S18 CDT Ciencia, Docencia y Tecnología

Suplemento CDyT
VOL. 15 | N° 18 | 2025 ISSN 2250-4559

PID 6205

Sistema de Información basado en norma DICOM para aplicaciones oftalmológicas orientadas a retinopatías del prematuro (ROP)

Adrián R. Salvatelli; Alejandro J. Hadad; Gustavo H. Bizai; Diego A. Evin

Autoras/es: Laboratorio de Sistemas de Información. Facultad de Ingeniería, Universidad Nacional de Entre Ríos. Ruta provincial 11, km 10 Oro Verde, Entre Ríos, Argentina.
Contacto: adrian.salvatelli@uner.edu.ar

ARK: https://id.caicyt.gov.ar/ark:/s22504559/me3jlbbp1

Resumen

La Retinopatía del Prematuro (ROP) es una enfermedad del desarrollo de los vasos retinianos y el vítreo, con anormal maduración y diferenciación celular. Afecta únicamente a los niños prematuros y especialmente a menores de 1500 g de peso al nacer (PN) y/o menores de 32 semanas de edad gestacional (EG). En Argentina nacen alrededor de 750.000 niños por año, el 10% son prematuros (75.000) y de éstos una tercera parte presentarán factores de riesgo para ROP (25.000), entre los que se incluyen los nacidos PN menor de 1500 g con mayor riesgo (nacimientos: 8.250). En este informe se presentan los avances realizados en este proyecto a través de técnicas de procesamiento digital de imágenes y videos, visión computacional y estándares de Informática médica (telemedicina), dando valor agregado al sistema ODI Smartphone (Oftalmoscopio Digital Indirecto) desarrollado por profesionales vinculados al Grupo ROP Argentina. En este contexto se avanzó es el desarrollo de una aplicación para smartphone orientada a la conformación de una modalidad oftalmológica para un PACS/HIS bajo estándares DICOM/HL7.

Palabras clave: Retinopatía del Prematuro (ROP), Oftalmoscopio Digital Indirecto (ODI), procesamiento digital de imágenes (PDI), Smartphone, Modalidad oftalmológica

Objetivos

A continuación se describen las actividades delineadas del proyecto y detalles de la ejecución efectiva de las mismas. A pesar de los imprevistos surgidos durante su implementación, que requirieron modificaciones en la planificación original, se logró completar exitosamente las etapas fundamentales, manteniendo la alineación con los objetivos iniciales del proyecto.

En primer lugar, se completaron en su totalidad las actividades iniciales (Análisis del estado del arte sobre aplicaciones ROP en smartphone). Se realizó una revisión de antecedentes bibliográficos, recursos de datos y software publicados relacionados al ROP, así como de aspectos de implementación del estándar DICOM empleando smartphones como medio de interoperabilidad. En esta línea se efectuó un análisis y búsqueda de tecnologías que pudieran ser aplicadas en la implementación del prototipo de software smartphone bajo estándares. También se realizó un análisis preliminar de los primeros videos e imágenes tomadas con el ODI (smartphone), para luego desarrollar una estrategia de implementación en el procesamiento digital de video e imágenes capturadas. Estas actividades se llevaron adelante en conjunto con la formación de recursos humanos en sendas tesinas de grado de Bioingeniería y una tesis de Maestría en Informática Médica. En la tesina del alumno Bruno Franseschini titulada, "*Implementación de una aplicación móvil orientada a la exploración del fondo de ojo*", se estudiaron las alternativas tecnológicas para la implementación del prototipo, se analizaron los recursos existentes para aplicaciones ROP y otras de similares características. Dicho estudio complementa la revisión del estado de arte que se había proyectado. Se analizó el estándar DICOM, el tipo y características de base de datos, lenguajes de programación compatibles con el desarrollo de aplicaciones smartphone, conversión a DICOM de videos, y accesos seguros a través de la web, entre otros aspectos.

Con dicha información se implementó la aplicación prototipo: "*Smartphone Fundoscopy*". Esta implementación tiene las siguientes funcionalidades:

- Toma imágenes o vídeos y los convierte a un archivo DICOM
- Etiqueta el estudio como modalidad OP (*ophtalmology*)
- Brinda la posibilidad que el profesional pueda buscar estudios por paciente, fecha de realización, modalidad
- Cuenta con un visualizador integrado de imágenes DICOM
- Flexibiliza la posibilidad de ajuste de la aplicación a las pantallas tanto de un smartphone como a las de PC

Para ello fue necesario definir la arquitectura de Smartphone Fundoscopy, las tecnologías utilizadas, la base de datos, el sistema de comunicación y archivado de imágenes PACS (*Picture Archiving and Communication System*), y el lenguaje de programación para llevar las imágenes al estándar DICOM.

En cuanto al análisis de imágenes y procesamiento digital de imágenes y video, desde el comienzo de la pandemia SARS-COV2 la interacción con los especialistas médicos y el envío de imágenes y video fueron muy discontinuas, dificultando los avances. En consecuencia, no se generaron suficientes imágenes (insumo del PID) por lo cual se consideró necesario redefinir los objetivos y actividades del proyecto en este aspecto. En el contexto de pandemia, se priorizó la formación de recursos humanos principal-

mente de los becarios. Se les dio un soporte virtual permanente para que continúen con el relevamiento de técnicas de procesamiento y algunos ensayos de prueba de algoritmos publicados en la literatura de referencia.

La becaria PID Maribel Gruber se enfocó en la primera etapa del procesamiento. Una vez analizados los videos disponibles procesó manualmente las regiones de interés con el apoyo de un script ad-hoc escrito en Matlab. Además participó en el estudio de técnicas de deconvolución basadas en modelos matemáticos de la PSF (*Point Spread Function*) para quitar las aberraciones ópticas de la lupa oftálmica presentes en las imágenes capturadas.

El becario CIN, Hernán Rodriguez Ruiz Diaz, completó su tesina de grado en Bioingeniería: *"Aplicación de técnicas de procesado de imágenes y vídeos adquiridos desde un smartphone a fondos de ojo de bebés prematuros con retinopatía del prematuro (ROP)"*. En el desarrollo avanzó en su formación en el procesamiento digital de imágenes y videos, así como en el análisis de imágenes mediante redes neuronales convolucionales (CNN). En esa actividad fue asistido por integrantes docentes del PID y también trabajó en el preprocesamiento de las imágenes y videos crudos provistos por los especialistas.

En cuanto a la tesis de Maestría en Informática Médica, se encuadra en la formación de posgrado del Dr. Guillermo Monteoliva. Esta tesis titulada: "*Evaluación de la Oftalmoscopia Digital Indirecta Smartphone en el diagnóstico de la Retinopatía del Prematuro mediante Telemedicina*", se encuentra en la última fase de su desarrollo. La misma propone evaluar la utilidad de las imágenes retinales obtenidas mediante un dispositivo smartphone empleando la técnica de registro "manos libres" (ODI) para el diagnóstico de Retinopatía del Prematuro. Este estudio está orientado a comparar los resultados diagnósticos de los datos de evaluación clínica convencional, respecto a resultados de la evaluación de las imágenes por Telemedicina, obtenidas de bases de datos separadas para cada grupo. Este procedimiento involucra la validación manual de los resultados por parte de oftalmólogos expertos externos al proyecto (equipos colaboradores de Colombia y México).

Como se mencionó, debido a las restricciones asociadas a la pandemia por Covid-19, el proceso de construcción de una base de datos de imágenes ROP obtenidas mediante ODI y la cadencia de adquisición de casos clínicos se vio sensiblemente afectado. Es por ello que se debió redefinir en las metas planteadas, cuya ejecución hacía imprescindible contar con los datos en cantidad y calidad suficiente. En ese sentido se pudo avanzar empleando técnicas de preprocesamiento que no requieren datos supervisados, a través de la definición de una secuencia de procesamiento (pipeline) para mejorar la calidad de los registros, acondicionándolos para su documentación y análisis automático posterior. Las tareas y resultados de este ítem se enmarcan principalmente en el proyecto final de carrera del becario Hernán Rodríguez, dirigido por Gustavo Bizai.

