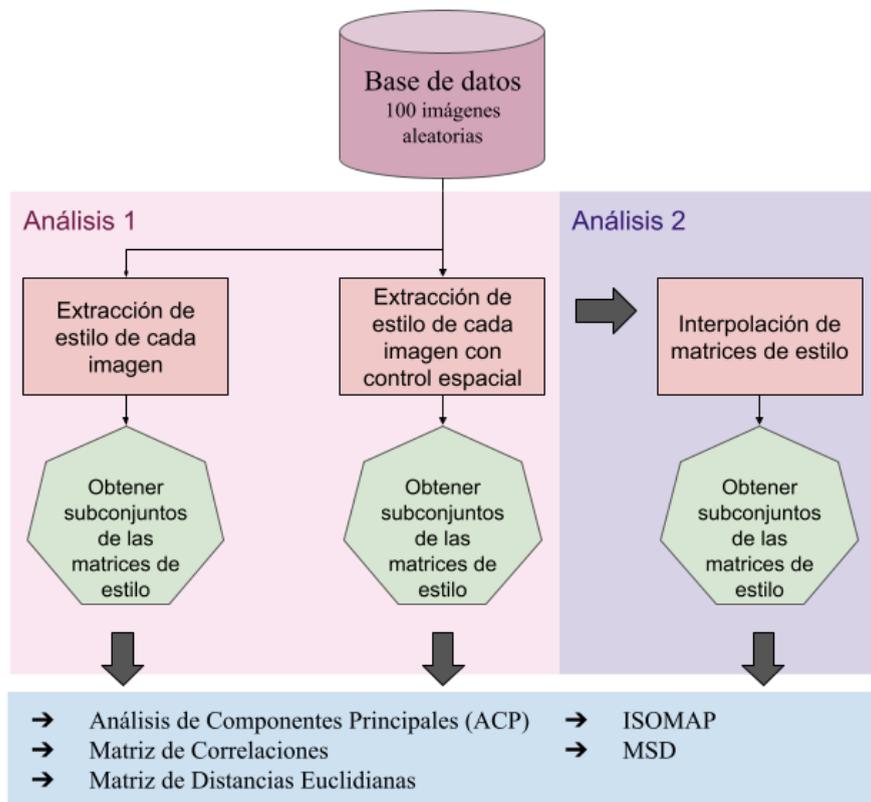
En esta sección se describen los experimentos realizados sobre las dos bases de datos mencionadas previamente. Se definen dos tipos de experimentos principales. El primero está enfocado en el análisis del estilo neuronal de las imágenes, mientras que el segundo utiliza dos estrategias basadas en Transferencia de Estilo Neuronal (TEN) para clasificación de diagnósticos (véase la Figura 9).



**Figura 9.** Descripción de la metodología.

### 3.2.1. Análisis del estilo neuronal

El análisis de los estilos neuronales se divide en dos partes principales, evaluando los resultados visualmente. El objetivo es explorar si los estilos extraídos de las imágenes, que contienen características que permitan discriminar entre diferentes diagnósticos. Para ello, se aplican técnicas de reducción de dimensionalidad, matrices de correlación y distancias euclidianas. En el Análisis 1, se estudian los estilos con y sin control espacial, y en el Análisis 2, se analizan los estilos interpolados con control espacial (véase la Figura 10).



**Figura 10.** Descripción del análisis del estilo.

### 3.2.1.1. Análisis 1: estilos con y sin control espacial

En esta primera parte del análisis, el objetivo es determinar si los estilos neuronales extraídos de las imágenes contienen características que faciliten la discriminación de diagnósticos. A partir del conjunto CA, se extraen sus estilos utilizando una red preentrenada VGG 19 con *ImageNet* [46] y el uso de matrices *Gram*. Los estilos neuronales se obtienen tanto con control espacial sobre las lesiones como sin él. Una vez extraídos los estilos, se analizan mediante técnicas de reducción de dimensionalidad, como el Análisis de Componentes Principales (ACP), Mapeo Isométrico (ISOMAP), y Escalamiento Multidimensional (MDS). Además, se calculan matrices de correlación y distancias euclidianas para evaluar las relaciones entre los estilos.

Para este análisis, se considera el uso de todos los subconjuntos posibles de las 5 matrices de estilo aplanadas, lo que permite explorar 31 combinaciones diferentes de estas matrices, es decir,  $\{(0), (1), (2), (3), (4), (0, 1), (0, 2), (0, 3), (0, 4), (1, 2), (1, 3), (1, 4), (2, 3), \dots, (0, 1, 2, 3, 4)\}$ , donde cada número del 0 al 4 representa a las matrices de estilo, para después aplanar cada matriz y concatenarla. Cada combinación de matrices se somete a las técnicas de reducción de dimensionalidad mencionadas, así como a las matrices de correlación y distancias euclidianas para tratar de medir la similitud entre los estilos neuronales y su capacidad de discriminar entre diagnósticos (véase la Figura 11).

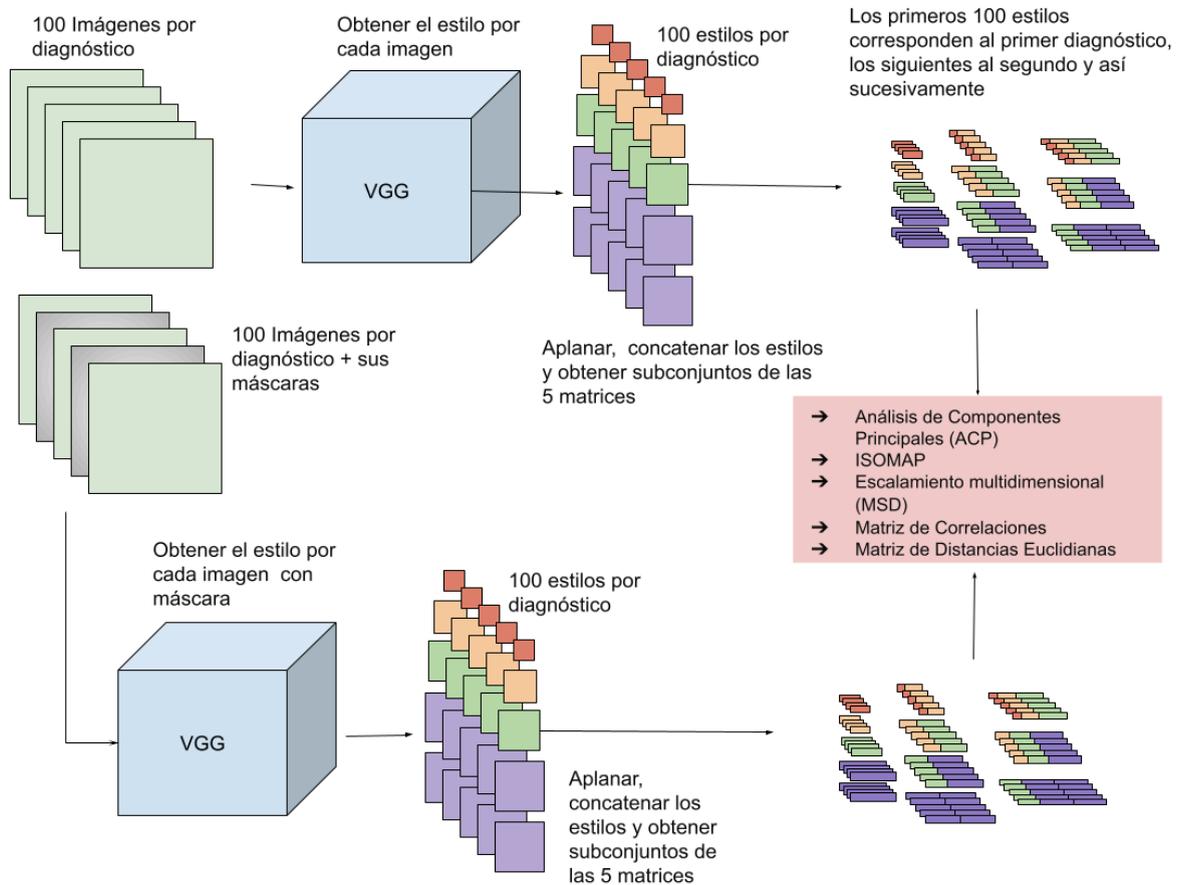
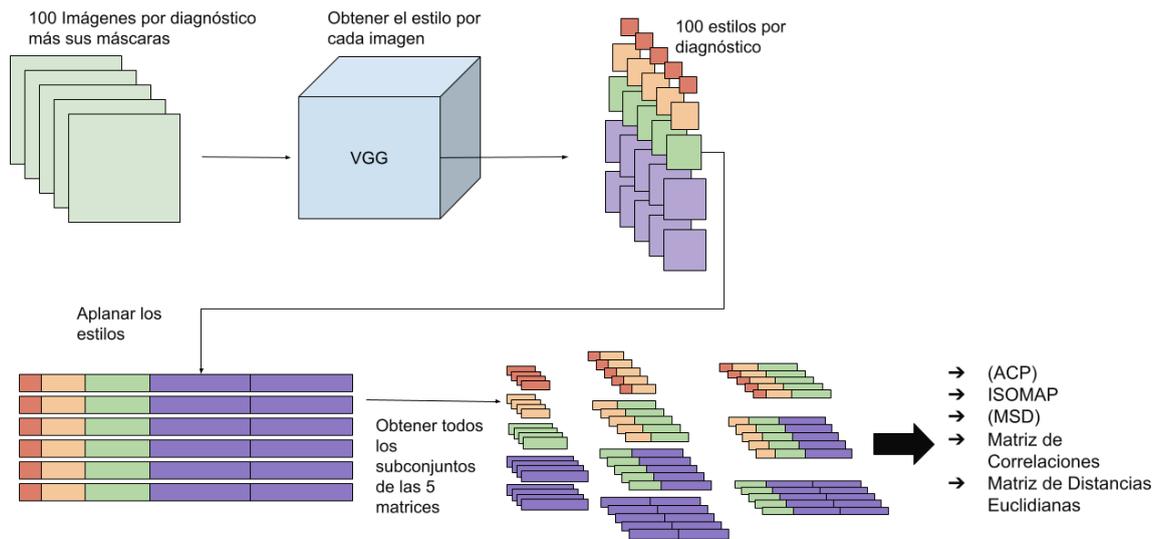


Figura 11. Proceso de análisis estilo con y sin control espacial.

### 3.2.1.2. Análisis 2: estilos interpolados

En la segunda parte del análisis, el objetivo es determinar si los estilos interpolados logran generar variaciones que mantengan la discriminación de diagnósticos. Para ello, se parte de la extracción de estilos con control espacial sobre el conjunto CA, y luego se realiza una interpolación de estos estilos con tres del mismo diagnóstico, seleccionados al azar. Este proceso genera 300 nuevos estilos por diagnóstico. Al igual que en el Análisis 1, los estilos interpolados se analizan utilizando los 31 subconjuntos posibles de las 5 matrices de estilo junto con técnicas de reducción de dimensionalidad, y las matrices de correlación y distancias euclidianas para evaluar las relaciones entre los estilos (véase la Figura 12).



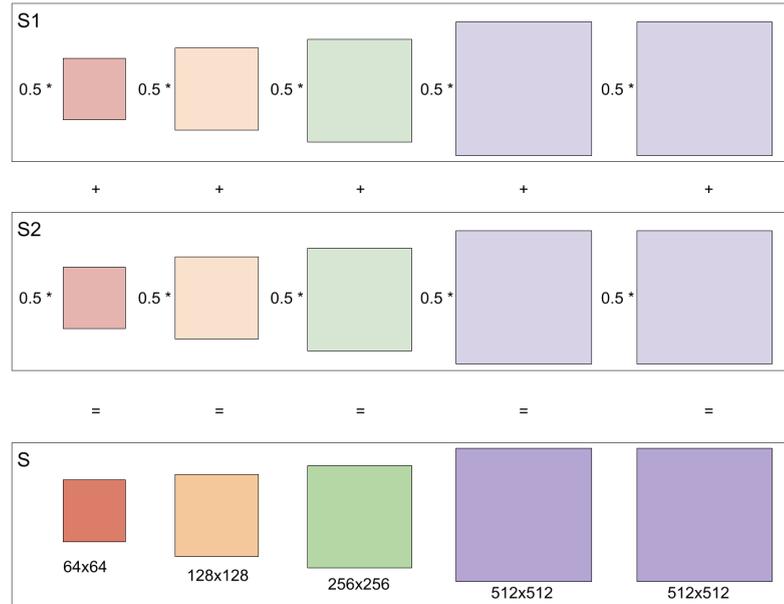
**Figura 12.** Proceso de análisis de estilos interpolados.

### 3.2.1.3. Interpolación de estilos

La interpolación entre 2 estilos se realiza entre cada una de sus 5 matrices, elemento por elemento, es decir, por cada matriz realizamos la ecuación 10:

$$S_i = (0.5 * S1_i) + (0.5 * S2_i) \quad (10)$$

Donde,  $S1$  y  $S2$  corresponden a los estilos que se van a interpolar,  $S$  es el estilo interpolado, e  $i$  corresponde a la matriz que forma el estilo, por lo que  $i \in [0, 1, \dots, 4]$  (véase la Figura 13).



