"Segmentación de Lesiones en Retinografías Usando U-Net"

Profesores: Olga Lucía López Acevedo, Hernán David Salinas Jiménez

Estudiantes: Jamir Moreno Salazar, Dany Durango Escobar

Fecha: 07/02/2025

Resumen

Se implementó un modelo U-Net para la segmentación de lesiones en retinografías, enfocándose en exudados duros, exudados blandos y hemorragias. Las imágenes fueron preprocesadas y redimensionadas para optimizar el cómputo. Se entrenó el modelo con binary crossentropy, el optimizador Adam y técnicas como early stopping y checkpointing. Las métricas mostraron valores adecuados de loss y accuracy, pero un Dice score bajo, posiblemente debido al número reducido de épocas y limitaciones de hardware. Se propone mejorar la segmentación aumentando las épocas y optimizando la arquitectura del modelo.

Palabras clave: Segmentación de imágenes, U-Net, retinopatía diabética, aprendizaje profundo, Dice score, procesamiento de imágenes.

Introducción

Este trabajo académico se centra en el reconocimiento de enfermedades oculares, específicamente en la detección de retinopatías, como hemorragias, microaneurismas, exudados duros y suaves, mediante segmentación de imágenes y el desarrollo de una solución basada en la arquitectura U-Net [1]. La retinopatía diabética (RD) [2] es, por ejemplo, la principal causa de pérdida de visión prevenible, afectando principalmente a la población en edad laboral a nivel mundial. El tamizaje para RD, junto con consultas y tratamientos oportunos, es una estrategia globalmente confiable para prevenir la pérdida de visión. Sin embargo, la implementación de programas de detección enfrenta desafíos debido a la escasez de profesionales médicos capacitados para examinar la creciente población diabética en riesgo de desarrollar la enfermedad.

El uso de diagnóstico asistido por computadora para analizar imágenes de la retina ofrece una solución sostenible a este problema a gran escala. Los recientes avances en capacidad computacional y aprendizaje profundo (Deep Learning) brindan una oportunidad para mejorar los métodos de detección y segmentación de lesiones oculares. Para contribuir al desarrollo de esta área, se llevó a cabo el reto "Retinopatía Diabética: Segmentación y Clasificación", organizado en conjunto con el Simposio Internacional IEEE sobre Imágenes Biomédicas (ISBI - 2018).

Este reto se basó en el conjunto de datos Indian Diabetic Retinopathy Image Dataset (IDRiD) [3] e incluyó tres subdesafíos principales: segmentación de lesiones, clasificación de la gravedad de la enfermedad y localización de estructuras retinianas. Estas tareas

permitieron evaluar la capacidad de generalización de los algoritmos propuestos, estableciendo un punto de referencia para futuras investigaciones.

El presente proyecto toma como base la información del reto IDRiD, así como las soluciones propuestas en el mismo y otros trabajos relevantes, con el objetivo de implementar una solución para la segmentación de lesiones en imágenes de retina utilizando la arquitectura U-Net. Es importante destacar que la solución adoptada en este trabajo representa sólo una de las múltiples posibles aproximaciones, además, hay que mencionar que la capacidad para el procesamiento de datos fue un factor limitante durante el desarrollo del proyecto.

Marco Teórico

La segmentación de imágenes médicas ha sido un campo de investigación clave en el desarrollo de modelos de aprendizaje profundo. La segmentación de imágenes de retina presenta retos específicos debido a la variabilidad en la apariencia de las lesiones, el ruido en las imágenes y la presencia de estructuras complejas dentro del ojo.

La arquitectura U-Net, introducida por Ronneberger et al. [4], se ha consolidado como una de las más eficaces para la segmentación biomédica. Su diseño de codificador-decodificador con conexiones de salto permite preservar la información espacial de alta resolución, lo que la hace ideal para segmentar lesiones en imágenes de retina. U-Net ha demostrado su efectividad en distintos contextos clínicos, facilitando la detección automática de estructuras anatómicas y patologías.

Segmentación de imágenes

La segmentación de imágenes es una técnica esencial en el procesamiento digital de imágenes y visión por computadora, cuyo objetivo es dividir una imagen digital en varios segmentos, simplificando y/o transformando su representación para hacerla más significativa y fácil de analizar. Esta técnica permite extraer detalles específicos de una imagen para su posterior análisis. Aunque a menudo se confunde con la detección de objetos y el reconocimiento de imágenes, la segmentación se distingue porque proporciona información más granular sobre los contenidos de la imagen. A diferencia del reconocimiento de imágenes, que asigna etiquetas a toda la imagen, o la detección de objetos, que localiza los objetos dibujando un cuadro delimitador, la segmentación clasifica cada píxel individualmente, dividiendo la imagen en partes llamadas "objetos de imagen" que se pueden analizar por separado.