En actividades posteriores, se realizó la evaluación del proceso para la transferencia de datos al PACS a través del sistema implementado. Para esas evaluaciones se emplearon fotografías y videos de distintas resoluciones y duración. Luego se realizó idéntica prueba con los estudios realizados con el sistema ODI.

En dichas pruebas se verificó la performance del sistema en lo que refiere a calidad

de imágenes, el tamaño del archivo DICOM, su correcta visualización y los tiempos de almacenamiento.

En relación a la actividad *Aplicación de técnica computacionales inteligentes a videos e imágenes ROP, para la conformación de un procesador de Imágenes DICOM* del plan de trabajo, la imposibilidad de contar con una cantidad suficiente de datos para entrenar modelos de clasificación llevó a redefinir los objetivos originales. Se estudiaron y evaluaron modelos y algoritmos del estado del arte para la segmentación de disco óptico, fóvea y vasos sanguíneos. Sin embargo, la mayoría de los sistemas publicados fueron entrenados empleando bases de datos de adultos, generalmente de buena calidad y condiciones controladas de adquisición, por lo que su aplicación sobre los datos de este proyecto no resulta directa. Como parte de esta actividad, se completó la tarea *Técnicas orientadas a la identificación de características geométricas y fotométricas de las imágenes que caracterizan a la ROP y sus grados*. En este sentido, se analizaron antecedentes y propuestas metodológicas para el procesamiento de videos e imágenes obtenidos mediante smartphones, particularmente para aplicaciones médicas. Sobre esos antecedentes se definió y elaboró un sistema de referencia, estableciendo una secuencia de pasos o pipeline de referencia, basado en técnicas de procesamiento digital de imágenes y visión computacional específicos para las características del dominio de ROP. También se completó la tarea *Evaluación de la aplicación de estas técnicas en las imágenes, particulares y tomadas de otra manera, para detectar estructuras de interés*, en el marco del pipeline de procesamiento definido se desarrollaron métodos para la detección y segmentación de la lupa y detección de frames útiles a partir de su contenido de información y medidas de calidad de imagen. Debido a la falta de datos propios anotados manualmente, no se logró entrenar modelos para la detección de disco óptico, zona foveal y vasos sanguíneos con la cadena de adquisición de imágenes propias. Sin embargo, sí se experimentó con diferentes modelos para estas tareas, pero entrenadas sobre bases de datos de retinografías de referencia. Se evaluó la aplicación de estos modelos sobre datos propios, confirmándose la necesidad de reentrenar los modelos a las características específicas de nuestros datos. Finalmente se avanzó en la identificación de las características de las imágenes recopiladas, se determinaron las fuentes de variabilidad y se estudiaron los efectos sobre las tareas de procesamiento y análisis correspondientes. Se experimentaron empíricamente las consecuencias de condiciones adversas sobre la detección de estructuras anatómicas de la retina. Dado que no se llegó a conformar un conjunto de datos etiquetado manualmente, ni modelos entrenados sobre estas condiciones, no se llegó a compilar un conjunto sistemático de casos que presenten dificultades para el diagnóstico médico ni el de los sistemas automáticos, como se había previsto.

Finalmente, se completaron la mayoría de las tareas correspondientes a la actividad *Conformación física del PACS*. Se pudo montar un servidor PACS/HL7 con acceso web con el objeto de enviar y recibir los estudios desde ubicaciones remotas. En el proceso se pudieron identificar potenciales inconvenientes y cuellos de botella del sistema, así como cuestiones de escalabilidad. En relación a este mismo punto, se estudió el comportamiento del sistema desarrollado, analizando tanto el desempeño del servidor ORTHANC como la aplicación web desarrollada.

Marco teórico y metodológico

El presente proyecto se desarrolló en torno al procesamiento, almacenamiento, análisis y gestión de datos obtenidos en pruebas ROP. La materia prima de todo el proceso son videofilmaciones de exámenes del fondo ocular de bebés prematuros, realizadas con teléfonos inteligentes. En estos estudios el teléfono funciona como un oftalmoscopio indirecto. El médico oftalmólogo manipula con una mano la cabeza y ojo del bebé, mientras que con la otra mano sostiene una lupa que permite observar el fondo del ojo, una vez que logra hacer foco. Por lo tanto, lo que captura la cámara es la imagen indirecta, que se forma en la lupa cuando se enfoca el fondo ocular. Este procedimiento de enfoque y manipulación del bebé no es sencillo y se ha reportado que apenas un cuarto del total de estudios de oftalmología indirecta por video ha podido utilizarse para evaluaciones de ROP [1].

Por otro lado, y en referencia a esta práctica, muchos profesionales afirman que al momento de realizar un estudio, pueden observar elementos o características que luego en el soporte de registro (imagen o video) se pierden. En ese sentido, se analizó y propuso una secuencia de procesamiento para mejorar la calidad de los registros, acondicionándolos para su documentación y análisis automático posterior, de forma tal que se minimice el nivel de pérdida referido por los profesionales. Las funcionalidades mencionadas requieren aplicar sobre imágenes y videos crudos una serie de operaciones que se pueden organizar en una secuencia o pipeline de procesamiento. Dado que los médicos oftalmopediatras han manifestado la necesidad de contar tanto con la documentación de imágenes como video de los estudios que realizan se definieron dos pipelines diferentes, y se presentan en las Fig. 1 y Fig. 2.

Figura 1. Pipeline para procesamiento de imágenes

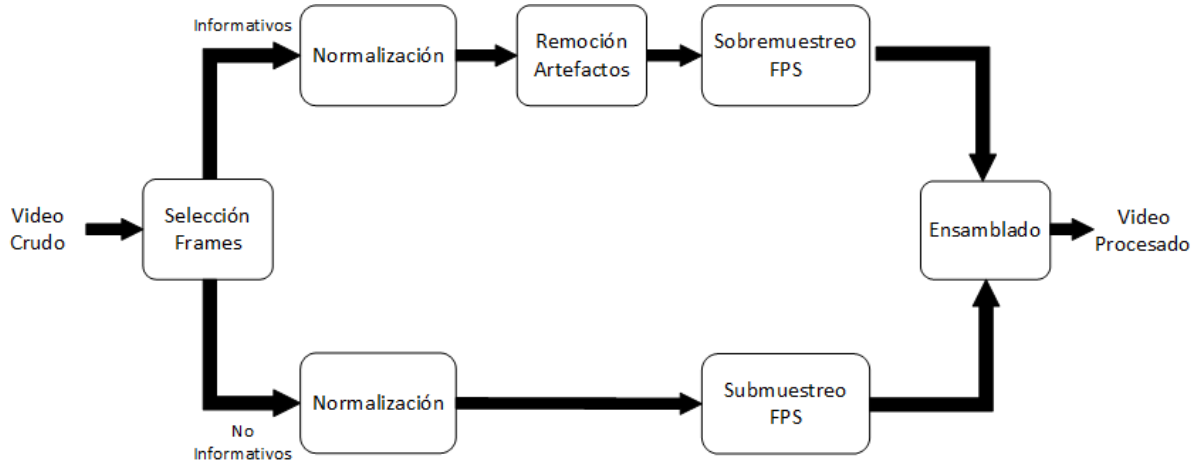

Figura 2. Pipeline para procesamiento de videos

Para la documentación de imágenes, como se muestra en la Fig. 1 el objetivo es seleccionar las tomas que contengan imágenes claras del fondo ocular en diferentes posiciones (campos) y con ellos, generar un mosaico del fondo en toda su extensión de acuerdo a las capturas realizadas. Esa selección admite dos alternativas, se puede dar a partir de imágenes ya preseleccionadas por el profesional o a partir del video crudo. En el último caso, dado que las secuencias de video, en general, poseen pocos cuadros útiles, es decir conteniendo una imagen adecuada del fondo ocular, la primera tarea a realizar es descartar los cuadros sin información útil, aquellos donde no hay lupa, donde la lupa aparece fuera de foco, o donde aparezcan aberraciones en la imagen de la misma como reflejos u oclusiones. Para determinar los frames donde aparece la lupa obviamente se debe identificar en primer lugar la lupa. Ésta puede segmentarse de manera sencilla mediante procedimientos de detección de curvas y/o por propiedades colorimétricas de los pixeles, o por métodos más actuales como los basados en aprendizaje profundo. Asimismo, una vez efectuada esa segmentación, es necesario determinar si la imagen está enfocada o no.

