

Eliminación de ruido (regla del visor) en imágenes dermatoscópicas

Autores :

Amanda Noris Hernández
Juan Miguel Pérez Martínez
Marcos Antonio Pérez Lorenzo

July 8, 2024



1 Introducción

El cáncer de piel representa uno de los principales problemas de salud pública en muchas regiones del mundo, caracterizado por su alta incidencia y la necesidad de diagnósticos tempranos para mejorar significativamente las tasas de supervivencia. La detección temprana del cáncer de piel es crucial para prevenir complicaciones graves y garantizar tratamientos efectivos. Sin embargo, el proceso de identificación temprana de lesiones cutáneas sospechosas puede ser desafiante debido a la variabilidad en la apariencia de estas lesiones y la experiencia requerida para realizar un diagnóstico preciso.

En este contexto, la Inteligencia Artificial (IA) emerge como una herramienta prometedora para abordar estos desafíos. Las técnicas de aprendizaje automático, un subcampo de la IA, han demostrado capacidades excepcionales en la interpretación de datos complejos y la identificación de patrones sutiles, lo que las hace ideales para la detección automatizada del cáncer de piel.

1.1 Motivación

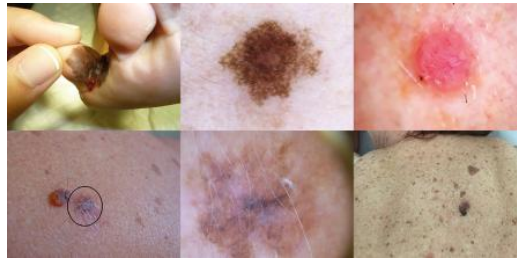


Figure 1: Ejemplos de cáncer

Con el aumento alarmante de casos de cáncer de piel a nivel mundial, la necesidad de métodos de diagnóstico más eficientes y precisos se ha vuelto urgente. Este incremento no solo pone de relieve la importancia de la detección temprana, sino también la necesidad de herramientas que puedan procesar rápidamente grandes volúmenes de imágenes dermatoscópicas para identificar lesiones potencialmente malignas. En respuesta a esta crisis, ha habido un notable auge en el desarrollo y la aplicación de modelos de clasificación automática basados en Machine Learning (ML), entre ellos destacamos el desarrollado en la Facultad de Matemática y Computación de la Universidad

de la Habana: Derma UH. Estos modelos prometen revolucionar la forma en que se aborda el diagnóstico del cáncer de piel, ofreciendo la posibilidad de realizar diagnósticos más rápidos y precisos. Sin embargo, el desafío radica en las limitaciones actuales de su uso en la práctica clínica diaria, donde la selección cuidadosa de imágenes adecuadas para el análisis por cómputo es esencial para garantizar la precisión del diagnóstico. La variabilidad en la captura de imágenes, influenciada por factores como la diversidad de los modelos de dermatoscopios y las condiciones de iluminación, puede introducir sesgos y reducir la fiabilidad de los modelos de clasificación. Por lo tanto, la principal motivación detrás de este trabajo es mejorar el proceso de clasificación automática para superar estas barreras, asegurando que los modelos existentes sean capaces de manejar la complejidad y la variabilidad inherentes en las imágenes dermatoscópicas, y proporcionar una herramienta confiable y eficaz para los profesionales de la salud en la lucha contra el cáncer de piel.

1.2 Problemática



Figure 2: Dermatoscopio del modelo DermaLite

Los dermatoscopios son dispositivos portátiles que han revolucionado la forma en que los dermatólogos examinan, diagnostican y tratan las afecciones de la piel. Mediante la iluminación y magnificación de la piel, estos dispositivos permiten a los profesionales de la salud descubrir detalles ocultos de las lesiones cutáneas y los cánceres, conduciendo a diagnósticos más precisos y tratamientos específicos.

La regla del visor en los dermatoscopios es una característica funcional que permite a los usuarios medir las lesiones cutáneas con precisión. Esta

regla, que suele estar incorporada en el dispositivo de manera magnética, es esencial para evaluar el tamaño de las lesiones y seleccionar el puntal adecuado para una biopsia. La presencia de esta regla facilita la toma de decisiones informadas durante el examen de las lesiones, asegurando que los procedimientos posteriores, como la biopsia, se realicen con la mayor precisión posible.

Aunque es indispensable su presencia en la herramienta, las diferencias en los estilos de regla entre distintos modelos de dermatoscopios pueden introducir variabilidad innecesaria en las imágenes, dificultando la tarea del clasificador. Por ejemplo, la longitud, el grosor y la posición de la regla pueden variar entre dispositivos, lo que puede afectar la forma en que las lesiones aparecen en las imágenes. Este tipo de variabilidad puede llevar a confusiones en el modelo de clasificación, especialmente si se entrenan con imágenes de un estilo de regla específico y luego se prueban con imágenes de otro estilo.

1.3 Objetivos

Al incorporar un paso de preprocesamiento que elimine o estandarice la regla del visor, se tiene como objetivo principal mejorar la precisión en la detección del cáncer de piel mediante la optimización del proceso de clasificación automática de imágenes dermatoscópicas, lo que se traduce a:

1. **Reducir la variabilidad:** Eliminar o estandarizar la regla del visor reduce la variabilidad en las imágenes, creando un conjunto de datos más uniforme y coherente para el entrenamiento y prueba de los modelos.
2. **Mejorar la precisión:** Al trabajar con imágenes más uniformes y consistentes, los modelos pueden realizar clasificaciones más precisas, lo que resulta en un diagnóstico más exacto del cáncer de piel.
3. **Aumentar portabilidad:** Los modelos optimizados para manejar imágenes estandarizadas son menos dependientes de las características específicas de un dispositivo dermatoscópico, lo que aumenta su portabilidad y utilidad en diferentes contextos clínicos.