**Figura 13.** Ejemplo práctico de interpolación de estilos elemento a elemento.

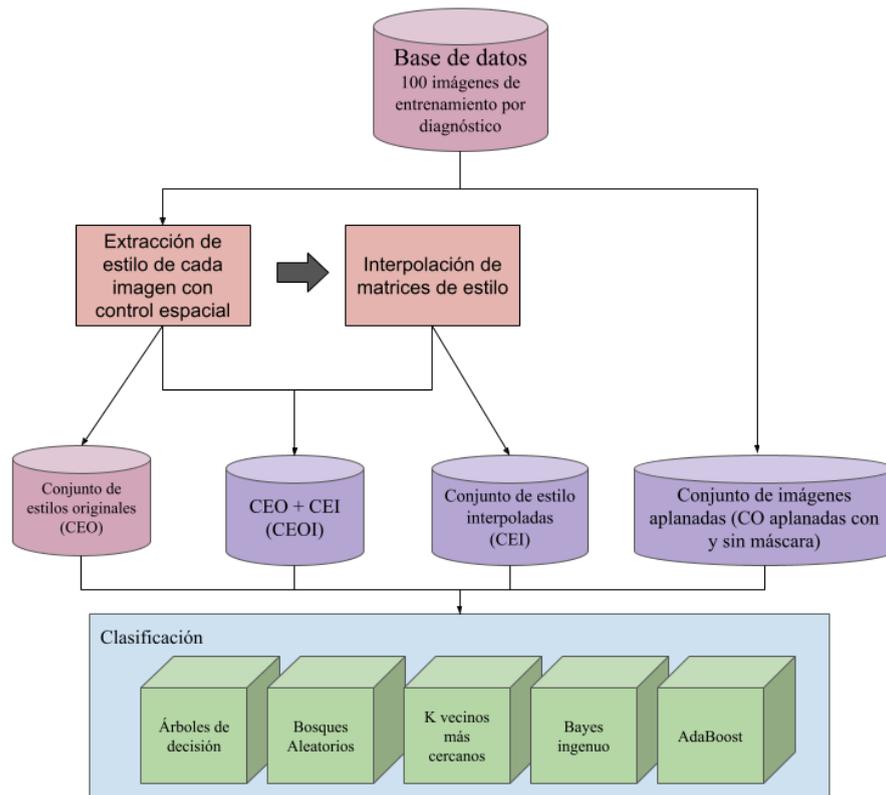
### 3.2.2. Experimentos de clasificación

El objetivo de estos experimentos de clasificación es determinar si las estrategias de Transferencia de Estilo Neuronal (TEN) pueden mejorar el rendimiento en la clasificación de pequeños conjuntos de datos. Para evaluar la efectividad de estas estrategias, se entrenan modelos utilizando imágenes originales y estilos generados mediante TEN. Posteriormente, se aplica un análisis estadístico para comparar los resultados obtenidos en términos de exactitud, con el fin de identificar si las diferencias en rendimiento entre las técnicas son significativas. Este enfoque permite evaluar el potencial de TEN para mejorar la clasificación en situaciones con conjuntos de datos limitados.

#### 3.2.2.1. Experimento de aprendizaje máquina AM

Para el experimento AM, es necesario contar con los estilos de cada imagen. Este experimento busca utilizar el estilo neuronal como técnica de extracción de características. Se definen cinco conjuntos de estudio: CO aplanados, CO aplanados con máscara, CEO, CEI y CEOI.

Este experimento busca comparar el rendimiento de los modelos al utilizar estilos neuronales como técnica de extracción de características frente a usar solo las imágenes aplanadas. Para integrar los estilos en los modelos de aprendizaje, primero se aplanan las cinco matrices que definen el estilo de cada imagen y luego se concatenan, formando un vector de 610,304 valores por estilo. Con estos conjuntos de estudio se entrenan cinco modelos de aprendizaje automático: árboles de decisión, bosques aleatorios, vecinos más cercanos, Naive Bayes y AdaBoost (véase la Figura 14).



**Figura 14.** Experimento de aprendizaje máquina con estilos. Para las bases de datos de ultrasonido de cáncer de mama y lesiones de la piel.

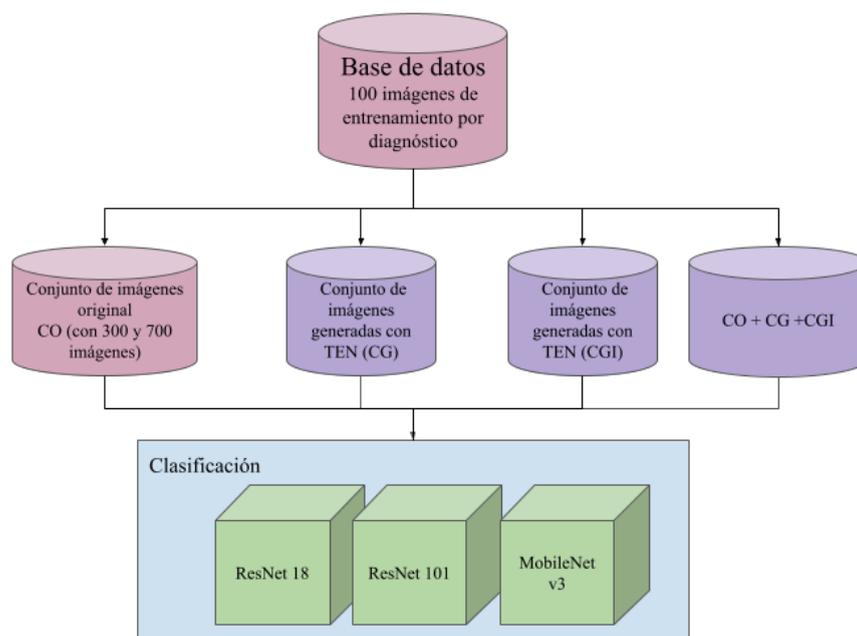
Se aplica la misma metodología estadística que en el experimento de AP, para evaluar la significancia de las diferencias en exactitud. Se utiliza Friedman, seguido de pruebas *post-hoc* donde utilizamos Nemenyi. Aquí, la prueba se realiza sobre todos los conjuntos de estudio para determinar si el uso de estilos neuronales como técnica de extracción de características ofrece una ventaja significativa en comparación con el uso de imágenes aplanadas. Definiendo las hipótesis para Friedman, tenemos:

- Hipótesis nula ( $H_0$ ): no hay diferencias significativas en la exactitud de los modelos entre los conjuntos CEO, CEI, CEOI, CO aplanados y CO aplanados con máscara.
- Hipótesis alternativa ( $H_a$ ): existen diferencias significativas en la exactitud de los modelos entre al menos dos de los conjuntos CEO, CEI, CEOI, CO aplanados y CO aplanados con máscara.

### 3.2.2.2. Experimento de aprendizaje profundo AP

Para el experimento de aprendizaje profundo (AP), se busca aumentar la cantidad de datos utilizando tanto técnicas de aumento de datos tradicional como estrategias basadas en TEN. El objetivo es entrenar y comparar modelos de AP con distintos conjuntos de estudio, observando su exactitud obtenida. Se definen seis conjuntos de estudio para cada una de las dos bases de datos: CO, CO 300, CO 700, CG, CGI, y CO + CG + CGI.

Estos seis conjuntos se emplean para entrenar tres modelos preentrenados con *ImageNet* [46] de redes neuronales convolucionales: ResNet 18, ResNet 101 y MobileNet. Al conjunto CO también se le aplica aumento de datos tradicional para generar conjuntos con diferentes tamaños (CO 300 y CO 700), permitiendo así una comparación exhaustiva de los resultados obtenidos (véase la Figura 15).



**Figura 15.** Experimento de aprendizaje profundo. Para las bases de datos de ultrasonido de cáncer de mama y lesiones de la piel.

Para este experimento, es necesario aplicar una prueba de significancia estadística que permita comparar los resultados de exactitud entre los diferentes conjuntos de datos. Se aplica la prueba de Friedman, que es adecuada para evaluar si existen diferencias significativas entre los conjuntos cuando los datos son apareados y no siguen una distribución normal.

- Hipótesis nula ( $H_0$ ): no hay diferencias significativas en la exactitud de los modelos entre los conjuntos CO, CO 300, CO 700, CG, CGI, y CO + CG + CGI.
- Hipótesis alternativa ( $H_a$ ): existen diferencias significativas en la exactitud de los modelos entre al menos dos de los conjuntos CO, CO 300, CO 700, CG, CGI, y CO + CG + CGI.

Posteriormente, para profundizar en los resultados, se realiza una prueba *post-hoc* de Nemenyi, cuyo objetivo es identificar qué pares de conjuntos de imágenes presentan diferencias significativas en la exactitud de los modelos. Este análisis abarca todos los conjuntos de estudio, con el fin de evaluar cómo influye en el rendimiento de los modelos el uso de imágenes originales con aumento de datos tradicional, imágenes generadas mediante TEN y estilos interpolados.

### 3.2.2.3. Comparación de modelo - conjunto

Finalmente, se comparan los resultados obtenidos en el experimento de aprendizaje automático (AM) con los del experimento de aprendizaje profundo (AP), utilizando los conjuntos de entrenamiento para cada modelo. Para analizar el rendimiento de los modelos entrenados con imágenes en el experimento de AP y los entrenados con estilos neuronales en el experimento de AM, se llevaron a cabo 2 análisis, el primero donde se hace una inspección visual por modelo de la mejor exactitud al usar los conjuntos, y luego comparándolas.

El segundo análisis fue una comparación formal entre los resultados de AM y AP, utilizando pruebas de significancia estadística mediante métodos no paramétricos. Debido a que los experimentos AM y AP difieren en aspectos metodológicos clave, como el valor de  $K$  y los modelos empleados, se empleó la prueba de Mann-Whitney U, adecuada para comparar distribuciones independientes. Esta prueba permite analizar datos sin necesidad de que sigan una distribución normal o tengan homogeneidad de varianzas, y se aplicó en este caso en una prueba de una cola para determinar si los modelos entrenados con estilos neuronales en AM superan significativamente en rendimiento a los modelos entrenados con imágenes en AP. Las hipótesis se definieron de la siguiente manera:

- Hipótesis nula ( $H_0$ ): la exactitud de los experimentos de AM no tiene un rendimiento mayor que los experimentos AP, (es decir, su rendimiento es igual o menor).
- Hipótesis alternativa ( $H_a$ ): la exactitud de los experimentos de AM tiene un rendimiento significativamente mayor que los experimentos AP.

### 3.3. Detalles Técnicos

Para que los experimentos puedan ser replicados, se describen a continuación los detalles técnicos y herramientas utilizadas:

1. Lenguaje de Programación: Python 3.8 para el análisis de estilos y los experimentos de clasificación. Python 2.7 para hacer la transferencia de estilo neuronal con y sin control espacial y la obtención de estilos.
2. Bibliotecas Utilizadas:
  - SciPy [60] y scikit-learn [61] para la reducción de dimensionalidad y la implementación de modelos de aprendizaje automático.
  - TensorFlow 1 [62] para la implementación de la Transferencia de Estilo Neuronal (TEN) con y sin control espacial, utilizando el repositorio de GitHub "Doodle" [63].
  - Fastai [64] y PyTorch [65] para la implementación y entrenamiento de los modelos de aprendizaje profundo.
3. Entrenamiento:
  - Los modelos de aprendizaje profundo se entrenaron utilizando validación cruzada con  $K = 5$  para cada conjunto de estudio definido. El proceso de entrenamiento se realizó mediante la función `fine_tune`, la cual congela las primeras capas del modelo (preservando el conocimiento general previamente aprendido) y ajusta las capas finales para que el modelo se adapte de manera óptima a la nueva tarea.
  - Los modelos de aprendizaje automático se entrenaron con validación cruzada  $K = 7$ , utilizando los estilos neuronales como vectores de características.
4. Procedimientos específicos:
  - Extracción de estilos. Se utilizó el modelo preentrenado VGG 19 junto con la aplicación de la matriz *Gram* con y sin control espacial para extraer los estilos neuronales de las imágenes. Para la interpolación, se utilizaron estilos con control

espacial de dos imágenes, para luego combinarlos a través de una interpolación ponderada de las matrices de estilo correspondientes.

- Generación de imágenes. Para la generación de imágenes mediante TEN, cada imagen del conjunto CO se utiliza como imagen de contenido. Para las imágenes de estilo, se seleccionan aleatoriamente imágenes del mismo diagnóstico dentro del conjunto CO. A continuación, se transfiere el estilo de las lesiones de las imágenes seleccionadas al contenido de las imágenes de contenido haciendo 200 iteraciones.
- Aumento de Datos Tradicional: para generar los conjuntos CO 300 y CO 700, se aplicaron transformaciones aleatorias como rotaciones y traslaciones a las imágenes originales del conjunto CO.
- Evaluación: el rendimiento de los modelos se midió con la exactitud de los modelos, comparando los resultados entre los distintos conjuntos de estudio.
- Pruebas de significancia estadística: se aplicó la prueba de Friedman en los resultados de los experimentos de aprendizaje automático (AM) y aprendizaje profundo (AP), para determinar si existían diferencias significativas entre los conjuntos de datos. Cuando Friedman mostró diferencias, se utilizó la prueba *post-hoc* de Nemenyi para identificar los pares de conjuntos con diferencias significativas. Finalmente, se empleó la prueba de Mann-Whitney U para comparar las distribuciones de la exactitud entre los experimentos AM y AP, evaluando las diferencias en la forma general de las distribuciones.