El módulo de normalización busca llevar las características de las imágenes (y video) de entrada a un rango uniforme de acuerdo a sus atributos fotométricos, geométricos así como de resoluciones. Se busca compensar las diferencias de iluminación en las salas de registro, así como las diferencias en la calidad de las cámaras, y en las características de las tomas. Los dispositivos de captura de imágenes pueden tener propiedades muy heterogéneas: diferentes resoluciones espaciales (asociadas entre otras cosas, con la cantidad de píxeles de los sensores), resolución de contraste (asociada con profundidad en bits por canal), ajustes de brillo y contraste automáticos, etc. El objetivo de esta etapa es proveer a la entrada del flujo de procesamiento, imágenes comparables desde los puntos de vista enunciados previamente. Por ejemplo, definir algo tan sencillo como la relación de aspecto de las imágenes a procesar, el número de pixeles por lado (si es necesario un resampleo con los correspondientes algoritmos de interpolación), si se trabajará con los tres canales RGB o se seleccionará uno de ellos o una combinación de dos canales ponderados, el rango de bits por canal, el nivel de brillo y contraste y características cromáticas. Otro punto no menos importante es conocer el modelo de ruido que afecta a las imágenes y el consecuente filtrado (suavizado Gaussiano, filtrado de mediana y tantos otros disponibles) para mejorar la relación señal ruido.

La tercera etapa, del pipeline de procesamiento de imágenes es la de remoción de artefactos. Tiene el propósito de corregir en la medida de lo posible, rasgos de las imágenes y videos que no tienen que ver con características del fondo ocular sino con el proceso de adquisición. Como el proceso de captura de las imágenes es dinámico en el sentido que por un lado se tiene el movimiento de la cámara (y la lupa) respecto al paciente y a esto se agrega el movimiento del propio paciente, creemos conveniente incluir un paso de compensación. Dado que parte de la información que no resulta útil desde el punto de vista médico puede resultar útil para la estabilización de la imagen, o para la registración, se propone aprovechar además la versión de las imágenes previas al enmascaramiento de la lupa. A su vez en el módulo de estabilización de imagen se considera interesante plantear una operación en dos fases como proponen Klemm et al. [2] para el caso de imágenes de flujo sanguíneo cerebral: una primera etapa en la que se busca eliminar el movimiento rígido de gran escala, causado por el movimiento de la cámara con respecto al paciente, y una segunda etapa de estabilización fina de movimientos causados por el movimiento del ojo y el movimiento residual del primer paso.

El siguiente paso intenta encontrar y segmentar los vasos sanguíneos en la imagen que proyecta la lupa y hacer un seguimiento en frames sucesivos.

El quinto está formulado con el propósito de componer a partir de múltiples proyecciones del fondo ocular una panorámica de mayor resolución con los diferentes campos oculares a fin de permitir una documentación resumida y enriquecida, así como la entrada al módulo de clasificación posterior.

Para el caso del pipeline de video, mostrado en la Fig. 2, el objetivo es almacenar a manera de documentación del estudio todo el video pero maximizando el nivel de información. De esta manera se propone por un lado mejorar la calidad del video crudo y por el otro acelerar la porción del video sin información relevante y hacerlo más lento en las porciones que presenten mayores detalles sobre el fondo ocular. Esto último genera la necesidad de clasificar los frames en informativos y no informativos. A los primeros, someterlos a un proceso de realce que les permita a los profesionales visualizar una secuencia mejorada de la información útil (constituida por los cuadros que contienen la lupa con la imagen del fondo ocular enfocado) y darles la posibilidad de sobremuestrear los cuadros por segundo para observar dicha información en modo "cámara lenta". Por otro lado, los cuadros no informativos no requieren ningún procesamiento adicional al de normalización, y a los efectos de dar continuidad al video, se propone submuestrear los cuadros por segundo no informativos de manera que esta parte no útil del estudio pueda observarse en "cámara rápida". En este pipeline las operaciones del módulo de normalización son compartidas con el pipeline para imágenes.

Un detalle no menor a tener en cuenta en la selección de la propuesta está relacionado con la complejidad y requerimiento de memoria/procesamiento de cada uno de los módulos, ya que ello determina si es posible llevar adelante parte del procesamiento del lado del cliente, o por el contrario si todas estas operaciones deberán efectuarse en el servidor.

Las fotografías capturadas por los teléfonos inteligentes para que puedan interoperar en sistemas HIS/RIS deben cumplir con estándares. El estándar reconocido mundialmente para el intercambio de imágenes médicas es DICOM. Fue publicado por primera vez en 1993 por la National Electrical Manufacturers Association (NEMA) y fue desarrollado conjuntamente con el American College of Radiology (ACR). El estándar DICOM original fue denominado el estándar ACR-NEMA en referencia a estas organizaciones. Cada año se publican versiones actualizadas, actualmente consta de 21 partes diferentes, que tratan del protocolo y formatos DICOM, así como la especificación de conformidad DICOM está basada en el entorno de la Programación Orientada a Objetos (OOP), se definen objetos que incluyen imágenes, y siempre se refiere a esos objetos en el contexto de las operaciones aplicables a ellos, tales como almacenarlos, moverlos o buscarlos, crearlos, imprimirlos, etc. Los objetos en DICOM son denominados Objetos de Información (IOD), y las operaciones o servicios son llamados Clases de Servicio (Service Class). Además, se define el Par de Objetos de Servicio (SOP) que consiste en una combinación de un Servicio y un Objeto, formando el Par.

A continuación se describen los criterios establecidos y elementos para la generación de la imagen en formato DICOM. Teniendo definido el lenguaje de programación, se utilizó la librería PyDicom[1] la cual facilitó el proceso de desarrollo. A continuación, se nombran y describen los TAGS utilizados para crear el DataSet de las imágenes DI-

1. https://pydicom.github.io/

COM, los cuales se agruparon en categorías: Sintaxis de Transferencia, Imagen, Paciente, Estudio y Equipo.

Por otro lado, la sintaxis de transferencia define la manera de cómo están escritos los *Data Element*. La imagen se envía en su formato nativo y se eligió la sintaxis de transferencia en formato Big Endian con VR explícito.

Imagen

Columnas (Columns) y Filas (Rows): se obtienen de la imagen cargada.

Interpretación fotométrica (PhotometricInterpretation): especifica la interpretación prevista de los píxeles. Se definen diferentes términos: MONOCHROME, PALETTE COLOR, RGB, YBR, etc. Se estableció por defecto RGB.

Muestras por pixel (SamplesPerPixel): es el número de planos separados en la imagen. Se definen uno y tres planos de imagen. Para la interpretación fotométrica MONOCHROME y PALETTE COLOR el número de planos es 1. Para RGB y otro modelo de 3 vectores es 3.

Paciente

Nombre de Paciente (PatientName), ingresados por el usuario.

Fecha de nacimiento (PatientBirthDate) ingresado por el usuario.

Género (PatientSex): género del paciente por el usuario. Los valores posibles: O, M o F.

Id de Paciente (PatientID): es el DNI del paciente, ingresado por el usuario.