El proceso de segmentación implica asignar a cada píxel de una imagen un objeto específico, lo que facilita la identificación y separación de elementos dentro de la imagen. Esta técnica permite agrupar píxeles con características comunes, lo que resulta en la creación de segmentos que representan diferentes objetos o regiones dentro de la imagen. Además, la segmentación es una etapa crucial en los sistemas de reconocimiento de imágenes, ya que ayuda a extraer los objetos de interés que luego se pueden utilizar para

tareas como el reconocimiento, la descripción y el entrenamiento de modelos de aprendizaje automático.

Existen diferentes tipos de segmentación, y cada uno tiene su propia utilidad dependiendo del caso de uso. La segmentación semántica, por ejemplo, asigna una etiqueta a cada píxel de la imagen que pertenece a una categoría general, mientras que la segmentación de instancias no solo etiqueta los píxeles, sino que también distingue entre diferentes instancias del mismo objeto. Ambos enfoques son fundamentales para entender el contenido de las imágenes en aplicaciones como la visión por computadora, donde es necesario analizar las fronteras y las relaciones entre los objetos presentes en una imagen [5].

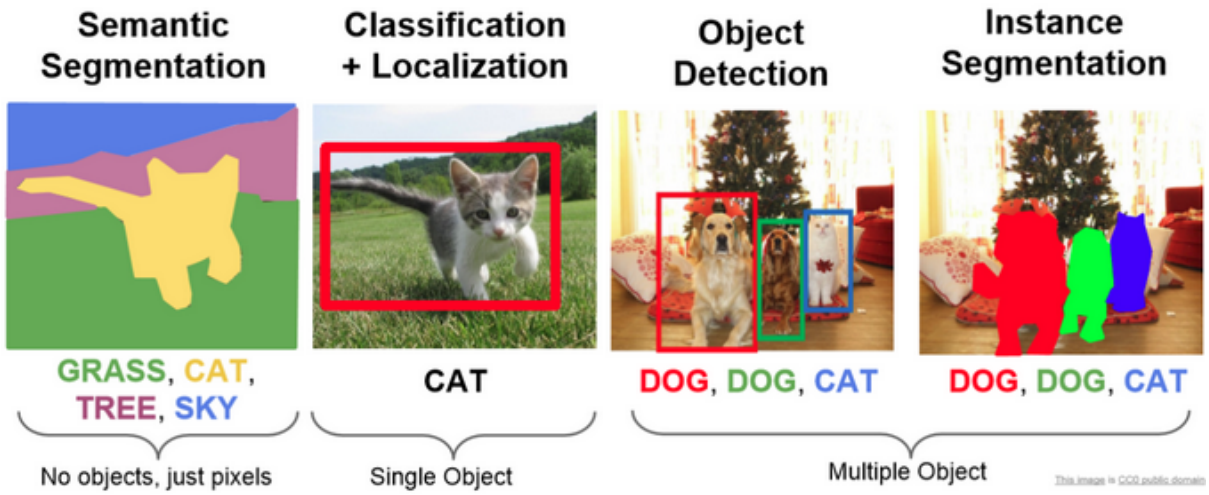

Imagen 1. Tipos de segmentación. Esta imagen describe diferentes tipos de segmentación de imágenes. Nótese como la segmentación semántica y la de instancias logran relacionarse, considerando que esta última no sólo diferencia los píxeles, sino también que logra diferenciar entre las diferentes instancias del mismo objeto.

Arquitectura U-NET

La arquitectura U-Net, propuesta por Olaf Ronneberger y colaboradores en 2015, ha revolucionado la segmentación de imágenes médicas debido a su diseño eficiente y su capacidad para aprender características detalladas a partir de imágenes. Esta red neuronal convolucional (CNN) se basa en un enfoque de entrenamiento de extremo a extremo, lo que significa que puede aprender directamente de los datos de entrada sin la necesidad de intervención manual para extraer características. El diseño de U-Net se organiza en forma de "U", lo que refleja su estructura de codificación y decodificación, en la que se combina reducción y ampliación de la imagen a través de las fases de "downsampling" (submuestreo) y "upsampling" (sobremuestreo) [1].

El proceso de la arquitectura U-Net comienza con una fase de codificación en la que las imágenes se reducen gradualmente en tamaño a medida que se extraen características de nivel superior. En esta etapa, las capas convolucionales y de pooling (submuestreo) permiten que la red aprenda representaciones abstractas de los datos a diferentes niveles.

Esta reducción progresiva de la resolución ayuda a que la red capture las características más complejas y profundas de la imagen, como los bordes, texturas y patrones.