## 4. Resultados y análisis

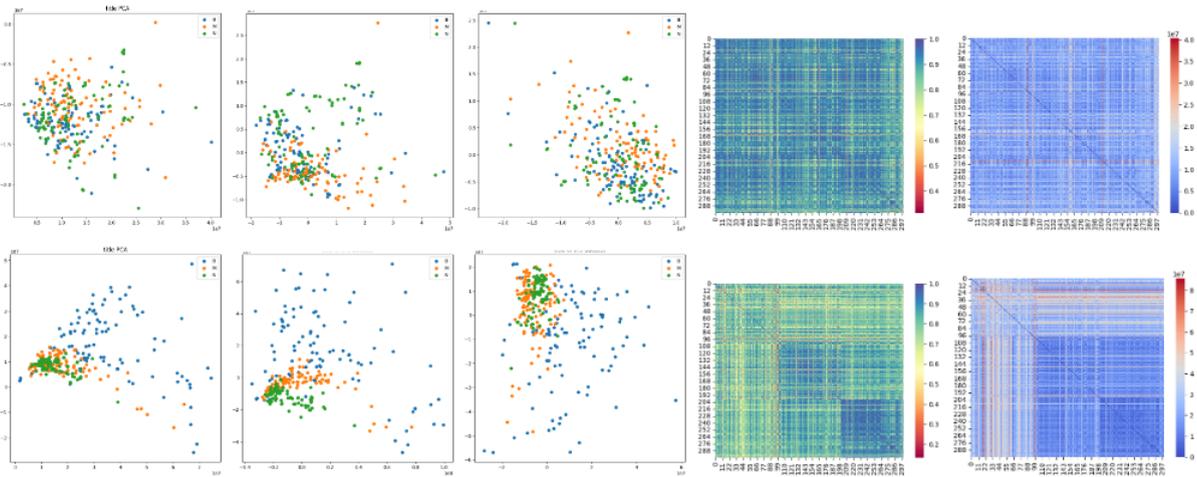
En este capítulo se presentan los resultados y análisis obtenidos para dos análisis sobre el estilo neuronal, el primero es del estilo con y sin control espacial, y el segundo, sobre los estilos interpolados. En la siguiente parte se muestran ejemplos de imágenes generadas con transferencia de estilo neuronal (TEN) con control espacial sobre las lesiones y TEN con estilos interpolados como objetivos. Después se abordan los experimentos de clasificación, se muestran los resultados para los experimentos de aprendizaje profundo (AP) y aprendizaje automático (AM). Por último, se muestran y se realiza un análisis de los resultados de ambos experimentos de clasificación.

### 4.1. Análisis de estilos neuronales

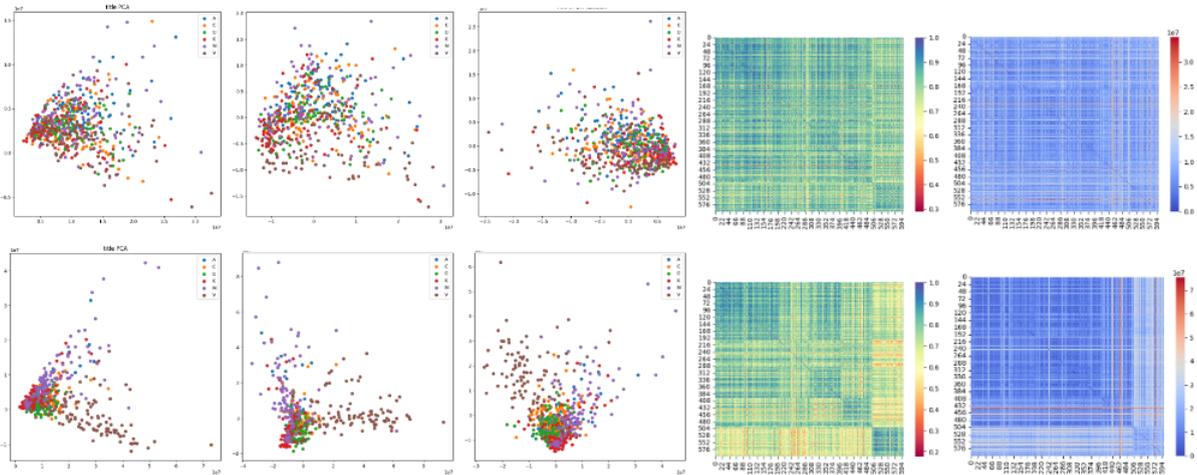
El análisis de estilos neuronales se divide en dos fases principales, donde se busca comprender cómo los estilos extraídos de las imágenes se agrupan y diferencian según los diagnósticos. Por lo que en ambos análisis, se aplican técnicas de reducción de dimensionalidad, como ACP, ISOMAP y MSD, junto con matrices de correlación y distancias euclidianas, para determinar la relación entre los estilos neuronales a través de una inspección visual de los resultados. A continuación, se detallan los dos enfoques: el análisis con y sin control espacial (Análisis 1), y el análisis con interpolación de estilos (Análisis 2).

#### 4.1.1. Análisis 1: estilos con y sin control espacial.

Para el Análisis 1, se obtuvieron 31 subconjuntos de matrices de estilo para el análisis sin control espacial y 31 subconjuntos para el análisis de estilos con control espacial, a estos subconjuntos se les aplicaron las técnicas de ACP, ISOMAP, MSD, y las matrices de correlación y distancias euclidianas. Por lo tanto, se obtuvieron 310 gráficas para la base de datos de cáncer de mama y 310 gráficas para la base de datos de lesiones de la piel. Ya que son demasiadas gráficas, se escogieron las gráficas donde utilizan todo el estilo, como representantes del comportamiento de los estilos (véase las Figuras 16 y 17).



**Figura 16. Análisis 1. Ultrasonido de cáncer de mama.** El renglón superior muestra los resultados de los estilos sin control espacial. El renglón inferior muestra los resultados de los estilos con control espacial. Las columnas indican de izquierda a derecha: ACP, ISOMAP, MSD, matrices de correlación, matrices de diferencia euclidiana. En las matrices, se dividen los índices por tercios, el primer tercio corresponde a estilos de benigno, luego maligno y al último normal.



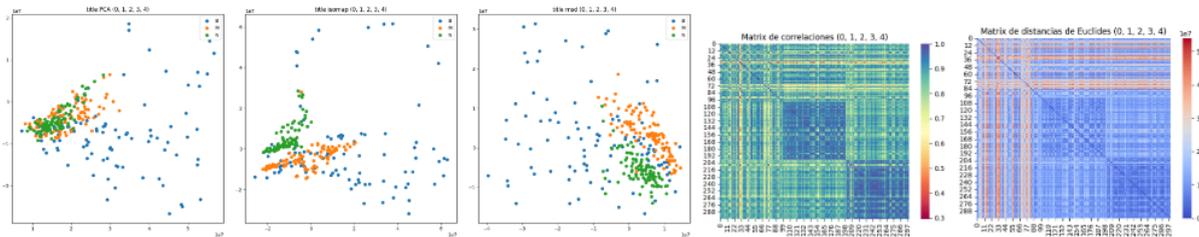
**Figura 17. Análisis 1. Lesiones de la piel.** El renglón superior muestra los resultados de los estilos sin control espacial. El renglón inferior muestra los resultados de los estilos con control espacial. Las columnas indican de izquierda a derecha: ACP, ISOMAP, MSD, matrices de correlación, matrices de diferencia euclidiana. En las matrices, se dividen los índices por sextos, el primer sexto corresponde a estilos de akiec, luego bcc, bkl, df, mel y al último vas.

En este primer análisis, se observa que el método sin control espacial no permite distinguir entre diagnósticos, ya que ninguna de las gráficas muestra una clara separación entre diagnósticos (puntos de colores). Sin embargo, en la parte del análisis con control espacial, los datos logran agruparse de acuerdo a sus diagnósticos (representados por puntos de colores) en las gráficas de ACP, ISOMAP y MSD. Además, en las matrices de distancia euclidiana y de correlación, se pueden distinguir algunos cuadrados de tonos más azules (menor distancia o mayor correlación) en la diagonal principal y en los índices correspondientes a los diagnósticos, lo que indica que en algunos casos, los datos son más cercanos o tienen mayor correlación cuando se comparan con datos del mismo diagnóstico.

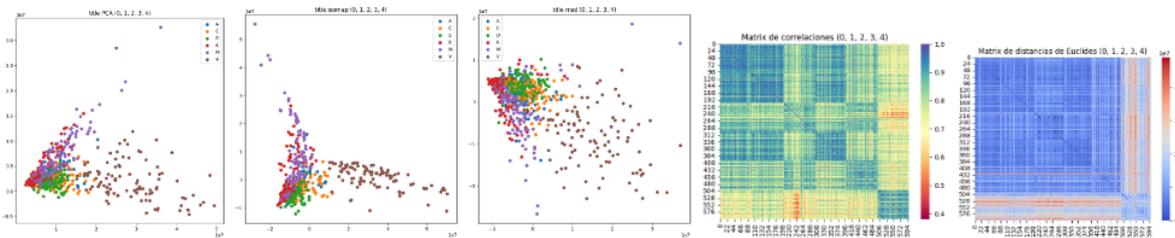
#### 4.1.2. Análisis 2: interpolación de estilos.

Para el análisis 2 del estilo, vemos el comportamiento de los estilos interpolados con control espacial. Estos estilos muestran un comportamiento, en general, más disperso que en el análisis 1 con control espacial, pero también muestra que para cada diagnóstico hay un parecido mayor, esto es más evidente al observar las matrices de correlación. Este comportamiento puede ser lo que estamos

buscando al hacer aumento de datos, ya que produce una mayor variabilidad en los datos, pero al mismo tiempo asegura que se mantienen agrupados por diagnóstico (véase las Figuras 18 y 19).



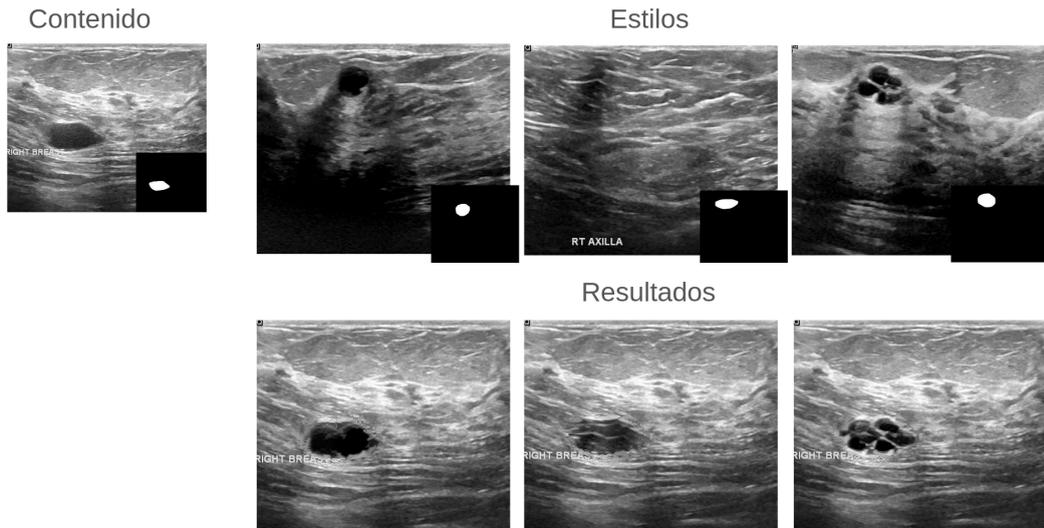
**Figura 18. Análisis 2, Estilos interpolados de ultrasonido de cáncer de mama. Las columnas indican de izquierda a derecha:** ACP, ISOMAP, MSD, matrices de correlación, matrices de diferencia euclidiana. En las matrices, se dividen los índices por tercios, el primer tercio corresponde a estilos de benigno, luego maligno y al último normal.



**Figura 19. Análisis 2, Estilos interpolados de lesiones de la piel. Las columnas indican de izquierda a derecha:** ACP, ISOMAP, MSD, matrices de correlación, matrices de diferencia euclidiana. En las matrices, se dividen los índices por sextos, el primer sexto corresponde a estilos de akiec, luego bcc, bkl, df, mel y al último vas.