1.4 Hipótesis y preguntas científicas

Nuestra hipótesis central sostiene que la adición de un paso de preprocesamiento en el que se elimine o estandarice la regla del visor en las imágenes dermatoscópicas puede incrementar significativamente la precisión en la clasificación de lesiones cutáneas por parte de los modelos de Machine Learning (ML). Este enfoque se fundamenta en la premisa de que la variabilidad introducida por la regla del visor, que puede variar considerablemente entre diferentes modelos de dermatoscopios, actualmente representa un factor de ruido que puede interferir con la capacidad de los modelos ML para identificar patrones y características clave en las imágenes. Al eliminar o estandarizar esta variabilidad, se busca crear un conjunto de datos más limpio y uniforme que pueda ser procesado de manera más efectiva por los algoritmos de ML, lo que, a su vez, debería traducirse en una mayor precisión en la clasificación de las lesiones cutáneas. Esta hipótesis se basa en la idea de que la mejora en la calidad y consistencia de las imágenes es un paso crucial para optimizar el rendimiento de los modelos ML en la detección temprana del cáncer de piel, un área de gran relevancia para la salud pública global.

2 Estado del arte

2.1 Tipo de problema

La traducción de imagen a imagen es una técnica de inteligencia artificial generativa que permite transformar una imagen fuente en una imagen objetivo mientras se preservan ciertas propiedades visuales de la imagen original. Esta tecnología emplea técnicas de aprendizaje automático y redes neuronales profundas, como las redes generativas adversarias (GANs), redes adversarias condicionales (cGANs) y redes neuronales convolucionales (CNNs), para aprender funciones de mapeo complejas entre imágenes de entrada y salida. La traducción de imagen a imagen facilita la conversión de imágenes de un formato a otro manteniendo características esenciales. El objetivo es aprender un mapeo entre los dos dominios y luego generar imágenes realistas en el estilo elegido por el diseñador. Este enfoque permite tareas como transferencia de estilo, colorización y super-resolución, una técnica que mejora la resolución de una imagen. La tecnología de traducción de imagen abarca una amplia gama de aplicaciones en arte, compromiso con la imagen, aumento de datos y visión por computadora, también conocida como visión

por máquina. Por ejemplo, permite a los fotógrafos cambiar una foto de día a noche, convertir una imagen satelital en un mapa y mejorar imágenes médicas para permitir diagnósticos más precisos.

2.2 Avances en el campo

Una de las principales innovaciones en este campo ha sido el desarrollo de arquitecturas como U-Net, que se especializa en la segmentación de imágenes mediante la combinación de conexiones laterales y capas de expansión/squeeze. Esto permite una mayor precisión al localizar objetos específicos dentro de una imagen.

Otro estudio relevante es "Image-to-Image Translation with Conditional Adversarial Networks", publicado en 2017 por Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, y Alexei A. Efros. Este paper introdujo el concepto de redes condicionales adversarias (cGANs), que permiten realizar traducciones de imágenes más precisas al condicionar la generación de imágenes a ciertas características o condiciones específicas. Por ejemplo, pueden transformar imágenes de día a noche o viceversa, manteniendo intactos otros aspectos relevantes de la imagen original.

Pix2Pix fue uno de los primeros en abordar la traducción de imágenes a imágenes de manera supervisada, utilizando una red condicional generativa adversaria (cGAN) para aprender un mapeo directo entre pares de imágenes.

Trabajos posteriores como "High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs" han extendido el concepto inicial de Pix2Pix, explorando cómo las cGANs pueden ser utilizadas para manipular imágenes de manera semántica, permitiendo cambios precisos en elementos específicos de una imagen sin afectar otras partes.

El estudio "Unsupervised Cross-Domain Single View to Multi-View Translation" ha aplicado principios similares a Pix2Pix para traducir imágenes entre diferentes vistas en un solo dominio, demostrando la versatilidad de los métodos de traducción de imágenes a imágenes no supervisadas.

Además, el trabajo "BigGAN: Large-Scale Generative Adversarial Networks" de 2018, presentado por Andrew Brock, Jeff Donahue, y Karen Simonyan, marcó un hito en la generación de imágenes realistas a gran escala. BigGAN utiliza arquitecturas profundas y optimización eficiente para producir imágenes de alta calidad, abriendo nuevas posibilidades en la traducción de imágenes.

Un ejemplo reciente es el trabajo "StyleGAN3: Training World Class Animal Face Generators on COCO-Stuff" publicado en 2023, donde los investigadores lograron generar imágenes de animales con una calidad y variedad impresionantes, utilizando técnicas avanzadas de aprendizaje profundo.

El trabajo presentado en "Deep Photo Enhancer: High Quality Enhancement of Real World Photos" utiliza CNN's en lugar de GAN's. Este estudio introduce una red convolucional profunda que mejora la calidad de las fotos reales mediante la aplicación de técnicas de superresolución y restauración. La arquitectura propuesta utiliza múltiples capas convolucionales y submuestreo espacial para aprender representaciones más ricas de las características de las imágenes, permitiendo una mejora significativa en la nitidez y el detalle de las fotos originales.

Otro avance notable es el uso de atenciones convolucionales (Convolutional Attention Mechanisms), que mejoran la capacidad de las CNNs para enfocarse en regiones específicas de la imagen durante la traducción. Esto resulta especialmente útil en aplicaciones donde el contexto o ciertos detalles son críticos para la tarea de traducción.