Estudio

Modalidad (Modality): Tipo de equipo que originalmente adquirió los datos utilizados para crear la imagen. Por defecto: OP (Oftalmoscopia) definido por el estándar DICOM.

Lateralidad del objeto fotografiado examinado (ImageLaterality): Ingresado por el usuario puede tomar los valores L, R o B.

Fecha de Estudio (StudyDate): fecha de creación de la imagen.

Hora de Estudio (StudyTime): hora de creación de la imagen.

Equipo

Fabricante (Make): del equipo que generó la imagen.

Modelo (Model): del equipo que generó la imagen.

Pipeline de procesamiento

El desarrollo de la evaluación del pipeline de procesamiento fue del tipo cualitativa-subjetiva. El grupo de expertos fue conformado por un grupo de 4 oftalmólogos, los cuales fueron contestando una serie de preguntas preestablecidas en formularios online. Estas preguntas estaban destinadas a calificar la practicidad de las herramientas brindadas por el algoritmo, es decir, el video procesado y el mosaico retinal. Además, los oftalmólogos contaban con una serie de rúbricas de forma de que puedan contestar las preguntas con un criterio definido.

A continuación se presentan las preguntas presentadas en los formularios y rúbricas:

Velocidad del video (informativo)		
¿Resulta útil la detención del video en puntos importantes?		
Es Impráctico (1)	Resulta indiferente (2)	Si, es útil (3)

Tabla 3.1: *Pregunta referida a los cuadros seleccionados como informativos por el video, en los cuáles se detiene momentáneamente.*

Velocidad del video (no informativo)		
¿Pasar en cámara rápida los cuadros no relevantes resulta práctico?		
Es Impráctico (1)	Resulta indiferente (2)	Si, es útil (3)

Tabla 3.2: Pregunta referida a los cuadros seleccionados como No informativos por el video, en los cuáles el video se reproduce más rápido.

Resaltado de vasos		
¿Qué tanto mejora la visualización de los vasos el resaltado de las imágenes?		
Empeora la visualización	Es indiferente	Mejora la visualización

Tabla 3.2: *Pregunta referida a calificar la visualización de los vasos resaltados.*

Características del video
No se detiene en algunos cuadros que son importantes: Esto se refiere a que pasa de largo algunos videos que se pueden identificar importantes para el diagnóstico y que tienen información que no se vio en ningún otro cuadro.
Muestra mucha información redundante: se detiene en cuadros que ya se fueron repitiendo a lo largo del video.
Hay muchos cuadros en donde se detiene el video y muestra información irrelevante: el video se detiene por error de procesamiento en cuadros que no muestran nada importante.

Tabla 3.3: Pregunta enfocada a preguntar las características de los cuadros mostrados en los videos.

Utilidad del Video Procesado		
Califique la practicidad del video procesado con respecto al video original:		
Poco útil (1)	Útil (2)	Excelente Utilidad (3)

Tabla 3.4: pregunta referida a la utilidad que les brinda el video

Diagnóstico de ROP	
¿Le sirve el video procesado para realizar diagnóstico de ROP?	
Si, sirve para diagnóstico(1)	No, no sirve para diagnóstico (2)

Tabla 3.5: pregunta referida a la posibilidad de hacer diagnóstico con el video.

3.4.1.2. Preguntas de Mosaico

Visualización de los vasos	
Calificación	Descripción
Excelente (3)	Los vasos pequeños son claramente visibles y definidos. La neovascularización y/o avascularización puede visualizarse de manera correcta, si existiese.
Buena (2)	Los vasos pequeños son claramente visibles pero no son nítidos
Inadecuada (1)	No hay estructuras anatómicas claramente visibles

Tabla 3.6: pregunta referida a calificar los vasos visualizados.

Nivel de Practicidad del Mosaico en el diagnóstico		
Califique la practicidad del Mosaico presentado:		
Poca utilidad (1)	Aceptable (2)	Excelente Utilidad (3)

Tabla 3.7: pregunta referida a calificar la practicidad del mosaico.

Artefactos en la imagen	
Calificación	Descripción
Excelente (3)	No son apreciables ningún artefacto.
Buena (2)	Los artefactos se distinguen pero no afectan al diagnóstico.
Inadecuada (1)	Los artefactos influyen de manera moderada o significativa en el diagnóstico.

Tabla 3.8: pregunta referida a calificar los artefactos encontrados en la imagen del mosaico.

Un desafío fue investigar la coherencia de puntuaciones establecidas entre oftalmólogos. La estadística Kappa propuesta por Cohen se utiliza para medir la concordancia entre dos evaluadores. Dado que en este caso fueron cuatro los evaluadores, se recurrió a Kappa de Fleiss, una generalización de la kappa de Cohen para más de dos evaluadores que determina el Nivel de concordancia entre calificadores según un

índice $\varkappa$ que toma valores entre 0 y 1. La máxima concordancia posible corresponde a $\varkappa = 1$. El valor $\varkappa = 0$ se obtiene cuando la concordancia observada es precisamente la que se espera a causa exclusivamente del azar. A la hora de interpretar el valor de $\varkappa$, se dispone de una escala como la siguiente:

Valor de $\varkappa$	**Nivel de concordancia**
< 0,20	Pobre
0,21 - 0,40	Débil
0,41 - 0,60	Moderada
0,61 - 0,80	Buena
0,81 - 1,00	Muy Buena

Más allá de los inconvenientes antes mencionados respecto de la conformación de la base de datos, como parte de las actividades de finalización del PID se realizó un relevamiento de datasets públicos relacionados y de los nuevos enfoques en cuanto a tareas de segmentación de imágenes que pueden ser de utilidad en una continuidad del presente proyecto.

Producto de la revisión bibliográfica se identificaron y obtuvieron algunos conjuntos de datos de imágenes de retina [2-13] que se emplearon en distintas fases del proyecto. La Tabla 1 resume alguna de las características de esos conjuntos de datos.

Dataset		Año	Resolución	N° Pacientes	Anotaciones	Condiciones
STARE		2000	605 x 700	20	Segmentación Vasos, Disco Óptico	10 SA, 10 PAT
DRIVE		2004	768 x 584	40	Segmentación Vasos	33 SA, 7 RD
ARIA		2006	576 x 768	161	Segmentación Vasos, Disco Óptico, Fóvea	61 SA, 59 RD,23 DME
REVIEW	HRIS	2008	3584 x 2438	4	Segmentación Vasos	16 RD
	VDIS		1360 x 1024	8		
	CLRIS		2160 x 1440	2		
	KPIS		288 x 119, 170 x 92	2		
CHASE_DB1		2011	990 x 960	28	Segmentación Vasos	28 SA
INSPIRE-AVR		2011	2392 x 2348	40	Segmentación Vasos, Disco Óptico	40 GLA

REFUGE	2018	2124 x 2056 1634 x 1634	1200	Disco Óptico, Fóvea	1080 SA, 120 GA
UoA-DR	2018	2124 x 2056	200	Segmentación Vasos, Disco Óptico, Fóvea	56 SA, 114 RD
IDRiD	2018	4288 X 2848	516	Disco Óptico, Fóvea	SA,DR
FIVES	2021	2048 x 2048	800	Segmentación Vasos	200 SA, 200 RD, 200 DME, 200 GLA
RVD	2022	1800 x 1800 (video)	1270	Segmentación Vasos Temporal	SA,PAT

Tabla 1. Resumen de conjuntos de datos públicos de imágenes de retina. SA: Sanos; PAT: Diferentes patologías; RD: Retinopatía Diabética; DME: Degeneración Macular Vinculada con la Edad; GLA: Glaucoma

Si bien, como se mencionó anteriormente, surgieron inconvenientes en el desarrollo del proyecto para la conformación de un conjunto de datos propios con anotaciones manuales por parte de los médicos oftalmólogos especializados en ROP, se pudo diseñar y desarrollar parcialmente la infraestructura de datos para el almacenamiento y gestión del material que se pudo adquirir en el lapso de ejecución del proyecto. Para ese desarrollo se utilizó un sistema de gestión de base de datos NoSQL. Dentro de los modelos de datos disponibles en este grupo de bases de datos se utilizó el modelo documental implementado a través de MongoDB. Desde el punto de vista del proyecto este modelo de datos jerárquico permite no solo gestionar las imágenes y videos sino también los metadatos que la describen. Asimismo también permite tratar con variaciones en la forma de presentación de las mismas dado que el modelo jerárquico utilizado se implementa en una estructura flexible que permite que los registros de paciente que se reciban puedan tener diferencias en los datos y sus tipos.