## 4.2. Ejemplos de imágenes generadas con TEN

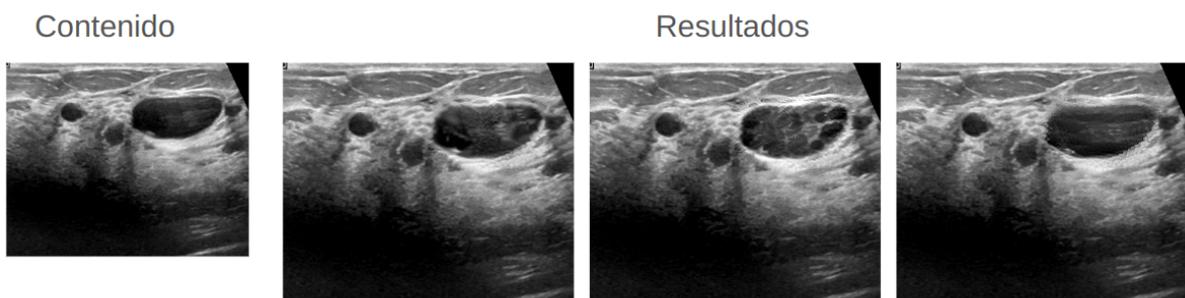
En esta sección vamos a observar algunos ejemplos de imágenes generadas con TEN usando control espacial (véase las Figuras 20 y 21) e imágenes generadas a partir de estilos interpolados (véase las Figuras 22 y 23), para las bases de datos de imágenes de cáncer de mama de ultrasonido y de lesiones de la piel. En las 2 siguientes figuras, la imagen de la esquina superior izquierda es el contenido, las imágenes del renglón superior son las imágenes de estilo y por debajo de cada uno se encuentran las imágenes de resultado. Para las figuras generadas con estilos interpolados, la imagen de la izquierda es la imagen de contenido y las otras 3 son los resultados.



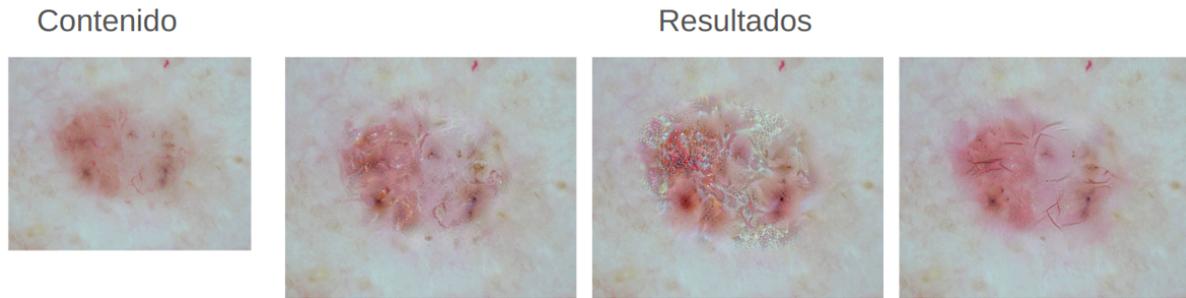
**Figura 20.** Ejemplos de imágenes generadas con TEN y control espacial para ultrasonido de cáncer de mama.



**Figura 21.** Ejemplos de imágenes generadas con TEN para lesiones de la piel.



**Figura 22.** Ejemplos de imágenes generadas con estilos interpolados para ultrasonido de cáncer de mama.



**Figura 23.** Ejemplos de imágenes generadas con estilos interpolados para lesiones de la piel.

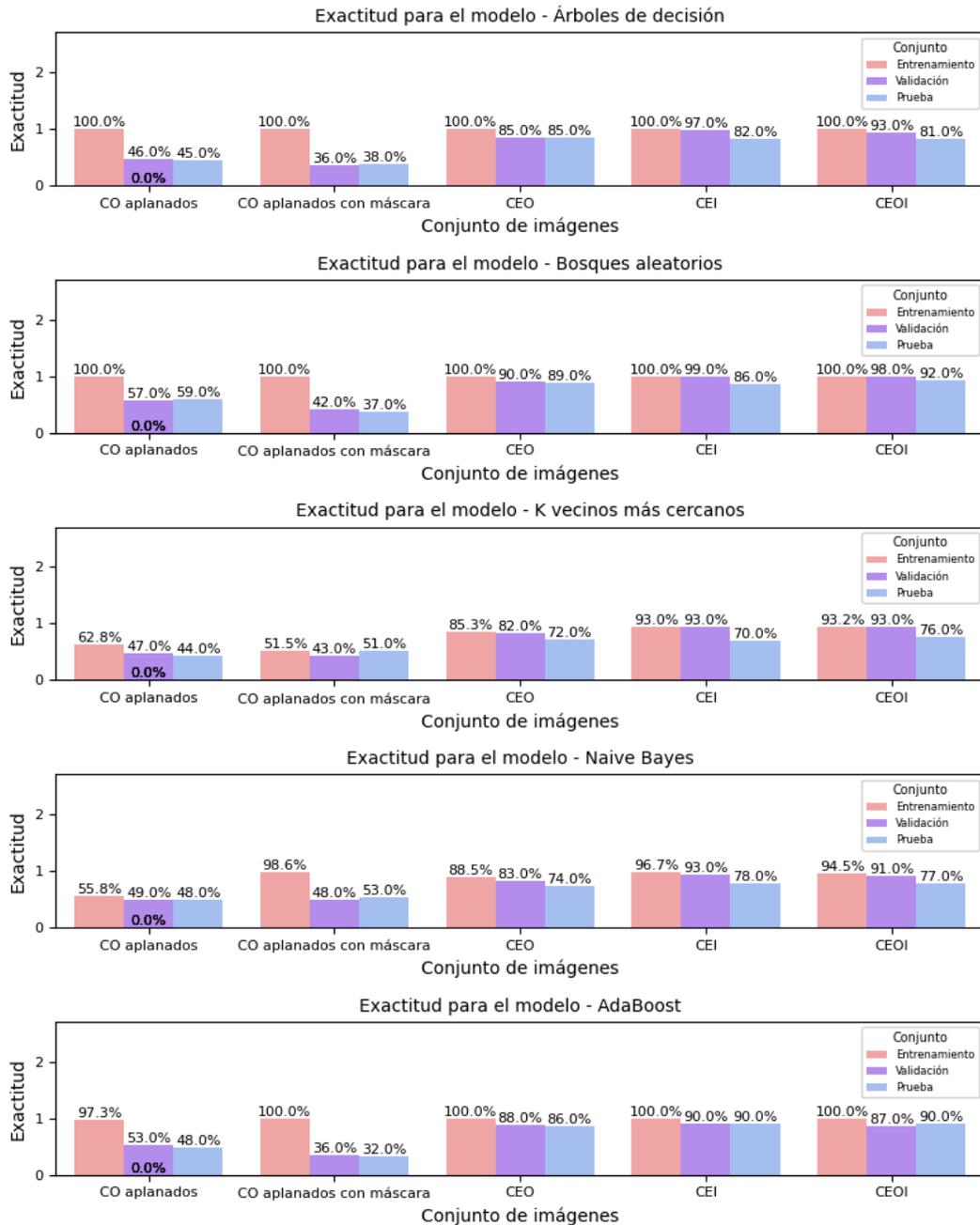
Como podemos observar, las imágenes generadas muestran una variación en el estilo solo en la región de la lesión. Estas variaciones y la calidad de la imagen final depende en gran medida del estilo objetivo. En las imágenes de ejemplo generadas con TEN con control espacial, se aprecia claramente el origen del estilo. Sin embargo, en las imágenes generadas con estilos interpolados, esto no es tan evidente. De hecho, se observa que algunos estilos interpolados pueden generar más artefactos, en comparación con las imágenes generadas mediante TEN con control espacial. Por ejemplo, en las imágenes de ultrasonido con estilos interpolados, las lesiones muestran un estilo más suave y parecido entre ellas, mientras que en la imagen central de lesiones de la piel, el resultado parece más sintético o con la presencia de posibles artefactos.

### 4.3. Experimentos de clasificación

En esta sección se presentan los resultados obtenidos en los experimentos de clasificación utilizando aprendizaje profundo (AP) y aprendizaje máquina (AM) con las bases de datos de imágenes de ultrasonido de cáncer de mama y lesiones de la piel. Los resultados obtenidos se analizan con base en los conjuntos de prueba, proporcionando una visión del posible impacto de los conjuntos de datos generados en el rendimiento de los modelos de clasificación. A continuación, se detallan los resultados obtenidos en los experimentos con AP y AM, así como una comparación entre los mejores modelos para cada técnica y base de datos.

#### 4.3.1. Resultados de experimentos AM

A continuación, se muestran los resultados obtenidos al entrenar los modelos de AM, solo para los subconjuntos de entrenamiento, validación y prueba (véase las Figuras 24 y 25).



**Figura 24.** Resultados de la base de datos *Breast Cancer Ultrasound* experimento aprendizaje máquina.

En el experimento de clasificación para ultrasonidos utilizando el aprendizaje automático (AM), observamos mediante una inspección visual que los resultados para los conjuntos de datos de imágenes aplanadas, donde se realiza la clasificación a nivel de píxel, son generalmente bajos, con valores de exactitud inferiores al 50%. En contraste, los resultados para los conjuntos que emplean estilos neuronales (CEO, CEI y CEOI) son significativamente más altos, alcanzando valores cercanos o superiores al 90% en modelos como bosques aleatorios y AdaBoost. Esto puede sugerir que el uso de estilo neuronal como técnica de extracción de características contribuye de manera positiva a la clasificación del diagnóstico, mejorando notablemente el rendimiento en comparación con las imágenes aplanadas.

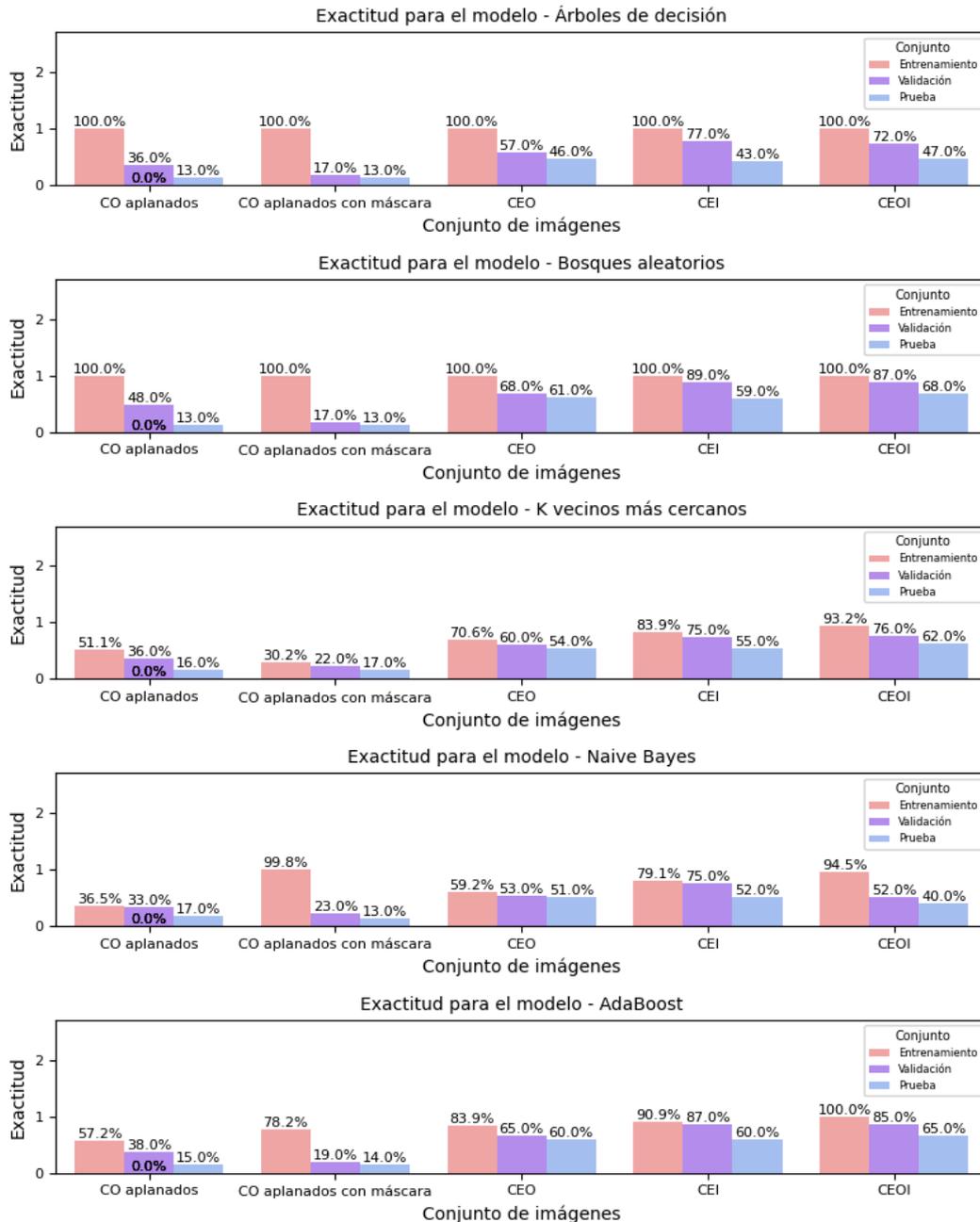
Para realizar un análisis formal de las diferencias en el rendimiento de los modelos entre los distintos conjuntos de estudio, se aplicó una prueba de Friedman. El estadístico de Friedman fue de 15.4747,

con un p-valor de 0.0038, lo que permite rechazar la hipótesis nula y concluir que existen diferencias significativas en la exactitud de los modelos entre al menos dos de los conjuntos evaluados ( $p < 0.05$ ).