3 Propuesta de solución

3.1 Red neuronal

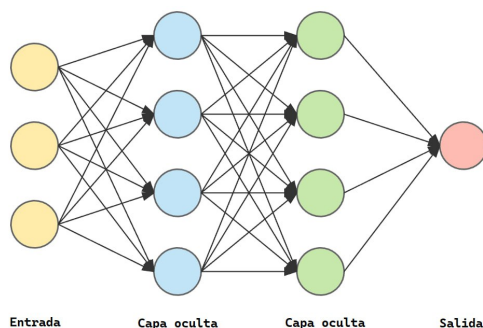


Figure 3: Red neuronal

En el campo del procesamiento de imágenes es bastante común utilizar filtros y elementos estructurantes para eliminar marcas como la regla del

visor, sin embargo, debido a la amplia diferencia de modelos de dermatoscopios existentes y las características únicas de las reglas que posee cada uno, es muy difícil lograr buenos resultados con estas técnicas. En este caso, las redes neuronales parecen ser una alternativa superior por varias razones:

1. **Generación de datos sintéticos:** A diferencia de los filtros y elementos estructurantes, que solo pueden modificar los datos existentes, las redes neuronales pueden generar datos sintéticos que son indistinguibles de los datos reales. Esto significa que pueden crear imágenes que no contienen marcas, incluso si las marcas están presentes en todas las imágenes de entrenamiento.
2. **Aprendizaje profundo:** Las redes neuronales utilizan aprendizaje profundo para aprender representaciones de datos complejas. Esto les permite capturar patrones y relaciones en los datos que podrían ser difíciles de capturar con métodos más tradicionales .
3. **Robustez:** Las redes neuronales son robustas frente a las variaciones en los datos de entrada. Esto significa que pueden funcionar bien incluso cuando las marcas aparecen en diferentes posiciones o tamaños en diferentes imágenes .
4. **Control sobre la calidad de la imagen generada:** Las redes neuronales permiten controlar la calidad de la imagen generada mediante la modificación de la función de pérdida durante el entrenamiento. Esto puede ser útil si se quiere mantener la calidad de la imagen mientras se eliminan las marcas.

3.2 Data set



Figure 4: ISIC Database

Los datos escogidos para el entrenamiento del modelo se extrajeron de una de los data sets más utilizados del mundo de la dermatología: ISIC (International Skin Imaging Collaboration). Esta base de imágenes dermatoscópicas es una colección de imágenes de lesiones cutáneas recogidas por dermatoscopios de todo el mundo. Esta iniciativa internacional se propuso como una plataforma única para fomentar la investigación y el desarrollo de tecnologías de aprendizaje automático aplicadas a la dermatología. La base de datos ISIC contiene miles de imágenes de lesiones cutáneas, incluyendo melanomas, nevos, moles, psoriasis, eccemas y muchas otras afecciones de la piel, todas ellas etiquetadas y anotadas por dermatólogos experimentados. La disponibilidad de esta vasta colección de imágenes ha sido fundamental para el avance de los modelos de Machine Learning en la detección temprana del cáncer de piel, permitiendo a los investigadores entrenar y probar sus algoritmos en un conjunto de datos representativo y diverso. La colaboración internacional que sustenta a ISIC ha hecho posible la acumulación de una de las bases de datos más completas y valiosas en el campo de la dermatología digital, contribuyendo significativamente a la mejora de la precisión y eficacia de los sistemas de diagnóstico automatizado.

3.3 Preparación de los datos de entrenamiento

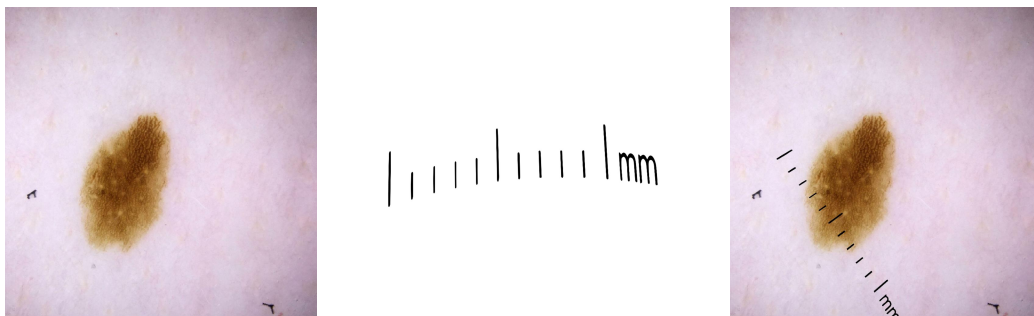


Figure 5: Adición de la mascara

Para abordar el problema de procesamiento de imágenes en nuestro proyecto, se decidió que la entrada ideal de entrenamiento del modelo es una imagen normal acompañada de otra con la regla del visor agregada, especialmente en casos donde las imágenes originales no presentan dicha regla. Se utilizaron dos técnicas clave para mejorar la calidad y la robustez de los datos:

La primera técnica es la fusión de imágenes mediante la superposición de una máscara, lo cual se logra mediante un algoritmo que combina una imagen con una máscara en TensorFlow, ajustando la posición y el tamaño de la máscara dentro de la imagen original. La utilidad de esta técnica radica en su capacidad para agregar variabilidad a los datos de entrenamiento, lo que puede mejorar la capacidad del modelo para generalizar y reconocer objetos bajo diferentes condiciones y orientaciones. La inclusión de una máscara en las imágenes de entrenamiento introduce una nueva dimensión de variabilidad, simulando la aparición de objetos en diferentes colores y formas, lo cual es particularmente útil en tareas de reconocimiento de objetos.