Por otro lado, se encuentra la fase de decodificación, en la que la imagen es aumentada de nuevo a su tamaño original mediante operaciones de upsampling. El upsampling ayuda a que la red reconstruya la imagen de alta resolución mientras mantiene las características aprendidas durante el proceso de codificación. La clave de la arquitectura U-Net es la existencia de conexiones de salto o "skip connections", que permiten que las características extraídas durante la codificación se transfieran a la fase de decodificación. Estas conexiones son fundamentales porque ayudan a preservar detalles espaciales importantes que podrían perderse en la reducción de resolución, lo que mejora la precisión en la segmentación de estructuras finas, como lesiones o áreas anatómicas pequeñas.

Una característica notable de la arquitectura U-Net es que, a pesar de las grandes reducciones de resolución durante el proceso de codificación, el modelo logra mantener una alta capacidad para generar segmentaciones precisas. Esto se debe a que la información contextual de la imagen se propaga eficientemente hacia las capas de mayor resolución en la fase de decodificación. Además, U-Net permite que las imágenes de tamaño arbitrario sean segmentadas, lo que lo hace particularmente útil para aplicaciones médicas, donde las imágenes pueden variar en tamaño y resolución dependiendo del tipo de examen (por ejemplo, resonancia magnética, tomografía computarizada o ultrasonido.

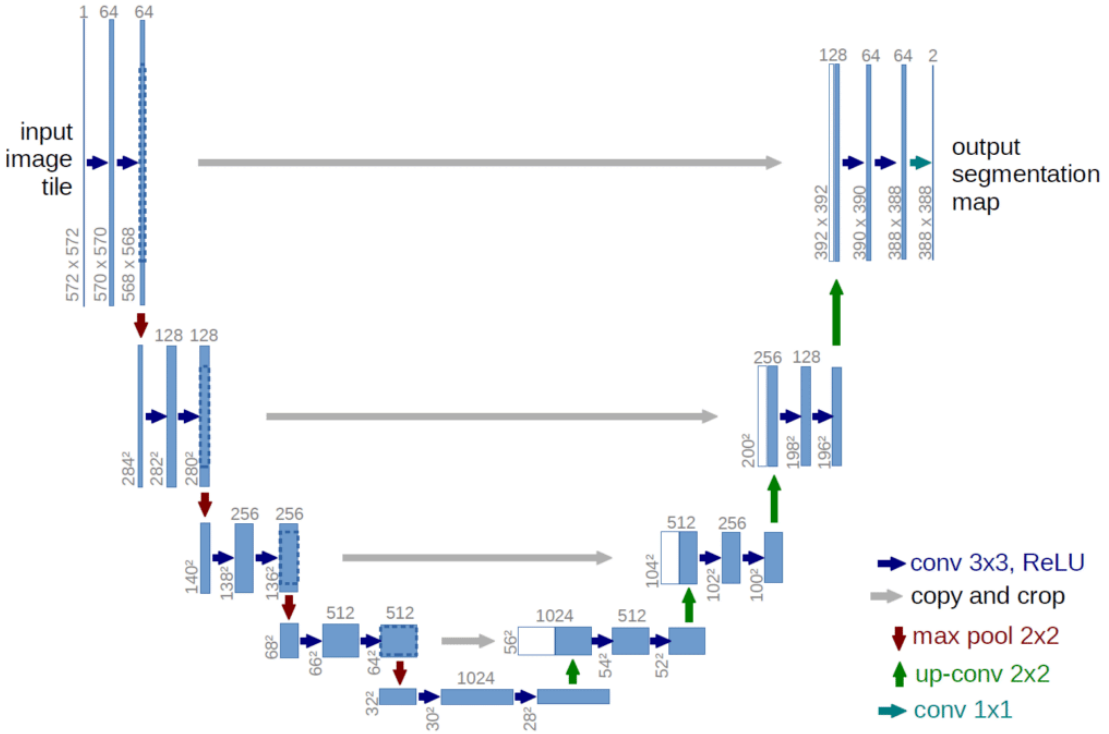

Imagen 2. Arquitectura U-Net. Esta imagen representa la arquitectura implementada en este trabajo, donde se logra apreciar cómo se logra apreciar la segmentación a partir de la codificación y decodificación de las imágenes.

Métricas de Evaluación

Para evaluar el desempeño del modelo U-Net en la segmentación de retinopatía diabética, se emplean métricas estándar en segmentación de imágenes médicas. Una de las más utilizadas es el **Dice Similarity Coefficient (DSC)**, también conocido como Dice-score, el cual mide la superposición entre la segmentación generada por el modelo y la segmentación de referencia (ground truth). Su fórmula es:

$$DSC = \frac{2|A \cap B|}{|A| + |B|}$$

donde A representa la región segmentada por el modelo y B la región real. Esta métrica varía entre 0 y 1, donde 1 indica una segmentación perfecta y 0 significa que no hay coincidencia entre la segmentación del modelo y la referencia.