Para identificar específicamente entre cuáles conjuntos existen estas diferencias significativas, se realizó una prueba *post-hoc* de Nemenyi. Los p-valores obtenidos en esta prueba se muestran en la siguiente Tabla 8, donde se observa que las comparaciones CEOI - CO aplanados y CEOI - CO aplanados con máscara presentan diferencias significativas, con p-valores de 0.053 y 0.030, respectivamente. Además, los pares CEO - CO aplanados y CEI - CO aplanados con máscara tienen p-valores relativamente bajos (0.179 y 0.09), lo que sugiere que, aunque no alcanzan el umbral de significancia, muestran una tendencia a diferencias en rendimiento. Estos resultados indican que el uso de estilos neuronales, especialmente en el conjunto CEOI, podría aportar beneficios en comparación con los conjuntos de imágenes aplanadas tradicionales.

Conjunto comparado	CO aplanados	CO aplanados con máscara	CEO	CEI	CEOI
CO aplanados	1.0	0.999	0.179	0.144	<b>0.053</b>
CO aplanados con máscara	0.999	1.0	0.115	<b>0.09</b>	<b>0.030</b>
CEO	0.179	0.115	1.0	0.999	0.987
CEI	0.144	<b>0.09</b>	0.999	1.0	0.994
CEOI	<b>0.053</b>	<b>0.030</b>	0.987	0.994	1.0

**Tabla 8.** Prueba *post-hoc* de Nemenyi para el Experimento AM con imágenes de ultrasonido de cáncer de mama. Los valores resaltados indican los p-valores más bajos.



**Figura 25.** Resultados de la base de datos HAM 10K experimento aprendizaje máquina.

En el experimento de aprendizaje automático (AM) para la base de datos de lesiones de la piel, observamos mediante una inspección visual que el uso de estilos neuronales como características resulta ser una estrategia prometedora, siempre y cuando se emplee el modelo adecuado. Por ejemplo, el conjunto de estilos CEOI, utilizado con el modelo de bosques aleatorios, alcanza una exactitud del 68.0%, lo cual es superior en comparación con los conjuntos aplanados.

Una observación relevante es que los conjuntos de datos CO aplanados y CO aplanados con máscara, junto con CEO, contienen solo 100 instancias. En cambio, los conjuntos CEO y CEOI contienen más datos, ya que representan un aumento de datos en términos de estilo. Esto permite evaluar el impacto de no realizar la extracción de estilo frente a realizar un aumento de datos específico para el estilo.

Para realizar un análisis formal de las diferencias en el rendimiento de los modelos entre los distintos conjuntos de estudio, se aplicó la prueba de Friedman. El estadístico de Friedman fue de 16.6185, con

un p-valor de 0.0022, lo que indica que se rechaza la hipótesis nula y existen diferencias significativas en la exactitud de los modelos entre al menos dos de los conjuntos de estudio ( $p < 0.05$ ).

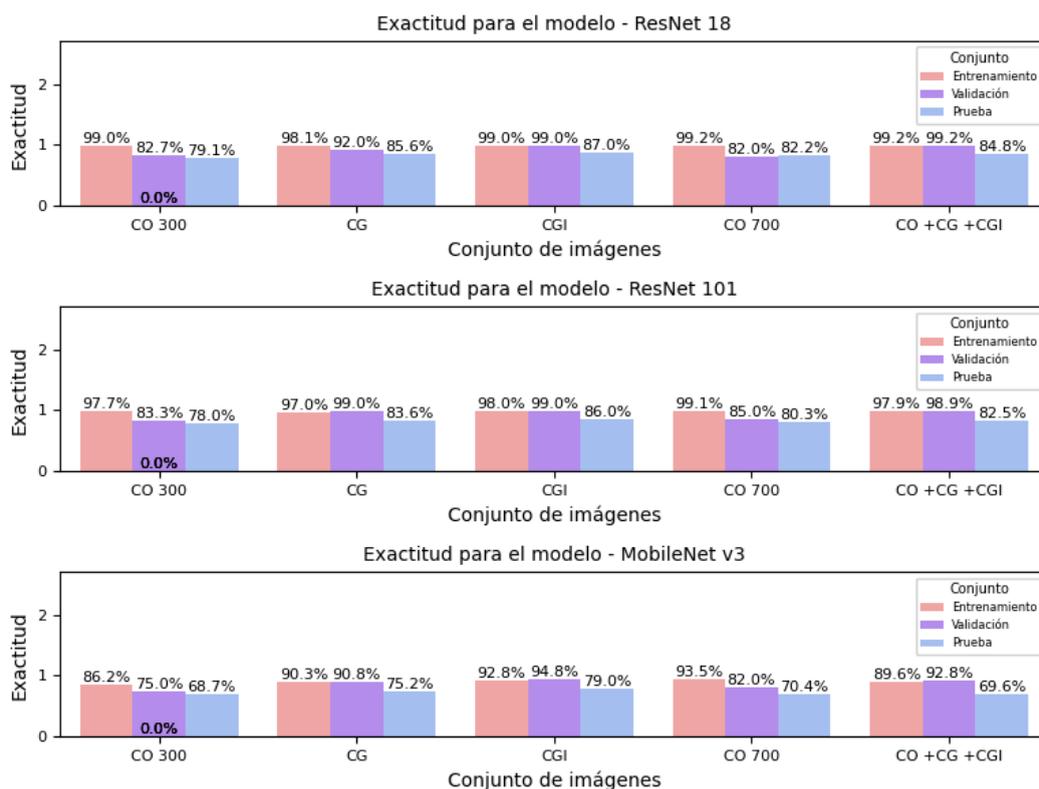
Para identificar entre cuáles conjuntos existen estas diferencias, se realizó una prueba *post-hoc* de Nemenyi, cuyos p-valores se presentan en la siguiente Tabla 9, destacando que el conjunto CEOI es el único que presenta diferencias estadísticamente significativas en comparación con los conjuntos aplanados ( $p < 0.05$ ). Esto sugiere que CEOI, que utiliza estilos interpolados, podría ofrecer una ventaja en el diagnóstico de lesiones de la piel frente a los conjuntos tradicionales, tanto con como sin máscara. En contraste, las comparaciones entre los conjuntos de estilos CEO, CEI y los conjuntos aplanados no muestran diferencias significativas, indicando que estos no aportan una mejora clara en rendimiento en este contexto.

Conjunto comparado	CO aplanados	CO aplanados con máscara	CEO	CEI	CEOI
CO aplanados	1.0	0.144	0.219	0.219	<b>0.022</b>
CO aplanados con máscara	0.144	1.0	0.144	0.144	<b>0.011</b>
CEO	0.219	0.144	1.0	1.0	0.896
CEI	0.219	0.144	1.0	1.0	0.896
CEOI	<b>0.022</b>	<b>0.011</b>	0.896	0.896	1.0

**Tabla 9.** Prueba *post-hoc* de Nemenyi para el Experimento AM con imágenes de lesiones de la piel. Los valores resaltados indican los p-valores más bajos.

#### 4.3.2. Resultados de experimentos AP

A continuación, se muestran los resultados obtenidos al entrenar los modelos de AP, donde el subconjunto de mayor interés es el de prueba, ya que muestra el rendimiento de los modelos frente a imágenes reales que nunca ha visto (véase las Figuras 26 y 27).



**Figura 26.** Resultados de la base de datos Breast Cancer Ultrasound experimento aprendizaje profundo.

Analizando los resultados del subconjunto de prueba en la base de datos de ultrasonidos de cáncer de mama, se observa, mediante una inspección visual, que el uso de los conjuntos de estudio CO 100, CO 300 y CO 700, generados mediante aumento de datos tradicional, no mejora de forma consistente el rendimiento de los modelos de aprendizaje profundo. En el único caso en el que se logra obtener siempre una mejor exactitud es con el modelo MobileNet.

Al comparar los conjuntos CG, CGI y CO+CG+CGI, se puede observar que estos siempre producen un aumento en el rendimiento de los modelos, siendo de igual forma los resultados en el modelo MobileNet los más destacables. Con este modelo, CG logra una exactitud 6.5% mayor que CO 300 y CGI un 10.3% superior. El mejor par modelo-conjunto es ResNet 18 con CGI, que alcanza un 87.0% de exactitud, lo cual representa una mejora de 7.9% en comparación con el rendimiento de ResNet 18 con CO 300 (79.1%), lo que puede sugerir que la estrategia de aumento de imágenes con estilos interpolados es una mejor técnica que las tradicionales. Un punto a destacar es que el conjunto CO+CG+CGI no obtiene un rendimiento superior a los conjuntos CG y CGI, y su diferencia con CO 700 es pequeña.

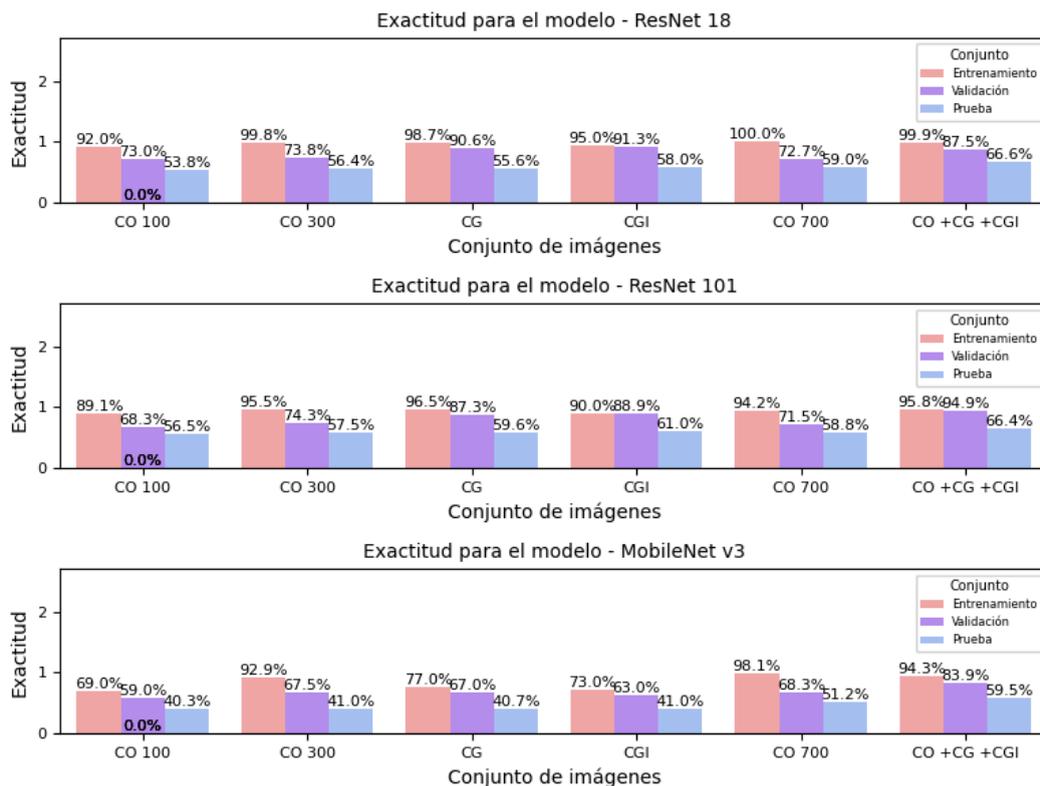
Para realizar un análisis más formal, se aplicó una prueba de Friedman para comparar la exactitud de los modelos en los distintos conjuntos de estudio. El estadístico de Friedman fue 13.476, con un p-valor de 0.0193, lo que indica que se rechaza la hipótesis nula y se concluye que existen diferencias significativas en la exactitud entre al menos dos de los conjuntos ( $p < 0.05$ ).

Para identificar específicamente entre qué conjuntos existen estas diferencias significativas, se realizó una prueba *post-hoc* de Nemenyi. Los p-valores obtenidos en esta prueba se presentan en la Tabla 10. Los resultados muestran que la única comparación con diferencia estadísticamente significativa es

entre los conjuntos CO y CGI. Esto sugiere que, aunque la mayoría de los conjuntos de aumento de datos no mejoran significativamente frente a CO, CGI al incorporar interpolación de estilos, ofrece un beneficio adicional en exactitud respecto al conjunto original.

Conjunto comparado	CO	CO 300	CO 700	CG	CGI	CO + CG + CGI
CO	1.0	0.998	0.986	0.363	<b>0.092</b>	0.885
CO 300	0.998	1.0	0.885	0.155	0.646	0.646
CO 700	0.986	0.885	1.0	0.780	0.998	0.998
CG	0.363	0.155	0.780	1.0	0.953	0.953
CGI	<b>0.092</b>	0.646	0.998	0.953	1.0	0.646
CO + CG + CGI	0.885	0.646	0.998	0.953	0.646	1.0

**Tabla 10.** Prueba *post-hoc* de Nemenyi para el Experimento AP con imágenes de ultrasonido de cáncer de mama. Los valores resaltados indican los p-valores más bajos.