La segunda técnica es la rotación de imágenes (máscaras incluidas) que permite girar imágenes en TensorFlow, cambiando su orientación. La rotación de imágenes es una técnica común en el preprocesamiento de datos para aumentar el tamaño del conjunto de datos y mejorar la robustez del modelo frente a imágenes presentadas en diferentes ángulos. Al rotar las imágenes, se expone al modelo a una variedad de perspectivas, lo que puede ayudarlo a aprender a reconocer objetos independientemente de su orientación en la escena.

Ambas técnicas, la fusión de imágenes con máscaras y la rotación de imágenes, son ejemplos de cómo el preprocesamiento de datos puede ser utilizado para mejorar la calidad y la robustez de los datos de entrenamiento en tareas de procesamiento de imágenes. Estas técnicas permiten introducir una mayor variabilidad en los datos, lo que puede conducir a modelos más precisos y capaces de generalizar mejor a partir de una menor cantidad de datos.

4 Experimentación y resultados

4.1 Red neuronal convolucional

La primera idea escogida para atacar el problema fueron las Redes Neuronales Convolucionales (CNN). Las CNN son especialmente adecuadas para tareas de procesamiento de imágenes debido a su capacidad para extraer características relevantes de las imágenes, como bordes, colores y profundidad, a través de capas ocultas que realizan funciones matemáticas específicas conocidas como convoluciones. Esto permite a las CNN detectar características específicas en una imagen, independientemente de su ubicación, lo que es

crucial para tareas como la clasificación de imágenes, la detección de objetos y la segmentación de imágenes.

Una de las principales ventajas de utilizar CNN para el procesamiento de imágenes es su alta capacidad para reconocer y comprender imágenes de manera similar a como lo hacen los seres humanos. Esto se debe a que las CNN están organizadas en capas convolucionales, donde cada neurona está conectada solo a una región local de la capa anterior, permitiendo que la red detecte características específicas en una imagen. Además, las CNN han demostrado ser extremadamente efectivas en una amplia variedad de aplicaciones de visión por computadora, incluyendo la clasificación de imágenes, la detección de objetos, el seguimiento de objetos y la segmentación de imágenes. Otra razón por la cual es beneficioso atacar problemas de procesamiento de imágenes con redes neuronales es su escalabilidad. Las CNN son altamente escalables y se pueden entrenar en grandes conjuntos de datos utilizando hardware especializado, como unidades de procesamiento gráfico (GPU) o unidades de procesamiento tensorial (TPU). Esto ha permitido a los investigadores y a las empresas abordar problemas de visión por computadora que antes se consideraban imposibles.

Sin embargo las primeras implementaciones de CNN realizadas no brindaron los resultados esperados, en vez de eliminar la regla la resaltaban y borraban el resto de la imagen, todo lo contrario a lo que se deseaba:

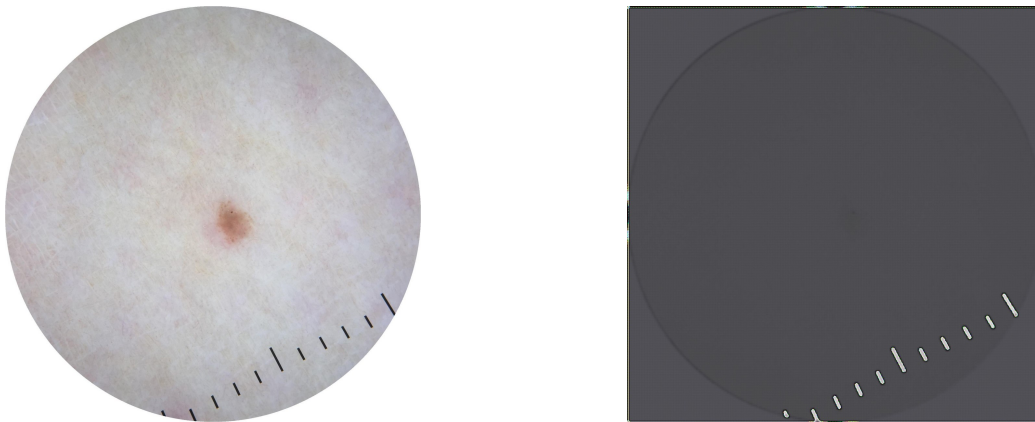


Figure 6: Primer entrenamiento con CNN

Luego de un análisis exhaustivo de las posibles causas de este mal funcionamiento, que tuvo en cuenta principalmente mala interpretación del con-

tenido estudiado en la literatura, errores de implementación y capacidad de cómputo disponible, y de varias jornadas extra de entrenamiento se llegó a la conclusión de que esta mediante esta vía no se iban a obtener los resultados esperados, como se muestra a continuación:

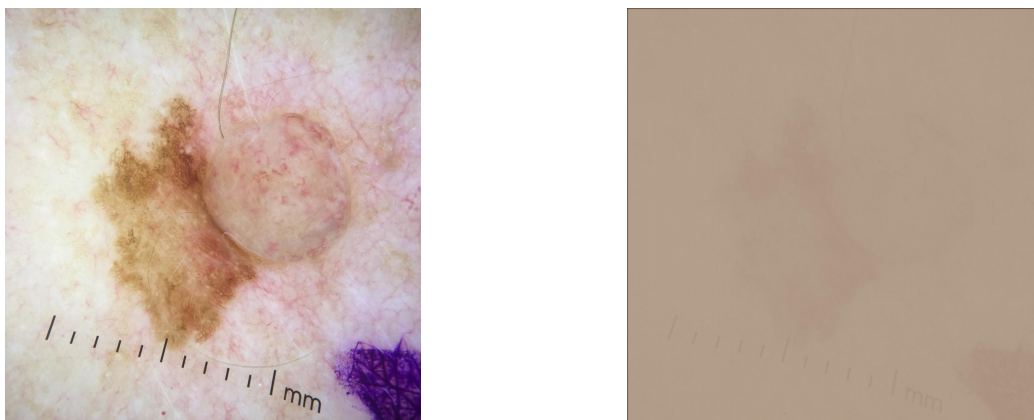


Figure 7: Segundo entrenamiento con CNN

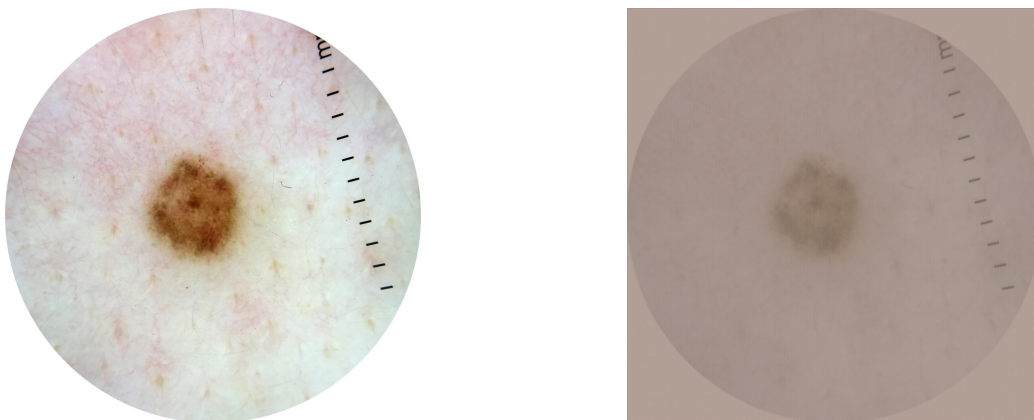


Figure 8: Tercer entrenamiento con CNN

Expongamos entonces una solución diferente, igual basada en redes neuronales, que mejoró muchísimo los resultados y cuya idea se extrajo de uno de los artículos más interesantes de los que se habla en el epígrafe del estado del arte.

4.2 Red neuronal generativa adversarial

Las cGAN son un tipo de modelo de aprendizaje profundo que se diferencia de las GAN tradicionales en que incorpora información condicional durante el proceso de generación. Esto significa que, además de recibir datos de entrada aleatorios, las cGAN también reciben etiquetas o condiciones específicas que guían la generación de datos. De esta forma, el modelo aprende a generar imágenes que coincidan con las características específicas indicadas por las etiquetas, permitiendo un mayor control sobre los resultados generados.

Dentro de una cGAN, hay dos componentes principales: el generador y el discriminador. El generador crea datos nuevos basándose en las entradas condicionales y en valores aleatorios, mientras que el discriminador intenta clasificar los datos como auténticos (del conjunto de entrenamiento) o falsos (generados). Durante el entrenamiento, tanto el generador como el discriminador se actualizan iterativamente para mejorar su rendimiento. El objetivo del generador es crear datos que el discriminador clasifique como auténticos, mientras que el discriminador busca diferenciar correctamente entre datos reales y generados. Este proceso adversarial permite a las cGAN generar datos que sean indistinguibles de los datos reales, pero con la ventaja adicional de que los datos generados pueden cumplir con condiciones específicas establecidas por el usuario.

En el presente trabajo, la cGAN Pix2Pix se usa para generar, a partir de las imágenes dermatoscópicas con la regla del visor, las imágenes correspondientes sin dicha regla. La implementación de dicha red se hizo siguiendo un artículo Image-to-Image Translation with Conditional Adversarial Networks publicado en Tensorflow (cuyo objetivo es una red que genera imágenes de fachadas de edificios a partir de un esbozo simple de sus componentes), dado que el modelo original se hizo en Lua, usando la biblioteca Torch. La implementación sigue el mencionado artículo al pie de la letra (aunque se generalizó la cantidad de filtros que tienen las capas convolucionales, de forma que su ajuste es más sencillo) en cuanto a la construcción de la red (actualizando el código para usar Tensorflow v2.0, dado que el subproyecto Keras fue movido fuera del namespace tensorflow a una biblioteca separada), pero con ciertos cambios en el proceso de entrenamiento, y cambiando totalmente el manejo del dataset (dado que el dataset usado por el artículo se compone de pares de fotos entrada-salida deseada, concatenadas en una sola).

Los cambios introducidos en el modelo Pix2Pix son: el proceso de entre-

namiento se modificó para que fuese interrumpible, es decir, el entrenamiento puede interrumpirse en cualquier momento, para luego continuarse partiendo del trabajo anterior, usando la capacidad de crear "checkpoints" de TensorFlow. El dataset usado (ISIC) contiene varios miles de imágenes dermatológicas, algunas con reglas y otras sin ella; estas últimas fueron extraídas para componer el dataset de entrenamiento como las entradas, a partir de las cuales se generaron las salidas. Para generar las salidas se tomó cada imagen si le añadió una regla sintética donde comúnmente puede encontrarse en las imágenes reales; así, el dataset que compuesto por pares imagen-imagen transformada.