El **Dice-score** es especialmente útil en contextos donde las regiones a segmentar son pequeñas en comparación con el fondo de la imagen, como en la detección de lesiones en la retina. Su principal ventaja radica en que penaliza tanto la sobre segmentación como la subsegmentación, proporcionando una evaluación equilibrada de la precisión del modelo.

Otras métricas complementarias incluyen la precisión (precisión), la exhaustividad (recall) y la Exactitud Global (Accuracy), que permiten evaluar distintos aspectos del desempeño del modelo en la segmentación de lesiones de la RD.

Accuracy (Precisión Global): La precisión, es una métrica que mide cuántos píxeles fueron clasificados correctamente en relación con el total de píxeles. Para mayor claridad se presenta la siguiente relación:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Donde, TN (True Negatives) = píxeles correctamente clasificados como fondo (no lesión).

Cabe destacar que si la imagen tiene mucha más área sin lesiones que con lesiones, un modelo puede obtener una precisión alta simplemente prediciendo "no lesión" en la mayoría de los píxeles. Métricas como DSC, IoU y F1-score suelen ser representativas en problemas de segmentación [6].

Metodología

1. Preprocesamiento de Datos

El primer paso en la implementación del modelo fue el preprocesamiento de los datos para garantizar su compatibilidad con la arquitectura U-Net y optimizar el tiempo de cómputo. El conjunto de datos utilizado estaba estructurado de la siguiente manera:

```
├── A. Segmentation/
│   ├── 1. Original Images/
│   │   ├── a. Training Set/  # Contiene archivos .jpg
│   │   └── b. Testing Set/   # Contiene archivos .jpg
│   ├── 2. Groundtruths/
│       ├── a. Training Set/
│       │   ├── Hard Exudates/       # Contiene archivos .tif
│       │   ├── Soft Exudates/       # Contiene archivos .tif
│       │   ├── Microaneurysms/      # Contiene archivos .tif
│       │   └── Hemorrhages/         # Contiene archivos .tif
│       └── b. Testing Set/
│           ├── Hard Exudates/       # Contiene archivos .tif
│           ├── Soft Exudates/       # Contiene archivos .tif
│           ├── Microaneurysms/      # Contiene archivos .tif
│           └── Hemorrhages/
```

Imagen 3. Estructura general del almacenamiento de los datos en diferentes carpetas. En esta imagen se ve claramente cómo estaban distribuidos los archivos tanto los .jpg como los .tif en las diferentes carpetas.

Tanto las imágenes originales como sus respectivas máscaras segmentadas disponibles en dichas carpetas, tenían una resolución aproximada de 4288×2848 píxeles; estas imágenes originales venian en formato .jgp, mientras que las máscaras en formato .tif. Sin embargo, este tamaño resultaba computacionalmente costoso y poco eficiente para el entrenamiento de una red neuronal. Para abordar esta problemática, se llevó a cabo un redimensionamiento de las imágenes; aunque cabe aclarar que no todas las lesiones fueron escaladas a la misma resolución. Para Hard exudates, su escalamiento fue de 256×256 píxeles, mientras que para hemorrhages y soft exudates fue de 128×128 píxeles. La elección de estas dimensiones no se basó en un análisis exhaustivo, sino en referencias de estudios previos en los que se trabajó con tamaños similares. Se optó por estas resoluciones también con la finalidad de reducir la carga computacional y mejorar la eficiencia en el entrenamiento sin comprometer la capacidad del modelo para capturar patrones relevantes.

Además de la redimensión, se aplicaron las siguientes técnicas de preprocesamiento, como normalización, y comprobación de consistencia. Tanto para las imágenes originales como las máscaras de segmentación fueron normalizadas a valores entre **0 y 1**, lo que facilita la convergencia del modelo durante el entrenamiento. Por otro lado, se verificó que las imágenes originales y sus máscaras correspondieran correctamente en términos de alineación y formato.

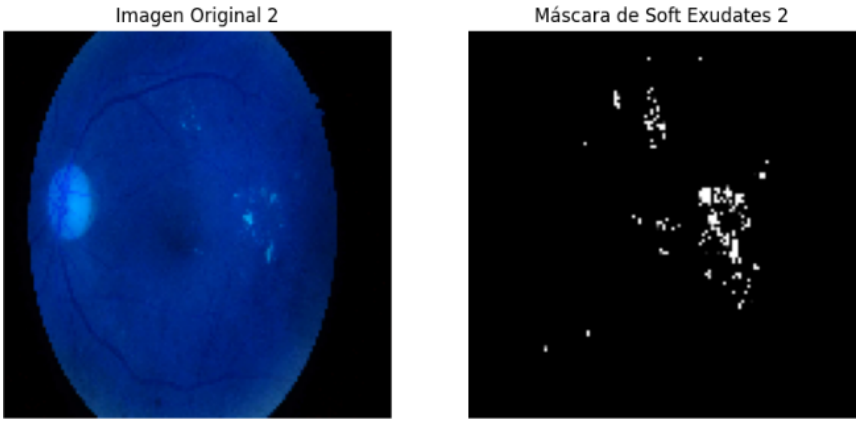

Imagen 4. Comprobación de lectura de alguna de las imágenes. Esta imagen representa una de las imágenes originales para la lesión Soft Exudates con su respectiva máscara.