Para la implementación de los métodos de visión artificial capaces de identificar estructuras anatómicas en las imágenes de retina, en primer lugar se realizó una revisión bibliográfica, que se fue actualizando conforme se desarrollaba el proyecto.

La segmentación es una tarea fundamental en el análisis de imágenes médicas, e implica identificar y delinear regiones de interés (ROI) dentro de una imagen capturada. Tales áreas de interés pueden ser por ejemplo órganos, lesiones o tejidos específicos. Una segmentación precisa es esencial para muchas aplicaciones clínicas, ya que suele constituir el insumo que emplean posteriormente los algoritmos de diagnóstico automático. En otros casos el tamaño y la forma de la estructura anatómica será un signo que pueda emplear el médico para la planificación del tratamiento y el seguimiento de la progresión de la enfermedad [14]. La segmentación manual ha sido durante mucho tiempo el patrón oro para delinear estructuras anatómicas y regiones patológicas, pero este proceso lleva mucho tiempo, requiere mucho trabajo y a menudo exige un alto grado de experiencia. Los métodos de segmentación semiautomáticos o totalmente automáticos pueden reducir considerablemente el tiempo y el trabajo necesarios, aumentar la coherencia y permitir el análisis de conjuntos de datos a gran escala.

Con respecto a los métodos más recientes para la segmentación de estructuras anatómicas en imágenes de retina, la mayoría están basados en algoritmos de aprendizaje profundo (*deep learning*). Los modelos basados en aprendizaje profundo demostraron ser muy útiles en la segmentación de imágenes médicas debido a su capacidad para aprender características sutiles o intrincadas de la imagen y ofrecer resultados

de segmentación precisos en una amplia gama de tareas, desde la segmentación de estructuras anatómicas específicas hasta la identificación de regiones patológicas [15]. Entre las aproximaciones basadas en aprendizaje profundo se puede apreciar una clara tendencia hacia la adopción de redes neuronales completamente convolucionales [16-21]. Estas redes permiten combinar la información semántica contenida en las capas profundas con información de apariencia, capturada por las capas superficiales, y demostraron mejorar la precisión de las segmentaciones para imágenes de retina [22]. La arquitectura preferida dentro de este grupo es la de U-Nets, sin embargo muchos trabajos se centran en modificaciones de esta estructura de base [23,24]. También se pueden encontrar otros modelos aplicados a la segmentación de vasos retinianos, como las redes generativas adversativas [25], redes convolucionales en grafos [26] y el aprendizaje de ensambles [27].

Finalmente, así como en otras tareas de visión artificial, recientemente se han propuesto varios algoritmos basados en el modelo de Transformers [28], como ViT [29], Swin [30] y Mask2Former [31]. También versiones híbridas entre U-Nets y Transformers [32], con resultados prometedores.

Una limitación importante de los modelos actuales de segmentación de imágenes médicas es su naturaleza específica. Estos modelos suelen diseñarse y entrenarse para estructuras anatómicas particulares, y su rendimiento puede degradarse significativamente cuando se aplican a nuevas tareas o a diferentes tipos de datos de imágenes. Esta limitación supone un obstáculo importante para la aplicación generalizada de estos modelos en la práctica clínica. En cambio, los avances recientes en el campo de la segmentación de imágenes naturales, que se han sucedido durante la ejecución del presente proyecto, trajeron consigo el surgimiento de los denominados *modelos fundacionales* de segmentación [33, 34], que muestran una notable versatilidad y rendimiento en diversas tareas y condiciones de segmentación. Sin embargo, su aplicación a la segmentación de imágenes médicas ha supuesto un reto debido a la importante brecha de dominio [35]. Por lo tanto, existe una demanda creciente de modelos universales en la segmentación de imágenes médicas: modelos que puedan entrenarse, y una vez entrenados, puedan aplicarse a una amplia gama de tareas de segmentación. Tales modelos no sólo mostrarían una mayor versatilidad en términos de capacidad del modelo, sino que también podrían dar lugar a resultados más coherentes en las distintas tareas, al beneficiarse de una misma arquitectura subyacente y un proceso de construcción unificado. En este sentido a continuación se describen brevemente algunas herramientas recientes con las cuales se han realizado algunos ensayos orientados a incorporar este nuevo enfoque en futuros proyectos relacionados con el área de ROP.

El objetivo de esta etapa exploratoria fue comparar el desempeño de modelos estándar, que formaran parte del estado del arte en la tarea de segmentación de imágenes, con los modelos de segmentación fundacionales. Para experimentar con los modelos de segmentación estándar se eligió la plataforma OpenMMLab [34]. Entre las razones que motivaron esta elección se pueden mencionar: su versatilidad y eficiencia de ejecución, la cantidad de modelos del estado del arte preentrenados y estrategias de entrenamiento como de evaluación que pone a disposición y las funcionalidades para manipulación de datos.

Esas cualidades surgen de los objetivos planteados por ese proyecto:

- Proporcionar bibliotecas de alta calidad para reducir las dificultades en la reimplementación de algoritmos.
- Crear pipelines de despliegue eficientes dirigidas a una variedad de backends y dispositivos.
- Crear una base sólida para la investigación y el desarrollo en visión por computadora.
- Tender puentes entre la investigación académica y las aplicaciones industriales con cadenas de herramientas completas.

Esta suite, que se lanzó en octubre de 2018, está basada en PyTorch, y proporciona un motor universal de entrenamiento y evaluación, así como operadores de redes neuronales y transformaciones de datos, que sirve como base para el despliegue de sistemas de visión artificial. Está compuesta por más de 30 bibliotecas, 300 algoritmos y 2000 modelos preentrenados. El módulo específico de esta suite que se utilizó en el proyecto es OpenMMSegmentation [37].

Por otro lado, como representante de los modelos fundacionales, se eligió MedSAM [35], el primer modelo de tipo fundacional para la segmentación universal de imágenes médicas. Como se dijo, a diferencia de los modelos previos de propósitos específicos entrenados con un tipo de caso particular y cuyo desempeño cae considerablemente para imágenes fuera de su contexto, los modelos fundacionales se caracterizan por exhibir un buen desempeño para una variedad de casos diferentes. Las características de las imágenes del campo médico suponen un dominio desafiante para los modelos fundacionales debido al gap sustancial respecto a las imágenes usadas para entrenar los modelos de propósitos generales.

MedSAM es una adaptación del modelo SAM [31] entrenados con más de un millón de pares imagen médica y sus correspondientes máscaras. En su desarrollo los autores reportan experimentos exhaustivos, en los que se emplearon más de 70 tareas de validación interna y 40 tareas de validación externa, y abarcan una variedad de estructuras anatómicas, condiciones patológicas y modalidades de imágenes médicas.