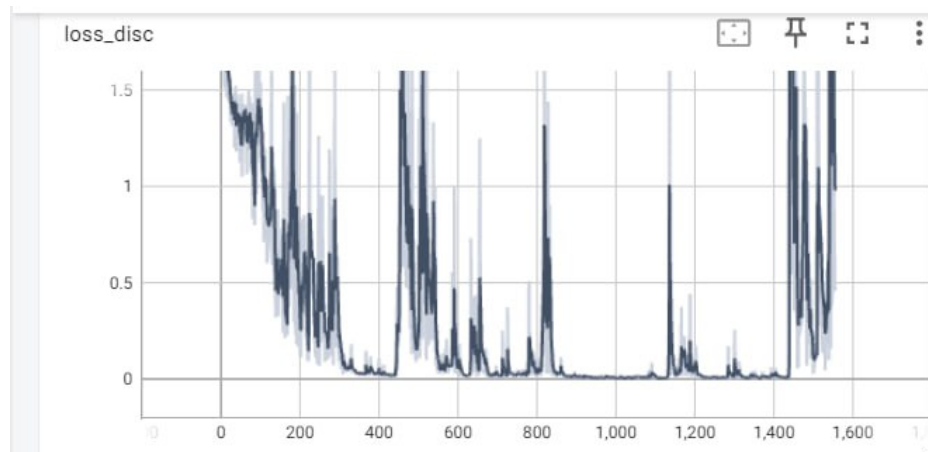


Figure 9: Grafica de perdida del discriminador

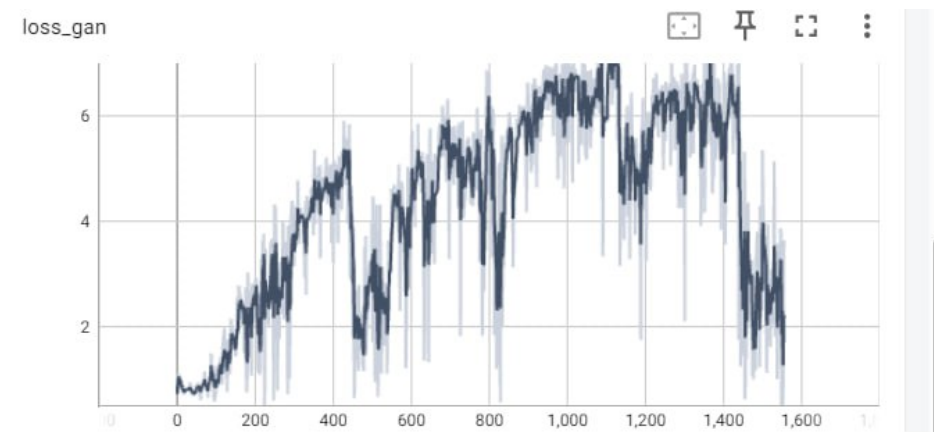


Figure 10: Grafica de perdida del generador

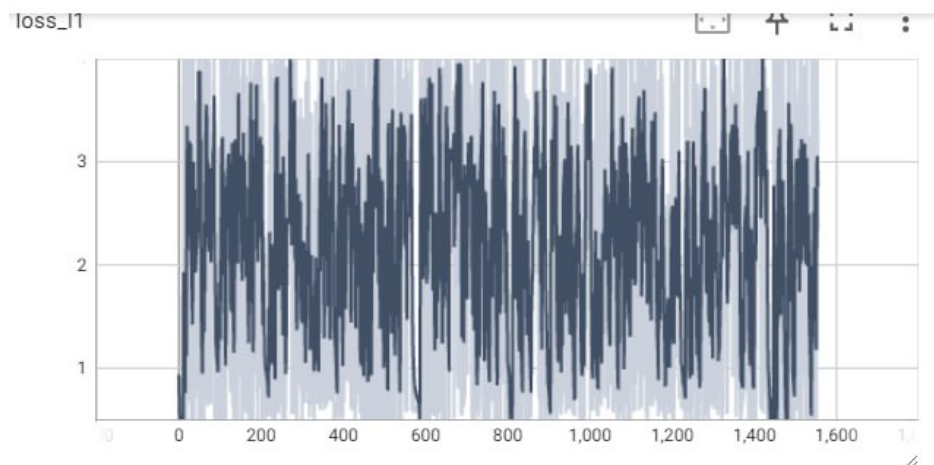


Figure 11: Grafica de perdida L1

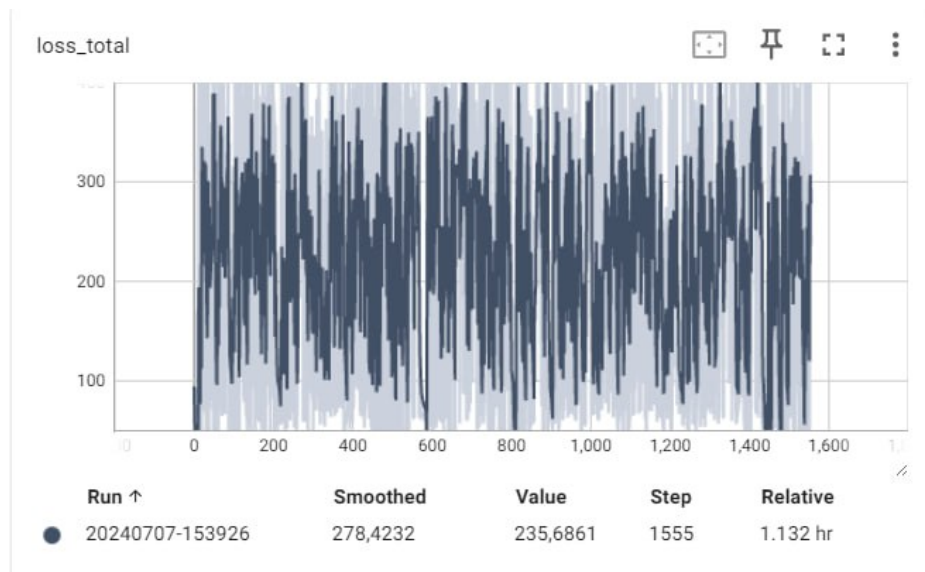


Figure 12: Grafica de perda total

Se realizaron una serie de experimentos en los que se ajustó la cantidad de filtros por capa convolucional, encontrando los mejores resultados en la cantidad sugerida en el artículo en TensorFlow: 64. Los mejores resultados se obtuvieron con 900+ pasos de entrenamiento, como se mostrara a continuación.