2. Implementación del Modelo

Para la segmentación de lesiones en las imágenes retinianas, se implementó la arquitectura U-Net, específicamente por su eficacia en tareas de segmentación biomédica. Por esta razón, se construyó un modelo individual para cada tipo de lesión, con la excepción de la lesión “Microaneurysms (MA)”, para el cual no se realizó segmentación en este trabajo debido a algunos inconvenientes en la memoria del computador usado al momento de correr y ejecutar los códigos.

La arquitectura U-Net se implementó con configuraciones como “binary_crossentropy” como función de pérdida, “Adam” como optimizador y 0.001 como la tasa de aprendizaje inicial. Luego, para mejorar la eficiencia del entrenamiento y evitar el sobreajuste, se utilizaron algunos callbacks como “Early stopping” en el cual se estableció un criterio de tres a cinco épocas, dónde se consideraba que después de este número de iteraciones el “validation_loss” no mejoraba, entonces el entrenamiento se detendría y se guardaba el mejor modelo obtenido hasta ese momento. Se implementó además, “Model checkpoint” como un mecanismo de guardado automático para evitar la pérdida de progreso en caso de fallos del sistema o interrupciones. Adicionalmente, se utilizó “TensorBoard” para visualizar métricas clave del entrenamiento, como la evolución de la pérdida y la precisión del modelo.

3. Entrenamiento y Evaluación

El modelo fue entrenado con todo el conjunto de imágenes y máscaras disponibles, dividiendo los datos en training set y testing set, siguiendo las consideraciones tenidas en cuenta en el reto mencionado en la introducción. Una distribución más clara para el manejo de la imágenes se muestra a continuación:

Lesion Type	Set - A Images	Set - B Images
MA	54	27
HE	53	27
SE	26	14
EX	54	27

Tabla 1. Distribución de las imágenes de entrenamiento y testeo. El set A y el set B representan el training test y el testing test respectivamente. Nótese que el conjunto que difiere considerablemente en número es el de sot exudates.

Es importante añadir que durante el entrenamiento, se midieron algunas métricas como el el “Dice Similarity Coefficient (DSC)” y el “Accuracy”, como se mencionan en el marco teórico. Dichas métricas permitieron evaluar el desempeño del modelo y comparar sus resultados con estudios previos en el área de segmentación de lesiones en imágenes retinianas.