Los resultados experimentales demuestran que MedSAM supera sistemáticamente al modelo fundacional de segmentación del estado del arte (SOTA), a la vez que alcanza un rendimiento a la par o incluso superior al de los modelos especializados. Estos resultados sugieren un potencial prometedor de MedSAM como herramienta para la segmentación de imágenes médicas. Su arquitectura utiliza una red basada en transformers [38], que ha demostrado una notable eficacia en diversos ámbitos, como el procesamiento del lenguaje natural y el reconocimiento de imágenes [29]. En concreto, la red incorpora un codificador de imágenes basado en *vision transformer*s (ViT) encargado de extraer las características de la imagen, un codificador de prompt para integrar las interacciones del usuario (bounding boxes), y un decodificador de máscaras que genera resultados de segmentación y puntuaciones de verosimilitud a partir de la codificación de la imagen, los bounding boxes y el token de salida.

El modelo ViT que utiliza consta de 12 capas de transformers, cada una de las cuales incluye un bloque de autoatención multicabezal y un bloque de perceptrón multicapa (MLP) que incorpora la normalización de capas. Su preentrenamiento se realizó mediante un modelo de autoencodificador enmascarado, seguido de un entrenamiento totalmente supervisado en el conjunto de datos SAM.

Las imágenes de entrada (1024×1024×3) se transforman en una secuencia de parches 2D aplanados con un tamaño de 16×16×3. Esto produce un tamaño de características en la codificación de la imagen de 64×64. Es decir, que tras pasar por dicho codificador la escala es reducida en un factor de 16. Los codificadores de imagen mapean el punto de esquina de la imagen del cuadro delimitador en codificaciones vectoriales de 256 dimensiones, cada cuadro delimitador se representa mediante un par de incrustaciones del punto de esquina superior izquierdo y el punto de esquina inferior derecho. Para facilitar las interacciones del usuario en tiempo real una vez calculada la codificación de la imagen el algoritmo usa una arquitectura de decodificación de máscaras, conteniendo 2 capas de transformers [9] para fusionar la incrustación de la imagen y la codificación puntual, y dos capas convolucionales transpuestas para mejorar la resolución de la incrustación hasta 256×256. Finalmente, se utiliza una activación sigmoidea, seguida de interpolaciones bilineales para ajustarse al tamaño de entrada.

Durante el preprocesamiento de datos se usaron 1.090.486 pares imagen médica-máscara para el desarrollo del modelo (no se incluyen los conjuntos de validación externa). Para la validación el conjunto de datos se dividió aleatoriamente en 80%, 10% y 10% como entrenamiento, ajuste y validación, respectivamente. Esta configuración les permitió controlar el rendimiento del modelo en el conjunto de ajuste y ajustar sus parámetros durante el entrenamiento para evitar el sobreajuste. Para la validación externa se emplearon conjuntos de datos no vistos por el modelo durante el entrenamiento. Estos conjuntos de datos proporcionan una prueba rigurosa de la capacidad de generalización del modelo, ya que representan nuevos pacientes, condiciones de imagen y, potencialmente, nuevas tareas de segmentación que el modelo no ha encontrado antes. Al evaluar el rendimiento de MedSAM en estos conjuntos de datos desconocidos, es posible esbozar cuál sería el desempeño esperado en entornos clínicos del mundo real, donde se registra una amplia gama de variabilidad en los datos.

Síntesis de Resultados y Conclusiones

Uno de los primeros resultados alcanzados fue el prototipo de la aplicación *Smartphone Fundoscopy.* El software fue concebido bajo un patrón de Arquitectura Cliente-Servidor combinado con capas, siguiendo la estructura MVC (*Model View Controller*). El patrón de arquitectura MVC se eligió para poder independizar la lógica de negocio de la interfaz y del acceso a la información. La separación de estas capas contribuye a la reutilización, versatilidad y mantenibilidad del proyecto. De esta manera, las vistas son susceptibles a cambios sin la necesidad de provocar que todo el sistema varíe, manteniendo la información y la lógica de negocio intacta. Una de las ventajas, que fue importante a la hora de la elección de este patrón, es la facilidad para la detección y corrección de errores, debido a que puede focalizarse en qué capa se encuentra una determinada falla.

El patrón de arquitectura Cliente-Servidor se escogió debido a que se ahorra tiempo y recursos económicos al realizarse actualizaciones, de este modo todos los clientes tienen la última versión del software publicado. Además, permite el acceso en cualquier parte del mundo donde se posea conexión a internet.

Además se definieron las tecnologías para la realización de la aplicación web, en base al patrón de arquitectura que se estableció en cada capa. Así se decidió separar a las tecnologías según la interacción con el usuario, en el *Frontend* que se ejecuta del

lado del cliente y el *Backend* del lado del servidor. Dentro este último, se hizo una división entre el lugar donde se implementa la lógica de negocio y donde se encuentran almacenados los datos. Al momento de seleccionar las tecnologías se tuvo en cuenta que las mismas sean de código abierto (*Open Source)*, gratuitas y que contribuyan al desarrollo de la aplicación de forma ágil.

En cuanto a la base de datos se decidió trabajar con una *no relacional* debido a que en la aplicación no hay transacciones, además le da al proyecto escalabilidad y flexibilidad. Al analizar las diferentes alternativas en base a lineamientos generales de las características deseadas, sobre todo que fueran de código abierto, se decidió utilizar *Mongo*, que adopta un modelo de datos en forma de colecciones y documentos.

Para la elección del PACS se tuvo en cuenta que fuera de código abierto, gratuito y de fácil configuración. Uno de los PACS más populares es DCM4CHEE, la desventaja del mismo es que se necesita tener una base de conocimientos y gran cantidad de horas para su configuración. Una alternativa que surgió para tratar de simplificar esos requisitos es *Orthanc*, que hace posible una instalación más sencilla y rápida, además tiene una curva de aprendizaje más accesible, y no requiere ningún requisito previo.

Para la generación de la imagen en formato DICOM, se consideró adecuada la biblioteca PyDicom, por lo que se optó por emplear el lenguaje Python, que además facilitó el proceso de desarrollo.

Toda la aplicación fue montada temporalmente en un servidor pago con recursos propios del laboratorio, probada y analizada por los expertos oftalmólogos, que dieron una muy buena evaluación al prototipo, de fácil uso y con grandes ventajas en la forma de trabajo. Otra ventaja es que no afecta a la modalidad de trabajo llevada a cabo por el médico oftalmólogo. En las Figuras 1 y 2 se muestran algunas de las interfaces desarrolladas para el prototipo.

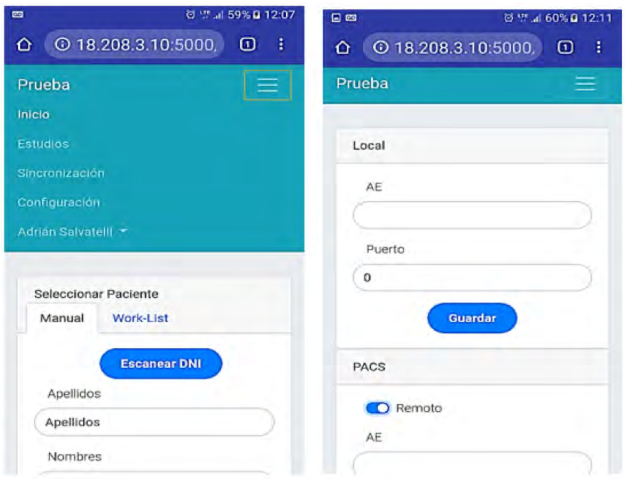

Figura 1. Interfaz del prototipo software, menú de inicio (izq.), estudios y configuración (der.)

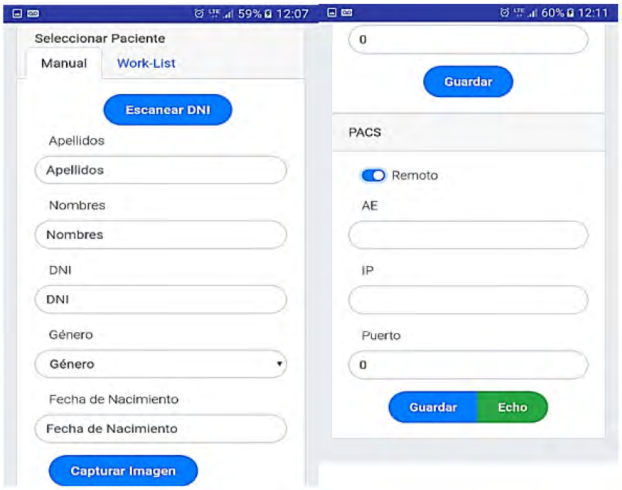

Figura 2. Interfaz del prototipo software, menú de carga de paciente (izq.) y configuración del PACS (der.)