Figure 14: 700 pasos de entrenamiento

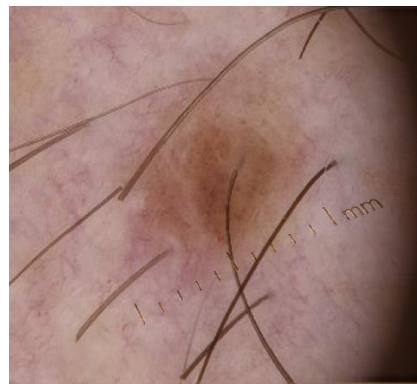


Figure 15: 900 pasos de entrenamiento

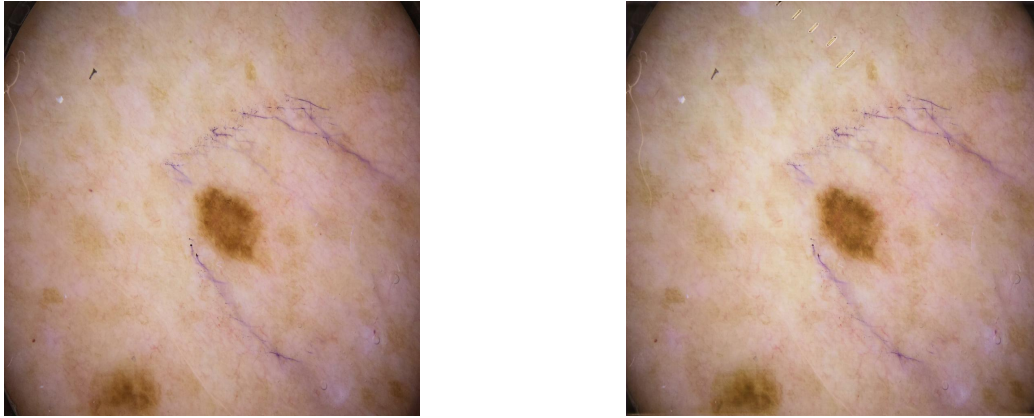


Figure 16: 900+ pasos de entrenamiento

5 Discusión de los resultados

Los resultados obtenidos fueron satisfactorios aunque no perfectos. En muchas de las imágenes, la regla se eliminaba por completo, logrando una efectividad notable. Sin embargo, en otras ocasiones, solo se observaba una sombra sobre el lugar donde se encontraba la regla, indicando que, aunque el modelo estaba en proceso de borrarla, aún quedaban áreas que necesitaban ajustes. A pesar de estas variaciones, el éxito parcial en la eliminación de la regla demuestra la capacidad del modelo para manejar esta tarea compleja, subrayando la importancia de futuros desarrollos y refinamientos para alcanzar una eliminación completa y consistente.

5.1 Validación del modelo

En muchos casos de problemas de image to image translation, se utiliza la validación humana para asegurar la calidad y utilidad de los resultados. En este contexto, se le mostraron imágenes devueltas por el modelo a un conjunto de especialistas médicos, quienes validaron su calidad para ser utilizadas en un diagnóstico de manera positiva. Esta etapa de validación humana es crucial, ya que permite verificar la precisión y relevancia de las imágenes generadas por el modelo, asegurando así su idoneidad para aplicaciones clínicas y diagnósticas.

Además de esta validación por expertos, se realizó un modelo de clasificación basado en una Red Convolutiva Multicapa que dada una imagen

devuelve si la regla se encuentra presente en la imagen o no. Este enfoque permite una evaluación cuantitativa de la presencia de la regla, facilitando la interpretación de los resultados y la comparación con otros métodos. Este modelo, respaldado por el artículo Image Classification Using CNN with Keras and CIFAR-10, tiene un accuracy de 0,85.

La clasificación de imágenes utilizando Redes Neuronales Convolucionales (CNN) implica la extracción de características de la imagen para observar patrones en el conjunto de datos. Las CNN utilizan filtros para la clasificación de imágenes, que existen en muchos tipos según su propósito. Los filtros nos permiten aprovechar la localidad espacial de una imagen imponiendo un patrón de conectividad local entre las neuronas, además de permitir al modelo capturar detalles intrincados y relaciones espaciales dentro de la imagen. La convolución básicamente significa una multiplicación punto a punto de dos funciones para producir una tercera función, aquí una función es nuestra matriz de píxeles de imagen y otra es nuestro filtro. Deslizamos el filtro sobre la imagen obteniendo el producto punto de las dos matrices. La matriz resultante se llama "Mapa de Activación" o "Mapa de Características".