Resultados y Discusiones

1. Soft Exudates

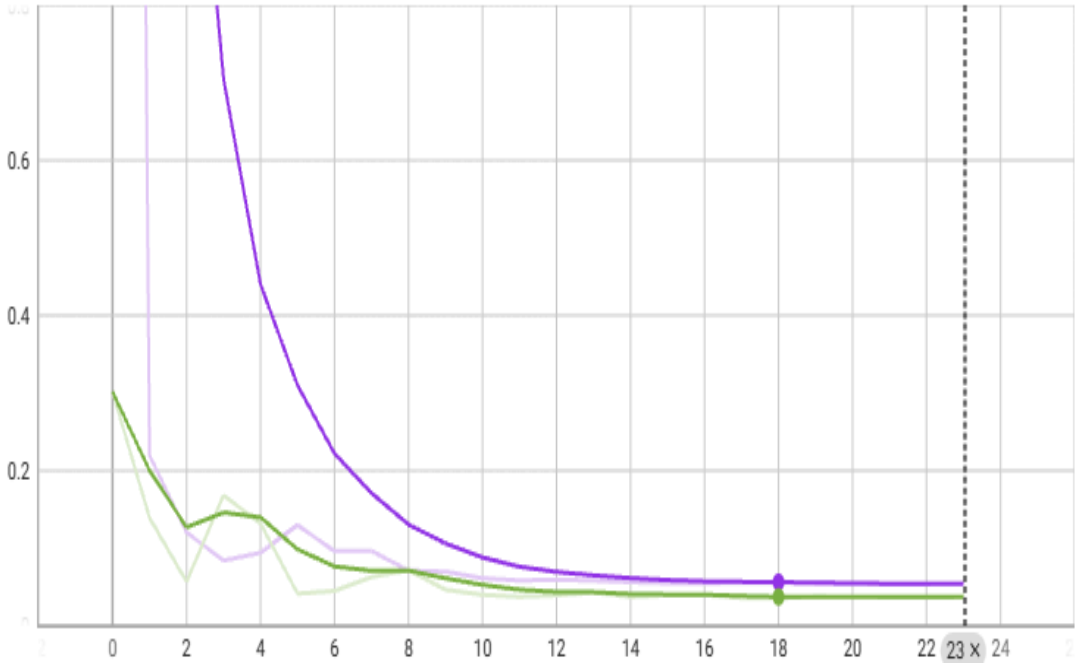

Imagen 5. Curvas de los valores de pérdida versus el número de épocas. En esta imagen, la curva verde representa el validation loss y la morada el training loss.

En la imagen anterior se puede apreciar cómo después de cierta cantidad de épocas la curvas se aplanan y tienden a un mismo valor, parando en 23 épocas debido al early stopping programado. En cuanto al test, este modelo presentó los valores de Loss= 0.0227, Accuracy = 0.9965 y un Dice Score = 0.0039. Aunque los valores de accuracy y loss estan bien el Dice score está muy bajo respecto a lo que se espera; sin embargo concluir algo en este punto no sería lo adecuado, teniendo presente que el numero de épocas es bastante bajo comparado con uno de 200 o 300 épocas como se ha considerado en el reto mencionado en la introducción.

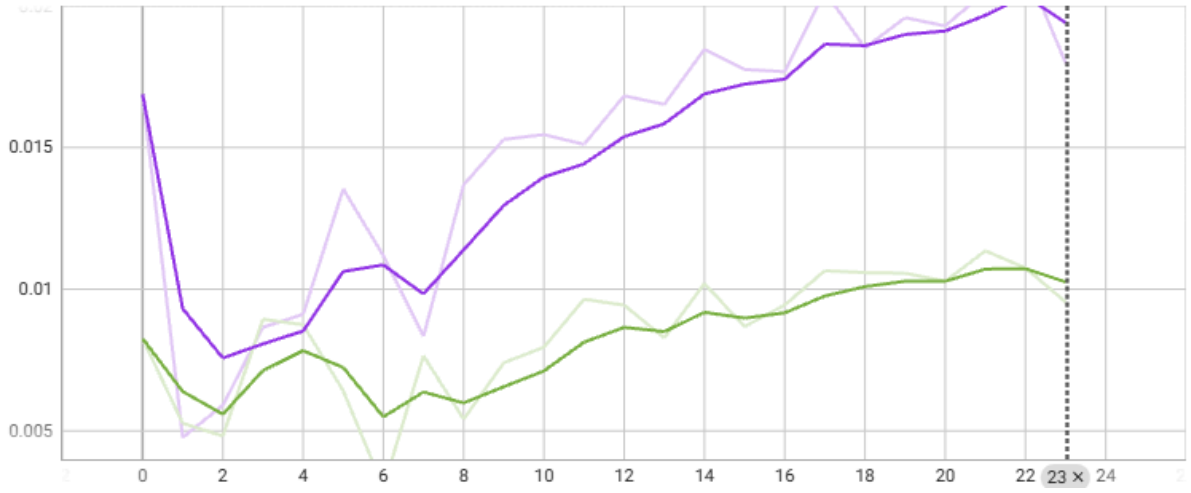

Imagen 6. Representación del Dice score en función de la épocas. Estas curvas representan como cambia el dice score respecto al número de épocas; en donde quizá haya cierta tendencia a mejorar pero con 23 épocas son pocos para notarlo como se menciona justamente antes.