**Figura 27.** Resultados de la base de datos HAM 10K experimento aprendizaje profundo.

Al realizar la inspección visual de los resultados obtenidos en la base de datos de lesiones de piel, observamos que los modelos no lograron superar el 70% de exactitud en el subconjunto de prueba. El mejor desempeño se registró con el modelo ResNet 18 en el conjunto CO+CG+CGI, alcanzando una exactitud del 66.6%, lo cual representa una diferencia de 7.6% en comparación con ResNet 18 en el conjunto CO 700. No obstante, en la mayoría de los casos, los modelos no lograron superar una

exactitud del 60%. Además, un aspecto notable es que, en esta base de datos, el aumento en la cantidad de imágenes parece mejorar el rendimiento, lo cual puede implicar que para esta base de datos es primordial tener más imágenes. Esto contrasta con los resultados en la base de datos de ultrasonido, donde el rendimiento tiende a disminuir al utilizar conjuntos de 700 imágenes por diagnóstico.

Para realizar un análisis estadístico formal y evaluar si las diferencias observadas en la exactitud entre los conjuntos de estudio son significativas, se aplicó una prueba de Friedman. El estadístico de Friedman fue de 12.788, con un p-valor de 0.0254, lo que permite rechazar la hipótesis nula y concluir que existen diferencias significativas en la exactitud de los modelos entre al menos dos de los conjuntos ( $p < 0.05$ ).

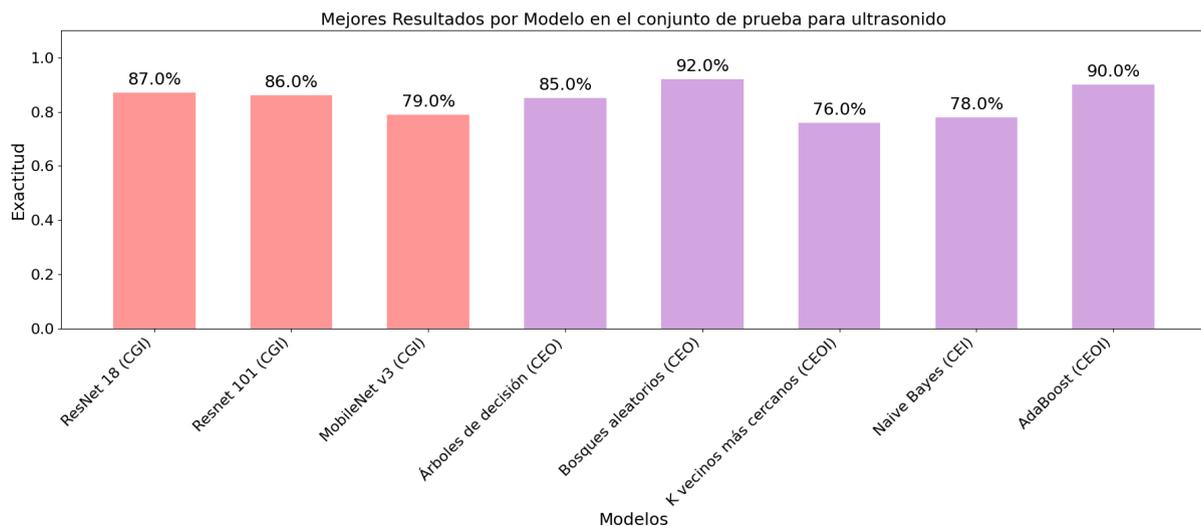
Para identificar cuáles pares de conjuntos presentan estas diferencias significativas, se realizó una prueba *post-hoc* de Nemenyi. Los p-valores de esta prueba se muestran en la siguiente Tabla 11. Los resultados indican que la única diferencia estadísticamente significativa se encuentra entre los conjuntos CO y CO + CG + CGI ( $p = 0.013$ ). Este hallazgo sugiere que, al combinar las imágenes originales (CO) con las generadas mediante TEN (CG y CGI), se logra una mejora significativa en comparación con el conjunto original solo (CO). Sin embargo, en las demás comparaciones, los p-valores son altos, lo que indica que no existen diferencias significativas en el rendimiento de exactitud entre el resto de los conjuntos y el conjunto CO.

Conjunto comparado	CO	CO 300	CO 700	CG	CGI	CO + CG + CGI
CO	1.0	0.836	0.246	0.885	0.301	<b>0.013</b>
CO 300	0.836	1.0	0.923	0.999	0.953	0.301
CO 700	0.246	0.923	1.0	0.885	0.999	0.885
CG	0.885	0.999	0.885	1.0	0.923	0.246
CGI	0.301	0.953	0.999	0.923	1.0	0.836
CO + CG + CGI	<b>0.013</b>	0.301	0.885	0.246	0.836	1.0

Tabla 11. Prueba *post-hoc* de Nemenyi para el Experimento AP con imágenes de lesiones de la piel. Los valores resaltados indican los p-valores más bajos.

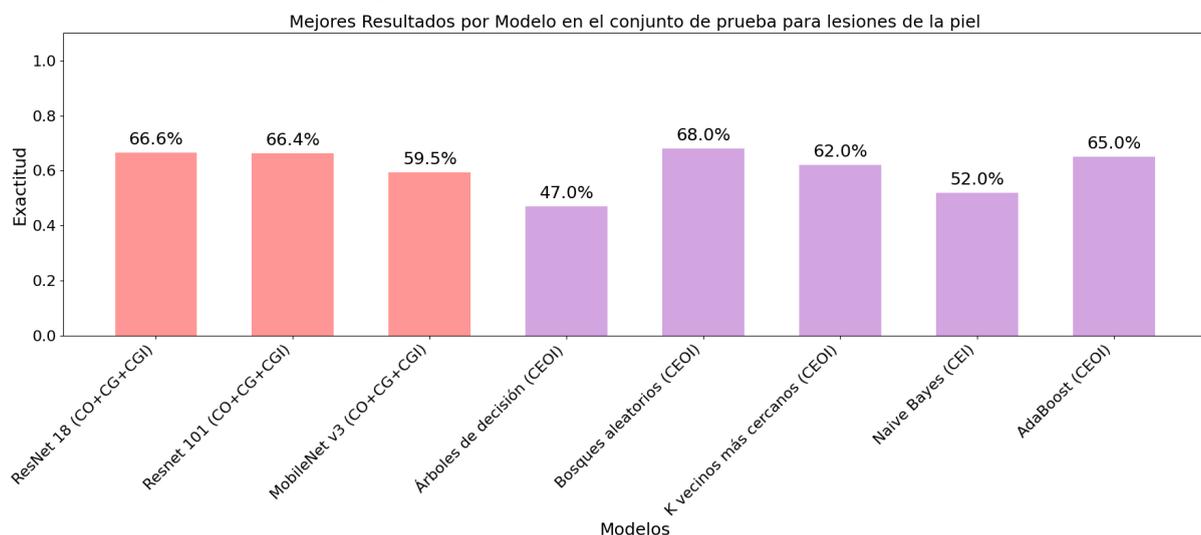
#### 4.3.3. Resultados por modelo (AP y AM)

A continuación se muestran los mejores resultados obtenidos para cada modelo al utilizar los diferentes conjuntos de estudio. Esta visualización de los datos tiene como objetivo realizar una comparación visual de los rendimientos de las mejores combinaciones modelo-conjunto (véase las Figuras 28 y 29).



**Figura 28.** Mejores resultados usando los conjuntos de estudio por cada modelo en el subconjunto de prueba de ultrasonido de cáncer de mama. Barras en **rojo pastel**: modelos de aprendizaje profundo (AP), barras en **morado pastel**: modelos de aprendizaje máquina (AM).

En la comparación entre los experimentos de aprendizaje automático (AM) y aprendizaje profundo (AP) en imágenes de ultrasonido de cáncer de mama, con un análisis visual podemos destacar que el modelo de bosques aleatorios con el conjunto CEO en AM logra un rendimiento del 92%, mientras que en AP el conjunto CGI proporciona los mejores resultados para redes neuronales estudiadas. Por otro lado, al aplicar la prueba estadística de Mann-Whitney U de una cola para evaluar las estrategias basadas en TEN con imágenes (CG, CGI, y CO+CG+CGI) y estilos (CEO, CEI y CEOI), se obtuvo un estadístico de 72.0 y un p-valor de 0.4056. Esto sugiere que no hay suficiente evidencia para rechazar la hipótesis nula, indicando que las diferencias de rendimiento entre las propuestas de AM y AP con TEN no son estadísticamente significativas.



**Figura 29.** Mejores resultados usando los conjuntos de estudio por cada modelo en el subconjunto de prueba de lesiones de la piel. Barras en **rojo pastel**: modelos de aprendizaje profundo (AP), barras en **morado pastel**: modelos de aprendizaje máquina (AM).

Para la base de datos de lesiones de la piel, con un análisis visual se puede observar que el mejor modelo es bosques aleatorios, con el conjunto CEOI en AM que alcanza un rendimiento del 68%, mientras que en AP el conjunto CO+CG+CGI (700 imágenes) ofrece el mejor desempeño. La prueba

de Mann-Whitney U de una cola aplicada a estas estrategias basadas en TEN (dejando de lado los datos originales y los aumentados tradicionalmente) arrojó un estadístico de 59.5 y un p-valor de 0.6934, lo que indica que no hay diferencias estadísticamente significativas en el rendimiento entre las estrategias de AM y AP.

Aunque algunos modelos en AM, como bosques aleatorios con CEOI, muestran un rendimiento destacable, no hay evidencia de que los modelos de AM superen de manera consistente a los modelos de AP cuando se emplean técnicas de TEN. Esto puede sugerir que ambas aproximaciones pueden ser adecuadas y que la elección de la técnica podría depender del tipo de imágenes y la capacidad de cada técnica para resaltar características relevantes en diferentes contextos.

## 5. Conclusiones

En este capítulo se presentan las conclusiones generales y específicas derivadas de los experimentos realizados a lo largo del trabajo. Además, se exploran y sugieren diversos caminos para trabajos futuros, destacando las oportunidades para optimizar estas técnicas en problemas similares.

### 5.1. Conclusiones por experimento

En esta sección, se desglosan las conclusiones específicas obtenidas de cada experimento, proporcionando una revisión de los resultados obtenidos en los experimentos de análisis del estilo neuronal, aprendizaje profundo, y aprendizaje automático.

#### 5.1.1. Análisis del estilo neuronal

Con estos análisis sobre el estilo neuronal, encontramos 2 elementos clave para los siguientes experimentos. El primero es que usar control espacial sobre las lesiones puede ayudar a discriminar los diagnósticos, mientras que no hacerlo no brinda información suficiente para saber si esto es posible. El segundo, es que los estilos neuronales interpolados con control espacial siguen conservando la propiedad anterior y, por lo tanto, pueden utilizarse como una estrategia de aumento de datos para estilos. Por otro lado, estos análisis no solo aportaron estos 2 elementos, sino que también sirvieron como fuente de inspiración para utilizar el estilo como extracción de características, y brinda la idea, para trabajos futuros, de hacer análisis de estilo neuronal para describir conjuntos de imágenes (véase Apéndice A.1).

#### 5.1.2. Experimentos de aprendizaje máquina

Con los resultados obtenidos en estos experimentos (véase Figuras 24 y 25), se observa que la extracción de estilo de la lesión aporta buenos resultados, especialmente en la base de datos de ultrasonido de cáncer de mama, donde se alcanzan niveles de exactitud superiores al 90%. Los análisis de significancia estadística, realizados mediante la prueba de Friedman, revelaron que para ambas bases de datos, se rechaza la hipótesis nula, indicando que por los menos 2 conjuntos de estudio tienen diferentes exactitudes. En ambas bases, el conjunto CEOI, que combina estilos neuronales y estilos interpolados, muestra una diferencia estadísticamente significativa frente a los conjuntos CO aplanados y CO aplanados con máscara (clasificación a nivel de píxel). Esto sugiere que la estrategia de CEOI es prometedora y justifica un estudio más exhaustivo en investigaciones futuras.

Además, es recomendable como trabajo futuro comparar este método propuesto con técnicas avanzadas de extracción de características para evaluar su efectividad en un contexto competitivo. Aun así, la estrategia propuesta, junto con la interpolación de estilos, representa un enfoque interesante y con gran potencial para trabajos futuros en conjuntos de datos pequeños.

### 5.1.3. Experimentos de aprendizaje profundo

Para este conjunto de experimentos, surgen dos panoramas al utilizar las dos bases de datos. En el análisis visual de la base de datos de imágenes de ultrasonido, observamos que las estrategias de aumento de datos mediante Transferencia de Estilo Neuronal (TEN) con control espacial sobre las lesiones entre imágenes del mismo diagnóstico, así como el incremento de estilos mediante estilos interpolados, muestran un mejor desempeño en comparación con el uso de solo las imágenes originales o el aumento de datos tradicional. Sin embargo, los resultados obtenidos con los conjuntos CO 700 y CO+CG+CGI, que contienen 700 imágenes por diagnóstico, sugieren un límite en la cantidad óptima de imágenes generadas para el entrenamiento, ya que la exactitud parece disminuir al utilizar estos conjuntos en comparación con los conjuntos CG y CGI que tienen 300 imágenes por diagnóstico.