Para realizar una clasificación binaria de imágenes se siguió un proceso que implica la preparación de los datos, la construcción de la red neuronal convolucional, el entrenamiento de la red y finalmente, la evaluación de su rendimiento. Primero, se cargó el conjunto de datos en dos clases y se etiquetaron como 0 (con regla) y 1 (sin regla). Luego, se divide el conjunto de datos en conjuntos de entrenamiento y prueba. La red neuronal convolucional se construye con capas convolucionales, capas de pooling y capas densas para la clasificación final. Se utiliza la función de activación ReLU para las capas convolucionales y softmax para la capa de salida. El modelo se entrena utilizando un optimizador como y una función de pérdida adecuada, como la entropía cruzada binaria, ajustando los pesos de la red para minimizar la diferencia entre las predicciones y las etiquetas reales. Finalmente, se evalúa el rendimiento del modelo en el conjunto de prueba para determinar su precisión, recall, F1-score, etc., proporcionando así una medida cuantitativa de su capacidad para clasificar correctamente las imágenes en las dos clases seleccionadas.

Epoch	loss	accuracy	val_loss	val_acc
Epoch 16/25	0.3297	0.8879	0.4876	0.8462
Epoch 17/25	0.3064	0.8862	0.4267	0.8462
Epoch 18/25	0.3498	0.9129	0.4203	0.8393
Epoch 19/25	0.3535	0.9176	0.4872	0.8462
Epoch 20/25	0.2311	0.9319	0.4337	0.8661
Epoch 21/25	0.2176	0.9386	0.4866	0.8386
Epoch 22/25	0.2413	0.9396	0.4388	0.8661
Epoch 23/25	0.1957	0.9338	0.4578	0.8571
Epoch 24/25	0.2217	0.9388	0.5013	0.8393
Epoch 25/25	0.1912	0.9338	0.4493	0.8661
Test accuracy		0.857542805578864		

Figure 17: Metricas del clasificador biclase

Aunque la pregunta que se busca responder con este trabajo es si la eliminación de la regla mejora las métricas de los modelos de clasificación de imágenes dermatoscópicas, no se pudo obtener acceso a un modelo para realizar los análisis pertinentes, ya que estos suelen estar registrados bajo derechos de autor y, por tanto, no son accesibles al público general.

5.2 Repercusión ética de la solución

El desarrollo de este trabajo presenta implicaciones éticas significativas, las cuales son equivalentes a las de los modelos existentes de clasificación de imágenes dermatoscópicas, ya que se prevee su uso como primer paso de un modelo más amplio de aprendizaje automático que involucre a ambos. Se destaca la importancia de proteger la privacidad y seguridad de los datos de los pacientes, garantizando que la información personal no pueda ser rastreada ni comprometida. Es crucial abordar la equidad en el acceso y uso de este modelo, asegurando que sea accesible y utilizable por profesionales de la salud en todo el mundo, comenzando por las instituciones cubanas. Además, se deben considerar los derechos de autonomía del paciente, comunicándole claramente cómo se utiliza este modelo en su atención médica y cómo puede afectar el diagnóstico y tratamiento recomendado.

5.3 Disclaimer

Este modelo fue entrenado bajo condiciones de recursos computacionales limitados, específicamente en un sistema operativo Linux equipado con 8GB de RAM y un procesador Ryzen AMD 5, sin contar con la adición de una tarjeta gráfica dedicada. En búsqueda de solucionar estas trabas, se utilizó el servicio gratuito de Google Colab mediante varias cuentas, que, aunque

brinde recursos significativamente mejores, cuentan con una restricción temporal de uso por cuenta. Con esto queremos dejar claro que los resultados obtenidos pueden no ser los óptimos para el modelo implementado y que será necesario un entrenamiento en condiciones ideales para confirmar y potencialmente mejorar estos resultados.

6 Conclusiones

En el caso específico de eliminar la regla del visor del dermatoscopio, el enfoque de utilizar una Red Generativa Adversaria Condicional (GAN c) ha demostrado ofrecer mejores resultados comparado con el uso de Redes Neuronales Convolucionales (CNN). Este hallazgo destaca la superioridad de las GAN c en tareas de traducción de imágenes a imágenes, especialmente cuando se trata de generar imágenes realistas que cumplan con criterios específicos definidos por el dominio objetivo. Las GAN c, al ser capaces de aprender representaciones mutuamente beneficiosas entre pares de dominios, permiten una traducción más precisa y detallada, superando así las limitaciones inherentes a las CNN en términos de flexibilidad y precisión en la generación de imágenes. Este resultado subraya la importancia de seleccionar el tipo de modelo adecuado según la naturaleza de la tarea y los requisitos específicos del dominio de aplicación, destacando el papel crucial de las GAN c en el ámbito de la traducción de imágenes a imágenes.

6.1 Trabajo futuro



Figure 18: Pagina web de DermaUH

Este proyecto se prevee que será utilizado por el equipo de Derma UH como primer paso hacia la construcción de un modelo de Machine Learning más robusto y preciso para la clasificación de lesiones cutáneas. Mediante la preparación y optimización de las imágenes dermatoscópicas, eliminando

o estandarizando la regla del visor, se busca mejorar la precisión del clasificador. Este enfoque estratégico está diseñado para validar nuestra hipótesis central: que la adición de un paso de preprocesamiento efectivo puede incrementar significativamente la precisión en la clasificación de lesiones cutáneas por parte de los modelos de Machine Learning. Al abordar la variabilidad introducida por la regla del visor, que puede variar considerablemente entre diferentes modelos de dermatoscopios, se busca reducir el ruido en las imágenes y facilitar la identificación de patrones y características clave por parte de los algoritmos de ML. Con este proyecto, Derma UH busca establecer una base sólida para futuros desarrollos, optimizando el rendimiento de los modelos ML en la detección temprana del cáncer de piel y contribuyendo así a la salud pública cubana y global.

7 Bibliografía

1. Image Classification Using CNN with Keras and CIFAR-10[link al paper](#)
2. Tabla de estado del arte[link al paper](#)