En cuanto a la actividad 3, en el período informado se ha comenzado con el punto 3.1. En tal sentido se han separado en fotogramas los vídeos ODI ROP por intermedio de los expertos en ROP, Dres. Saidman y Monteoliva. Dichos profesionales obtuvieron 50 archivos que incluyen el video en crudo del procedimiento, los fotogramas que a su criterio muestran la patología y una planilla donde se establece el diagnóstico ROP según la clasificación estándar. De estos 50 archivos se han descartado 4 por no tener la información completa.

La mayoría de los videos tienen una resolución digital de 3840 x 2160 (8 Mpíxels aprox.) con una duración de entre 7 y 32 segundos en formato MOV y MP4. Un 10% de estos estudios, fueron realizados con resolución digital de 1920 x 1080 (2 Mpíxels). Ambos estudios fueron obtenidos con un equipo iPhone 8. Iphone y Ipad son dos dispositivos móviles aceptados para uso médico de imágenes por FDA, quien establece una resolución digital mínima, de la imagen completa capturada para uso oftalmológico de 4 Mpixels.

La cantidad de fotogramas seleccionados por los expertos varía entre 3 y 4 mayormente representativos por video capturado. El formato del fotograma es RGB con 8 bits por canal de color. Dentro de estos fotogramas, sólo una parte son píxeles de interés para un futuro diagnóstico. Estos “píxeles útiles” se encuentran dentro de la imagen que proyecta la lupa de 28 o 66 Dioptrías. Para un análisis cuantitativo de esta cantidad de píxeles útiles se ha desarrollado un script de Matlab que segmenta dicha lupa de cada fotograma y enmascara con color negro la información irrelevante. Luego, obteniendo el centroide de la imagen del disco de la lupa, fue recortado el fotograma en su región rectangular más próxima.

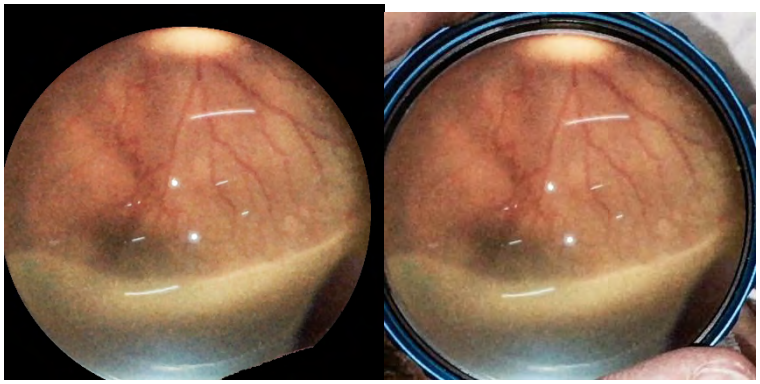

Figura 3. Imagen con ROP original (izq.) y procesada (der.)

Analizando cada fotograma procesado y recortado se ha determinado que el 58% de los fotogramas tienen una resolución digital de 2 Mpixels, mientras que el resto una resolución muy inferior. De este 58% la cantidad de píxeles útiles es inferior a 1,2 Mpixels, es decir un promedio aproximado del total de píxeles del 55%.

En este mismo análisis se han incluido parámetros obtenidos de los histogramas de los píxeles útiles, resultando sus medias entre los colores 60 y 140 con desvío estándar promedio de 77 valores. Morfológicamente los histogramas presentan más de 3 modas y los rendimientos de color de cada fotograma es muy disperso.

En una etapa posterior se trabajó sobre distintos modelos gaussianos de PSF para ser utilizados como filtros de deconvolución. De esta manera se han quitado algunas aberraciones ópticas mejorando la calidad visual de la imagen. Se sigue trabajando en este sentido analizando la calidad de la imagen, pero desde el punto de vista computacional.

Dentro de esta actividad 3, el becario Rodríguez Ruiz Días Hernán propuso trabajar en una secuencia o pipeline de procesamiento, que consta de etapas de enmascarado de imágenes, registro, normalización, remoción de artefactos, submuestreo, realce de vasos y ensamblado, para la conformación de un mosaico de las imágenes de la retina. De estas tareas, comenzó con la etapa de normalización: todas las imágenes y videos se llevan a un formato estándar y se asegura de que los niveles de intensidad de las imágenes se encuentren dentro de los marcos definidos.

Luego, se centró en la clasificación de los cuadros informativos (que contienen imagen del fondo ocular) para desechar los no informativos (sin información diagnóstica, lo que sucede durante la manipulación de los párpados del bebé y búsqueda del cuadro de imagen). Comenzó haciendo un análisis en el espacio de colores HSV, referido a tonalidad (Hue), saturación (Saturation) e intensidad (Value). La clasificación en el espacio HSV es más robusta a los cambios de iluminación, sombras y las variaciones de textura que en otros espacios de color. Se especifica un contorno cerrado S en el espacio HSV, donde los pixeles con valores dentro del contorno, se clasifican como retinianos, y se obtiene un puntaje. La puntuación HSV de un fotograma viene dada por la proporción de píxeles de la retina en la imagen con respecto a la cantidad de píxeles totales de la imagen.

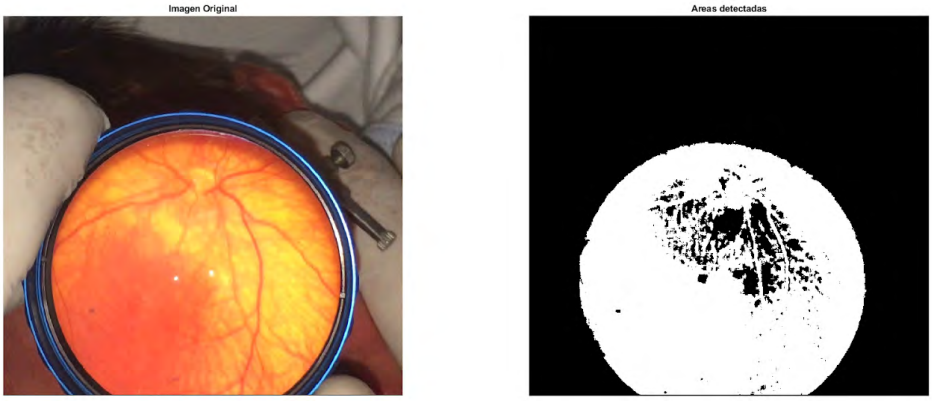

Figura 4: Imagen original a la izquierda. Imagen con máscara HSV

Una vez clasificados los cuadros con imágenes del fondo, se hizo el análisis del enfoque con análisis frecuencial. Se observa que la información anatómica relevante en los fotogramas está presente en frecuencias espaciales intermedias, mientras que las frecuencias altas se encuentran presentes en imágenes con artefactos de movimiento y de entrelazado debido al video. Por esto, se procede al cálculo de la Transformada Rápida de Fourier y se aplican dos filtros Gaussianos para segmentar así el espectro de frecuencias bajas, medias y altas. Una vez hecho esto, se procede a elaborar un puntaje a partir de la proporción entre las frecuencias medias y las frecuencias altas, que determina el grado de enfoque/desenfoque.