Estas conclusiones se validan al analizar las pruebas de significancia estadística, especialmente al comparar los conjuntos de estudio con aumento de datos contra el conjunto de imágenes originales CO. Inicialmente, la prueba de Friedman indica que se rechaza la hipótesis nula, lo que significa que al menos dos conjuntos presentan diferencias significativas en sus exactitudes. En la prueba *post-hoc* de Nemenyi, aunque no hay pares de grupos con un p-valor menor a 0.05, la comparación entre CGI y CO muestra una diferencia que es estadísticamente significativa o cercana al umbral de significancia, y algo similar ocurre al comparar CG con CO, que tiende a una diferencia significativa. Por el contrario, los conjuntos CO 300, CO 700 y CO+CG+CGI son muy parecidos al conjunto CO, lo cual es consistente con las observaciones del análisis visual.

Los resultados para la base de datos de lesiones de la piel ofrecen un caso particularmente interesante, a pesar de que el desempeño de los modelos es bajo en general. Este bajo rendimiento podría estar relacionado con las limitaciones en la generación de imágenes en una base de datos que abarca seis diagnósticos diferentes. En el análisis visual, se observa que los conjuntos con 700 imágenes por diagnóstico alcanzan una mayor exactitud en comparación con los conjuntos de 300 imágenes por diagnóstico, lo que sugiere la posibilidad de realizar un aumento de datos adicional. Una evidencia de esto es que los conjuntos generados con TEN, CG y CGI muestran mejoras mínimas en comparación con CO 300, mientras que CO+CG+CGI aumenta la exactitud en más de un 7% respecto a CO 700. Esto sugiere que la naturaleza de las imágenes y la cantidad de diagnósticos contribuyen significativamente a la complejidad de la base de datos.

Esta interpretación visual de los resultados se puede apoyar en el análisis estadístico. La prueba de Friedman rechaza la hipótesis nula, indicando que al menos dos conjuntos presentan diferencias significativas en la exactitud. Además, la prueba *post-hoc* de Nemenyi identifica que los únicos conjuntos con una diferencia estadísticamente significativa son CO y CO+CG+CGI, mientras que no se observan diferencias claras entre los demás conjuntos, lo cual refuerza las observaciones visuales.

Un experimento interesante para estudiar la baja exactitud en el conjunto de lesiones de la piel, que puede deberse a la complejidad de esta base de datos, sería reducir la cantidad de diagnósticos basándose en la naturaleza de las imágenes, es decir, realizar un análisis de estilos sobre las imágenes

(véase la figura 17), seleccionar los diagnósticos que menos se parezcan entre sí y entrenar los modelos (véase Apéndice A.1).

#### 5.1.4. Comparación de mejores Modelos-Conjunto

Con una inspección visual, los resultados donde se comparan todos los modelos de clasificación con el mejor modelo (véase las Figuras 28 y 29), muestran que los modelos de bosques aleatorios y AdaBoost, en combinación con los conjuntos de estudio CEOI y CEO, logran los mejores rendimientos en la clasificación para ambas bases de datos. No obstante, los enfoques de aprendizaje profundo con aumento de imágenes también muestran resultados alentadores, como los de las ResNet 18, aunque necesitan un estudio más profundo sobre el comportamiento, naturaleza y complejidad de las bases de datos.

Las pruebas de Mann-Whitney U no muestran una diferencia significativa en el rendimiento entre las técnicas de aprendizaje automático (AM) y aprendizaje profundo (AP) para ambas bases de datos, lo que sugiere que, aunque algunos modelos de AM obtienen buenos resultados con estilos neuronales, no hay evidencia suficiente para afirmar que estos enfoques superen consistentemente a los modelos de AP en términos estadísticos. Por lo que, estos hallazgos sugieren dos enfoques prometedores:

1. La extracción de características mediante TEN y estilos interpolados, combinados con el modelo de aprendizaje máquina más adecuado como bosques aleatorios o AdaBoost, puede alcanzar buenos resultados.
2. La generación de imágenes sintéticas utilizando TEN y estilos interpolados puede mejorar significativamente el rendimiento de los modelos de aprendizaje profundo, en especial cuando los modelos son pequeños como MobileNet.

## 5.2. Conclusiones generales

Los enfoques presentados en este trabajo son el resultado de numerosos experimentos que, aunque en su mayoría no tuvieron éxito y no se incluyeron en este documento, aportaron nuevas perspectivas sobre la transferencia de estilo neuronal, el aumento de datos, la extracción de características y el uso óptimo de modelos de inteligencia artificial para la resolución de problemas. Aunque el objetivo inicial era aplicar únicamente el aumento de datos mediante transferencia de estilo neuronal (TEN) en imágenes médicas, la investigación y los experimentos realizados permitieron mejorar y ampliar el enfoque original.

Entre las principales contribuciones de este trabajo se destacan:

- La implementación de TEN con control espacial sobre lesiones en imágenes médicas para aumentar los datos y mejorar modelos de clasificación.
- La exploración del comportamiento de los datos con estilos neuronales, tanto con como sin control espacial.
- La propuesta de un método de interpolación de estilos para incrementar la diversidad estilística, superando las limitaciones en la cantidad y variedad de imágenes disponibles.

- El uso de los estilos de las imágenes originales y los estilos interpolados como datos de entrada para los modelos, aprovechando sus características para mejorar el rendimiento de los modelos de aprendizaje automático.

Los resultados obtenidos demuestran que el aumento de datos utilizando transferencia de estilo es factible. El potencial de esta técnica para la ampliación de datos depende tanto del estilo como del contenido. Si bien el estilo permite modificar las imágenes para generar más variantes estilísticas, la falta de suficientes imágenes de contenido puede llevar a que los modelos de aprendizaje profundo se ajusten al ruido y a los artefactos, provocando un sobreajuste.

Una de las propuestas más destacables es la interpolación de estilos, que permite generar una gran cantidad de estilos objetivo, prácticamente ilimitados. Los resultados sugieren que esta estrategia puede mejorar el rendimiento de los modelos de aprendizaje automático y aprendizaje profundo; sin embargo, es esencial contar con suficientes imágenes de contenido para evitar el sobreajuste. Además, el uso del estilo neuronal como técnica de extracción de características ha demostrado potencial para mejorar los resultados en modelos de aprendizaje automático.

En resumen, este trabajo se centró en analizar y utilizar la transferencia de estilo neuronal para estudiar su impacto en dos pequeños conjuntos de imágenes médicas. Más allá de abordar únicamente el problema de la escasez de datos para entrenar modelos de aprendizaje profundo y aprendizaje automático en ciertos tipos de imágenes médicas, este estudio revela un área prometedora que abre nuevas posibilidades sobre cómo emplear técnicas de transferencia de estilo para resolver diversos retos en la creación de modelos de inteligencia artificial.

### 5.3. Trabajo futuro

A partir del trabajo realizado y de las observaciones obtenidas en los experimentos de análisis y clasificación, surgen varias direcciones interesantes para futuros trabajos. Estas sugerencias incluyen tanto mejoras metodológicas como del análisis estadístico en todas las fases experimentales:

- Caracterización y análisis de estilo en bases de imágenes: explorar técnicas de análisis de estilo y reducción de dimensionalidad para caracterizar bases de datos de imágenes de alta dimensión.
- Optimización del número de imágenes generadas: determinar el límite óptimo de imágenes que se pueden generar en función del número de imágenes de estilo y contenido. Se proponen dos estrategias:
  - Realizar una serie de experimentos, variando el número de imágenes de contenido y generadas para identificar la cantidad óptima que maximice el rendimiento en redes neuronales profundas.
  - Definir una imagen de contenido base y evaluar el número de estilos necesarios para generar imágenes efectivas, basándose en el impacto en el rendimiento de las redes neuronales profundas.

- Avances en técnicas de TEN: incorporar técnicas más recientes de transferencia de estilo neuronal que sean más rápidas y/o conserven mejor las características de las imágenes originales. Además, se puede explorar una interpolación o aumento de estilos que resulte en imágenes de mayor calidad y diversidad, maximizando su impacto en el rendimiento de los modelos.
- Estudio del estilo como extracción de características:
  - Comparar el uso del estilo frente al contenido como técnica de extracción de características para evaluar cuál enfoque es más efectivo en distintas tareas.
  - Implementar métodos de selección de características para reducir el tamaño de los vectores de estilo, optimizando su utilidad y reduciendo la complejidad computacional.
  - Probar alternativas a las matrices *Gram* para la obtención de estilo, considerando técnicas que puedan capturar otros aspectos importantes de la textura y apariencia.
- Comparación con otras técnicas generativas: evaluar y comparar el rendimiento de GANs, modelos de difusión y TEN en la generación de imágenes sintéticas y en la mejora de modelos de clasificación. Además, experimentar con combinaciones de GAN y modelos de difusión junto con métodos de TEN, evaluando el impacto en la calidad de las imágenes y en el uso de memoria.

## A. Apéndice

### A. 1. Experimento de aprendizaje profundo sobre solo 3 diagnósticos del conjunto *HAM 10K*

El propósito de este experimento es evaluar el rendimiento de varios modelos de aprendizaje profundo (AP) aplicados a la base de datos *HAM 10K* [58], pero limitando la clasificación a solo tres diagnósticos seleccionados por su clara diferenciación entre ellos con el análisis de los estilos neuronales. Los diagnósticos elegidos fueron Carcinoma de células basales (BCC), Melanoma (MEL) y Lesiones vasculares (VASC), ya que como se ilustra en la Figura 17, estos pueden considerarse diferentes entre sí cuando se analiza el estilo con control espacial.

#### A. 1. 1. Hipótesis

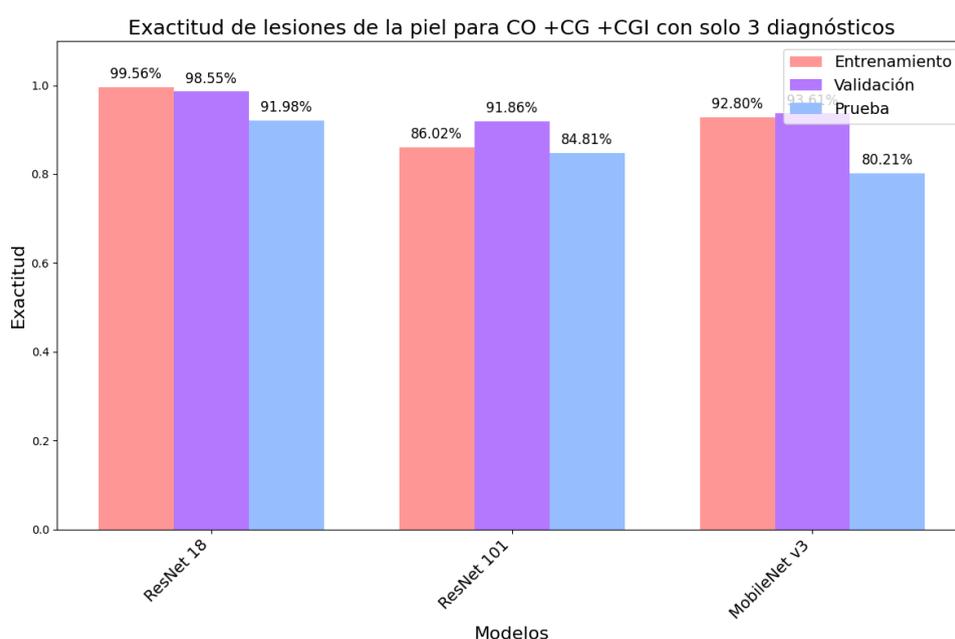
La disminución de la complejidad en la clasificación, mediante el análisis de estilos neuronales con control espacial, mejorará el rendimiento de los modelos de aprendizaje profundo

#### A. 1. 2. Metodología

Este análisis sigue la estructura del experimento descrito en la Figura 15, pero se simplifica para estudiar la hipótesis seleccionando solo imágenes de los diagnósticos BCC, MEL y VASC del conjunto CO+CG+CGI. El subconjunto de prueba utilizado corresponde al descrito en la Tabla 7, pero restringido exclusivamente a estas tres clases. El conjunto de entrenamiento, por tanto, se denominó "CO+CG+CGI solo 3".

Posteriormente, se entrenaron las redes ResNet 18, ResNet 101 y MobileNet utilizando validación cruzada con  $k = 5$ . La métrica principal evaluada fue la exactitud en cada caso.

#### A. 1. 3. Resultados de exactitud para *HAM 10K* con solo 3 diagnósticos



#### A. 1. 4. Discusión y conclusiones

Los resultados obtenidos muestran que los modelos de aprendizaje profundo alcanzan una exactitud superior al 90% para los tres diagnósticos seleccionados, lo que representa una mejora significativa en comparación con el mejor rendimiento obtenido al clasificar los seis diagnósticos. Esta mejora puede atribuirse a una reducción en la complejidad de la tarea, ya que al disminuir el número de diagnósticos, los modelos pueden enfocarse mejor en las características particulares de cada uno.