2. Hemorrhages

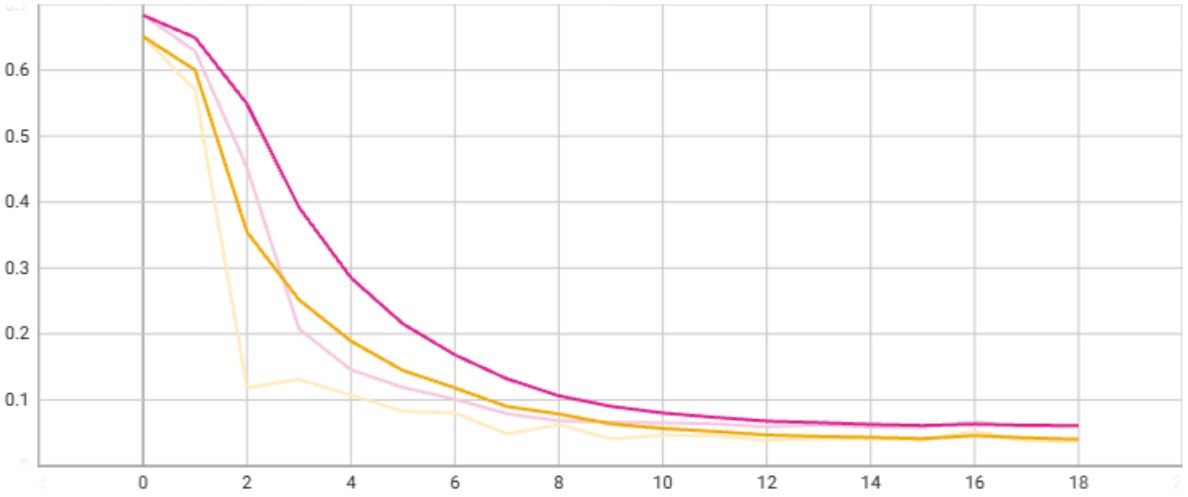

Imagen 7. Curvas de los valores de pérdida versus el número de épocas.. la curva amarilla representa el validation loss, mientras que la naranja representa el training loss. Como se puede ver, las curvas se estabilizan alrededor de 18 épocas.

En cuanto al conjunto de prueba, se obtuvo un valor de pérdida igual a 0.579, para una precisión del 0.9890 y un Dice score de 0.0122. Nótese que el Dice score sigue siendo muy bajo después de 18 épocas, lo que se considera de igual manera que el caso anterior que el número de época no es lo suficientemente alto como para concluir algo acerca del modelo. Una imagen más representativa de lo anteriormente dicho se enseña a continuación:

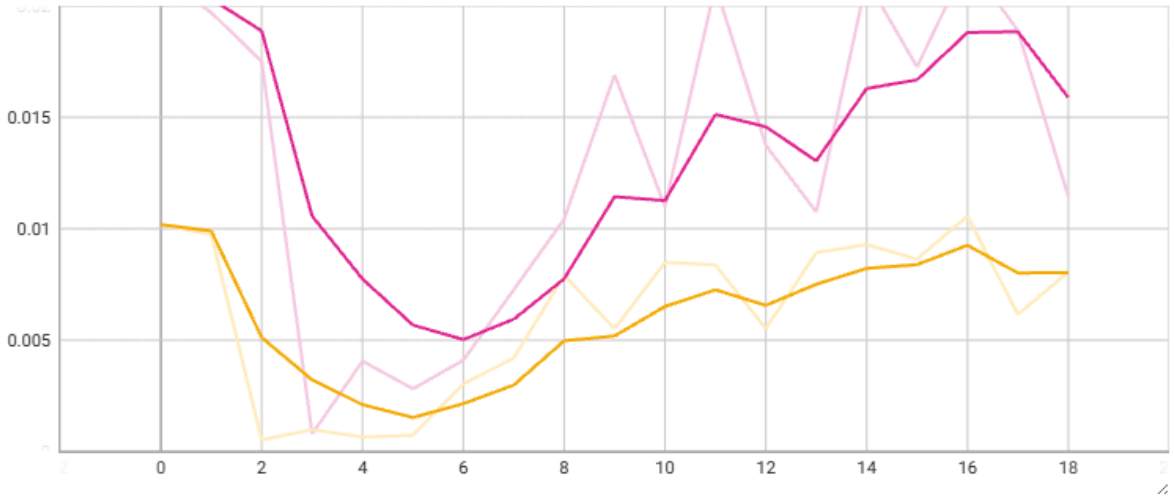

Imagen 8. Representación del Dice score en función de la épocas.