Un punto a resaltar, es que los resultados obtenidos en los experimentos anteriores (véase Figura 27), los modelos para clasificar 6 diagnósticos de la base de datos *HAM 10K*, sugieren que se requiere un mayor número de imágenes para alcanzar buenos resultados. Sin embargo, en este experimento, combinando técnicas como el aumento de datos por la transferencia de estilo neuronal , junto con la reducción de clases, se logra mejorar el rendimiento sin necesidad de incrementar significativamente el tamaño del conjunto de imágenes.

## 6. Bibliografía

1. C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, 2019, doi: 10.1186/s40537-019-0197-0.
2. P. K. Mall *et al.*, "A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities," *Healthcare Analytics*, vol. 4, p. 100216, 2023, doi: <https://doi.org/10.1016/j.health.2023.100216>.
3. R. Ceron and R. Ceron, "AI, machine learning and deep learning: What's the difference?," *IBM Blog*, Mar. 8, 2021. [En línea]. Disponible en: <https://www.ibm.com/blog/ai-machine-learning-and-deep-learning-whats-the-difference>.
4. W. H. L. Pinaya *et al.*, "Generative AI for Medical Imaging: extending the MONAI Framework," Jul. 2023, doi: <https://doi.org/10.48550/arXiv.2307.15208>.
5. L. A. Gatys, A. S. Ecker, and M. Bethge, "A Neural Algorithm of Artistic Style", Agos. 2015, doi: <https://doi.org/10.48550/arXiv.1508.06576>.
6. L. A. Gatys, A. S. Ecker, M. Bethge, A. Hertzmann, and E. Shechtman, "Controlling Perceptual Factors in Neural Style Transfer," Nov. 2016, doi: <https://doi.org/10.48550/arXiv.1611.07865>.
7. A. Mikołajczyk and M. Grochowski, "Style transfer-based image synthesis as an efficient regularization technique in deep learning," 2019, doi: <https://doi.org/10.48550/arXiv.1905.10974>.
8. Z. Liu *et al.*, "Remove Appearance Shift for Ultrasound Image Segmentation via Fast and Universal Style Transfer," Feb. 2020, doi: <https://doi.org/10.48550/arXiv.2002.05844>.
9. Z. Liu *et al.*, "Generalize Ultrasound Image Segmentation via Instant and Plug & Play Style Transfer," Jan. 2021, doi: <https://doi.org/10.48550/arXiv.2101.03711>.
10. Y. Li, R. Zhu, A. Qu, and M. Yeh, "Dermoscopic Image Classification with Neural Style Transfer," May 2021, doi: <https://doi.org/10.48550/arXiv.2105.07592>.
11. M. Haenlein and A. Kaplan, "A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence," *California Management Review*, vol. 61, p. 000812561986492, Jul. 2019, doi: 10.1177/0008125619864925.

12. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998, doi: 10.1109/5.726791.
13. I. J. Goodfellow *et al.*, "Generative Adversarial Networks," Jun. 2014, doi: <https://doi.org/10.48550/arXiv.1406.2661>.
14. V. Kaul, S. Enslin, and S. A. Gross, "History of artificial intelligence in medicine," *Gastrointestinal Endoscopy*, vol. 92, no. 4, pp. 807–812, Oct. 2020, doi: 10.1016/J.GIE.2020.06.040.
15. E. H. Shortliffe y B. G. Buchanan, "A model of inexact reasoning in medicine," *Mathematical Biosciences*, vol. 23, núms. 3–4, pp. 351–379, 1975. doi: 10.1016/0025-5564(75)90047-4.
16. C. A. Kulikowski y S. M. Weiss, "Representation of Expert Knowledge for Consultation: The CASNET and EXPERT Projects," en *Artificial Intelligence in Medicine*, P. Szolovits, Ed. Boulder, Colorado: Westview Press, 1982 Disponible en: <https://groups.csail.mit.edu/medg/ftp/psz/AIM82/ch2.html>.
17. O. Ronneberger, P. Fischer, y T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," en *Proceedings of MICCAI*, 2015. doi: 10.1007/978-3-319-24574-4\_28.
18. A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, y S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115-118, 2017. doi: 10.1038/nature21056.
19. E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nature Medicine*, vol. 25, no. 1, pp. 44-56, 2019. doi: 10.1038/s41591-018-0300-7.
20. T. M. Mitchell, *Machine Learning*. New York: McGraw-Hill, 1997. doi: 10.5555/541177.
21. K. P. Murphy, *Probabilistic Machine Learning: An Introduction*, Cambridge, MA: MIT Press, 2022, pp. 1-18. ISBN 9780262046824.
22. S. Surabhi, "Get started with Naive Bayes Algorithm: Theory & implementation," *Analytics Vidhya*, Jul. 14, 2023. [En línea]. Disponible en: <https://www.analyticsvidhya.com/blog/2021/01/a-guide-to-the-naive-bayes-algorithm/>.
23. R. D. Singh, "What is the time complexity of training a Bayesian Classifier?" LinkedIn, Nov. 3, 2023. [En línea]. Disponible en: <https://www.linkedin.com/pulse/what-time-complexity-training-bayesian-classifier-r-d-singh-upqpf/>.
24. V. Jayaswal, "Understanding Naïve Bayes Algorithm," *Towards Data Science*, 8 Nov. 2020. [En línea]. Disponible en: <https://towardsdatascience.com/understanding-naïve-bayes-algorithm-f9816f6f74c0>.
25. T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967, doi: 10.1109/TIT.1967.1053964.

26. K. P. Murphy, *Probabilistic Machine Learning: An Introduction*, Cambridge, MA: MIT Press, 2022, pp. 603-6023. ISBN 9780262046824.
27. H. M. Sani, C. Lei, and D. Neagu, "Computational complexity analysis of decision tree algorithms," in *Artificial Intelligence XXXV*, M. Bramer and M. Petridis, Eds. Cham: Springer, 2018, pp. 191–197. doi: 10.1007/978-3-030-04191-5\_17.
28. "What is a Random Forest?," *NVIDIA Data Science Glossary*. [En línea]. Disponible en: <https://www.nvidia.com/en-us/glossary/random-forest/>.
29. A. Anshul, "Guide on AdaBoost Algorithm," *Analytics Vidhya*, 2024. [En línea]. Disponible en: <https://www.analyticsvidhya.com/blog/2021/09/adaboost-algorithm-a-complete-guide-for-beginners/>.
30. IBM, "What is a Confusion Matrix?" IBM, Jan. 19, 2024. [Online]. Available: <https://www.ibm.com/topics/confusion-matrix>.
31. A. Bhandari, "Confusion Matrix in Machine Learning," *Analytics Vidhya*, Aug. 28, 2024. [Online]. Available: <https://www.analyticsvidhya.com/blog/2020/04/confusion-matrix-machine-learning/>.
32. K. P. Murphy, *Probabilistic Machine Learning: An Introduction*, Cambridge, MA, USA: MIT Press, 2022, pp. 125-126.
33. P. E. Black, "Euclidean distance", in *Dictionary of Algorithms and Data Structures* [online], P. E. Black, ed. 17, Dic. 2004. [En línea]. Disponible en: <https://www.nist.gov/dads/HTML/euclidndstnc.html>.
34. J. J. Berman, "Chapter 4 - Understanding Your Data," in *Data Simplification*, J. J. Berman, Ed., Boston: Morgan Kaufmann, 2016, pp. 135–187. doi: <https://doi.org/10.1016/B978-0-12-803781-2.00004-7>.
35. T. Bock, "What is a Correlation Matrix?," Displayr, Jan. 2, 2024. [En línea]. Disponible en: <https://www.displayr.com/what-is-a-correlation-matrix/>.
36. "What is Dimensionality Reduction?," IBM, [En línea]. Disponible en: <https://www.ibm.com/topics/dimensionality-reduction>.
37. K. P. Murphy, *Probabilistic Machine Learning: An Introduction*, Cambridge, MA, USA: MIT Press, 2022, pp. 665–702.
38. A. Maćkiewicz and W. Ratajczak, "Principal components analysis (PCA)," *Computers & Geosciences*, vol. 19, no. 3, pp. 303–342, 1993, doi: [https://doi.org/10.1016/0098-3004\(93\)90090-R](https://doi.org/10.1016/0098-3004(93)90090-R).
39. F. Anowar, S. Sadaoui, and B. Selim, "Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE)," *Computer Science Review*, vol. 40, p. 100378, 2021, doi: <https://doi.org/10.1016/j.cosrev.2021.100378>.

40. M. A. Nielsen, *Neural Networks and Deep Learning*, Determination Press, 2015. [En línea]. Disponible en: <http://neuralnetworksanddeeplearning.com/index.html>.
41. Z. Zhang, A. Lipton, M. Li, and A. J. Smola, *Dive into Deep Learning*, 2023. [En línea]. Disponible en: <https://d2l.ai/index.html>.
42. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2015. doi: <https://doi.org/10.48550/arXiv.1409.1556>.
43. A. Géron, *Hands-On Machine Learning With Scikit-Learn and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd ed., Sebastopol, CA, USA: O'Reilly Media, 2019, pp. 456-459.
44. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv preprint arXiv:1512.03385*, 2015. doi: <https://doi.org/10.48550/arXiv.1512.03385>.
45. A. Howard *et al.*, "Searching for MobileNetV3," *arXiv preprint arXiv:1905.02244*, 2019. doi: <https://doi.org/10.48550/arXiv.1905.02244>.
46. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, Jun. 2009, pp. 248-255, doi: [10.1109/CVPR.2009.5206848](https://doi.org/10.1109/CVPR.2009.5206848).
47. K. P. Murphy, *Probabilistic Machine Learning: An Introduction*, Cambridge, MA, USA: MIT Press, 2022, pp. 628–630.
48. "What is transfer learning?," IBM, [En línea]. Disponible en: <https://www.ibm.com/topics/transfer-learning>.
49. L. A. Gatys, A. S. Ecker and M. Bethge, "Image Style Transfer Using Convolutional Neural Networks," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 2414-2423, doi: [10.1109/CVPR.2016.265](https://doi.org/10.1109/CVPR.2016.265).
50. T. Henighan, "Spatial Control in Neural Style Transfer." Stanford Physics.[En línea]. Disponible en: <https://tomhenighan.com/pdfs/cs231n-project.pdf>.
51. T. Lin *et al.*, "Drafting and revision: Laplacian pyramid network for fast high-quality artistic style transfer," *arXiv preprint arXiv:2104.05376*, 2021. doi: <https://doi.org/10.48550/arXiv.2104.05376>.
52. N. Kolkin, J. Salavon, and G. Shakhnarovich, "Style transfer by relaxed optimal transport and self-similarity," *arXiv preprint arXiv:1904.12785*, 2019. doi: <https://doi.org/10.48550/arXiv.1703.06868>.
53. X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," *arXiv preprint arXiv:1703.06868*, 2017. doi: <https://doi.org/10.48550/arXiv.1703.06868>.
54. M. U. Akbar, W. Wang, and A. Eklund, "Beware of diffusion models for synthesizing medical images -- A comparison with GANs in terms of memorizing brain MRI and chest x-ray images," May 2023, doi: <https://doi.org/10.48550/arXiv.2305.07644>.

55. A. Staff, "GANs vs. Diffusion Models: Putting AI to the test," Aurora Solar, 18 de abril de 2024. [En línea]. Disponible: <https://aurorasolar.com/blog/putting-ai-to-the-test-generative-adversarial-networks-vs-diffusion-models/>. [Accedido: 22 de septiembre de 2024].
56. R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2022, pp. 10674-10685.
57. W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data in Brief*, vol. 28, p. 104863, 2020, doi: <https://doi.org/10.1016/j.dib.2019.104863>.
58. P. Tschandl, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Harvard Dataverse*, 2018, doi: 10.7910/DVN/DBW86T.
59. G. D. Finlayson and E. Trezzi, "Shades of Gray and Colour Constancy," 2004.
60. SciPy, "GitHub - scipy/scipy: SciPy library main repository," GitHub. [En línea]. Disponible en: <https://github.com/scipy/scipy>.
61. F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
62. Martín~Abadi *et al.*, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems." 2015. [En línea]. Disponible en: <https://www.tensorflow.org/>.
63. Gargimahale, "GitHub - gargimahale/Doodle: 🎨 Semantic Style Transfer and Turning Two-Bit Doodles into Fine Artworks," GitHub. [En línea]. Disponible en: <https://github.com/gargimahale/Doodle>.
64. Fastai, "GitHub - fastai/fastai: The fastai deep learning library," GitHub. [En línea]. Disponible en: <https://github.com/fastai/fastai>.
65. PyTorch, "GitHub - pytorch/pytorch: Tensors and Dynamic neural networks in Python with strong GPU acceleration," GitHub. [En línea]. Disponible en: <https://github.com/pytorch/pytorch>.