3. Hard Exudates

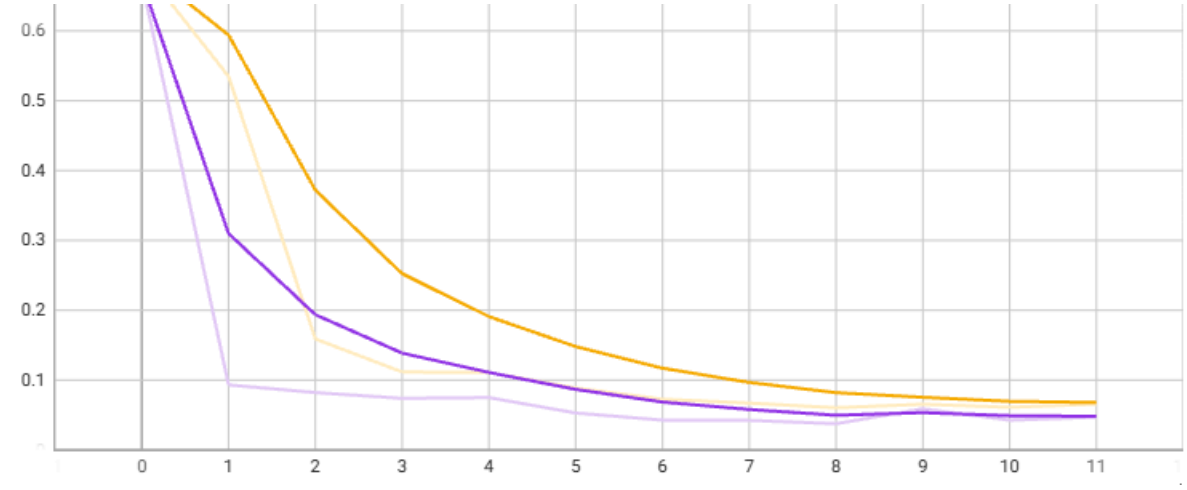

Imagen 9. Curvas de los valores de pérdida versus el número de épocas. Para esta imagen, la curva morada representa el validation loss.

Se resalta que para esta segmentación no se obtuvo el dice score, y los valores del test fueron de 0.0706 para valor de pérdida, con una precisión de 0.9884.

Es importante mencionar que la segmentación de la lesión MA no se realizó debido a limitaciones de hardware. Se decidió interrumpir el proceso para evitar posibles daños en el equipo. Se considera que estas restricciones también influyeron en la capacidad de entrenamiento del modelo, limitando la posibilidad de obtener un Dice Score más representativo. Además, el bajo número de épocas contribuye a la dificultad de extraer conclusiones significativas sobre el rendimiento del modelo en la segmentación de esta categoría de lesiones.

Conclusiones

Se observó que, en los tres casos analizados (Soft Exudates, Hemorrhages y Hard Exudates), el entrenamiento se detuvo en un número de épocas relativamente bajo (18-23), lo que pudo haber afectado negativamente la capacidad del modelo para aprender patrones más complejos. Se sugiere aumentar el número de épocas para evaluar si el Dice Score mejora significativamente.

La falta de recursos computacionales impidió la ejecución de un entrenamiento más prolongado y la segmentación de todas las categorías de lesiones. Un hardware más potente podría permitir un mejor ajuste del modelo y una evaluación más precisa.

Aunque los valores de loss y accuracy fueron aceptables, el Dice Score se mantuvo en niveles bajos en todos los casos. Esto sugiere que el modelo puede estar aprendiendo características generales, pero sin lograr una segmentación precisa de las lesiones.

La implementación de early stopping permitió evitar sobreajuste, pero es posible que se haya detenido el entrenamiento antes de que el modelo alcanzara su máximo potencial. Se recomienda realizar pruebas con diferentes configuraciones para determinar si una mayor cantidad de épocas puede mejorar los resultados sin incurrir en sobreajuste.

Como recomendaciones para mejorar la segmentación de las lesiones se deja a consideración aumentar el número de épocas y optimizar los hiperparámetros del modelo. Utilizar técnicas de aumento de datos para mejorar la generalización. Implementar modelos más robustos o arquitecturas más avanzadas, como U-Net con mayor capacidad de extracción de características y contar además con un hardware más potente para evitar limitaciones en el entrenamiento.

Referencias

[1] Du, G., Cao, X., Liang, J., Chen, X., & Zhan, Y. (2020). Medical image segmentationbased on U-Net: A review. *Journal of Imaging Science and Technology, 64*(2), 020508. https://doi.org/10.2352/J.ImagingSci.Technol.2020.64.2.020508

[2] Khojasteh, P., Aliahmad, B., & Kumar, D. K. (2018). Fundus images analysis using deep features for detection of exudates, hemorrhages and microaneurysms. *BMC Ophthalmology, 18*(1), 288. https://doi.org/10.1186/s12886-018-0954-4

[3] Porwal, P., Pachade, S., Kokare, M., Deshmukh, G., Son, J., Bae, W., Liu, L., Wang, J., Liu, X., Gao, L., Wu, T., Xiao, J., Wang, F., Yin, B., Wang, Y., Danala, G., He, L., Choi, Y. H., Lee, Y. C., ... & Mériaudeau, F. (2019). IDRiD: Diabetic retinopathy – Segmentation and grading challenge. *Medical Image Analysis, 59*, 101561. https://doi.org/10.1016/j.media.2019.101561

[4] Çiçek, O., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In S. Ourselin et al. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016* (Vol. 9901, pp. 424–432). Springer. https://doi.org/10.1007/978-3-319-46723-8_49

[5] Extraído de: https://www.optisolbusiness.com/insight/an-overview-of-image-segmentation-part-1

[6] Huynh, N. (2023, March 1). *Understanding evaluation metrics in medical image segmentation*. Medium. https://medium.com/@nghihuynh_37300/understanding-evaluation-metrics-in-medical-image-segmentation-d289a373a3f