Una vez hecha la clasificación entre imágenes informativas y no informativas, se procedió a enmascarar la imagen en la zona de la retina, es decir, donde se encuentra la lente del dispositivo. Para ello, se preparó una primera máscara binaria donde los pixeles de color gris toman el valor 1 y los valores restantes toman el valor 0. Para mejorar el contraste, se puede realizar una deconvolución Lucy-Richardson, donde se modela la función de dispersión de puntos como una función gaussiana. Por último en el preprocesamiento, se utiliza un filtro de acentuación de bordes Sobel.

Una vez hecho esto, se realizó la detección de la lupa mediante la transformada de Hough. Este enfoque se utiliza debido a su robustez en presencia de ruido, oclusión e iluminación variable.

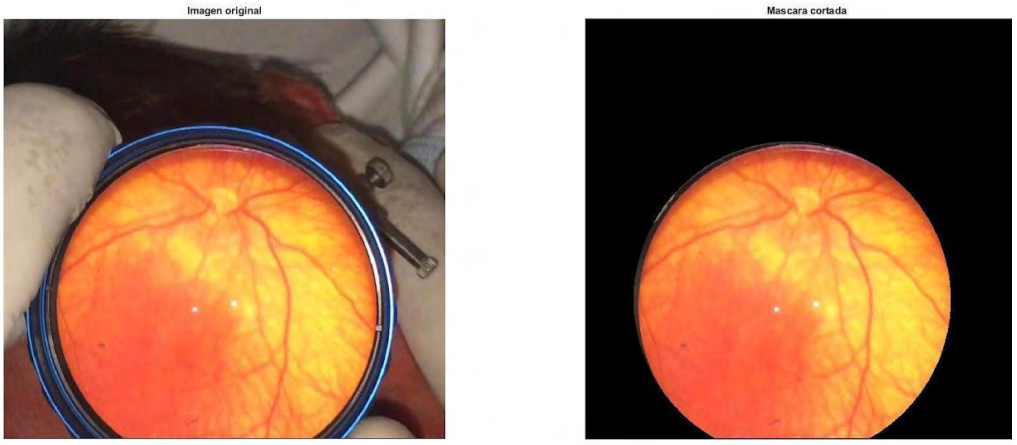

Figura 5: *Imagen original a la izquierda. Imagen con máscara a la derecha.*

Otra de las tareas realizadas fue la remoción de artefactos (principalmente, reflejos de la lupa). Se utilizó el valor de la mediana, en lugar de la media, debido a su mayor solidez frente a los valores atípicos. La medida de contraste de Weber está definida para las imágenes en escala de grises. Se aplicó esta metodología a las imágenes RGB considerando sólo el canal verde, como es la práctica habitual para las imágenes de la retina debido a su mayor contraste de vasos. La figura 3 muestra los resultados de la aplicación de este filtrado. Se puede ver que los artefactos de iluminación, que se encuentran en el centro de la lupa fueron eliminados, a costo de obtener una imagen ligeramente difuminada. Sin embargo, con un ajuste adecuado del nivel de sensibilidad del filtro puede mejorar.

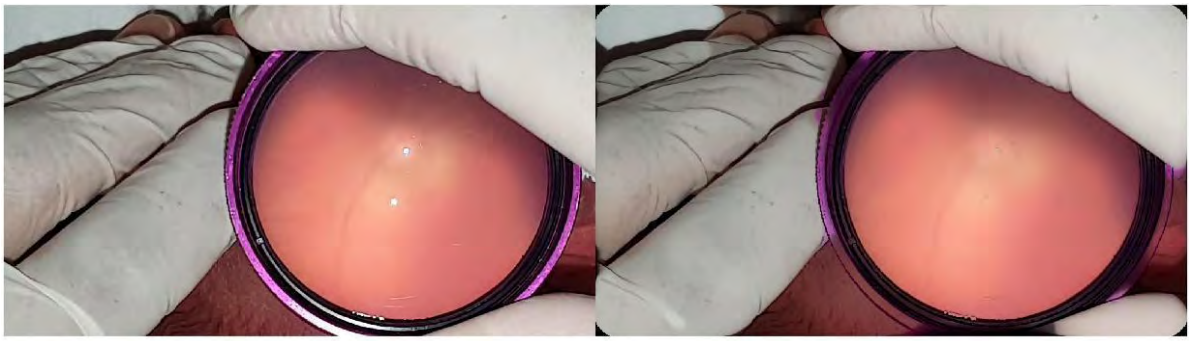

Figura 6: *Imagen original (izquierda) e imagen sin artefactos de iluminación en la lente (derecha)*

Finalmente, se procedió al realce de los vasos retinianos a los efectos de utilizarlos como marcadores para lo que será la etapa de registración y armado del mosaico. Para marcar los vasos de forma individual, se realiza en primer lugar el filtrado LoG. Para mejorar los vasos a diferentes escalas, se convoluciona la imagen original con filtros que varían en el valor de su parámetro de escala σ (entre 0.11 a 0.51 con pasos de 0.1) y conservan la respuesta máxima en cada píxel de forma de crear un banco de filtros. En la figura 4, se puede ver el resultado de la aplicación del banco de filtro a la imagen. El resultado del mismo luego se pasa por otro banco de filtros del tipo Gabor de forma análoga con la parte anterior, configurado con un ángulo de rotación variables de entre 10° a 170° con pasos de 10, donde se conserva la respuesta máxima del píxel.

Con respecto a la actividad vinculada a estandarizar las imágenes obtenidas en cuanto a brillo, contraste, realce y restauración (que no afecten al diagnóstico ROP), las principales tareas y resultados se enmarcaron en el proyecto final de la carrera Bioingeniería del becario Hernán Rodríguez. Ese trabajo titulado: *"Aplicación de técnicas de procesado de imágenes y vídeos adquiridos desde un smartphone a fondos de ojo de bebés prematuros con retinopatía del prematuro (ROP)"*, fue desarrollado en el marco de este PID y aprobado en junio de 2022.

Con respecto al ítem 2.1, se propuso y se avanzó en la implementación de un pipeline de procesamiento para mejorar la calidad de los registros, acondicionados para su documentación y análisis automático posterior.

El mismo consta de las siguientes etapas:

- Normalización
- Detección de lupa

- Clasificación de cuadros
- Remoción de artefactos
- Mapeo de vasos
- Generación de vídeo procesado
- Conformación de un mosaico retinal

La estrategia o técnicas que se decidió aplicar en cada etapa y que se mencionan a continuación se evaluaron en relación a los datos y las alternativas de procesamiento digital de imágenes disponibles.

En la etapa de normalización en caso de ser necesario se extraen los cuadros del video y se normaliza la resolución de los datos, sus relaciones de aspecto y los formatos de almacenamiento.

En la etapa de detección de lupa se realiza el reconocimiento automático de la lente de aumento que se emplea para inspeccionar la retina. Se implementó usando una secuencia de pasos que incluye: enmascaramiento de la imagen con los colores de la lente, filtrado pasa alto para realce de bordes, detector de circunferencias por Transformada Hough Circular, creación de máscara de lupa.

En la etapa de clasificación de cuadros, se busca reconocer si el cuadro considerado posee información pertinente para el análisis de ROP o no. Dicho análisis, una vez detectada la presencia de la lupa, por un lado, comprende la definición para establecer si se puede acceder a una región de la retina, y por otro respecto a si el nivel de enfoque de la imagen es adecuado. Para hacerlo se clasifica en el espacio de color HSV la información retinal y por otro lado se hace una clasificación de enfoque basada en filtros orientables.

En la etapa de remoción de artefactos se identifican y remueven regiones de la imagen afectadas por reflejos de la fuente de iluminación sobre la lupa y está implementada empleando un método estadístico basado en un umbral sobre el nivel de brillo de los píxeles en la imagen.

En la etapa de mapeo de vasos se efectúa un realce de vasos sanguíneos empleando una secuencia de filtros: Laplacianos de Gaussianos (LoG), banco de filtros Gabor, filtrado morfológico.

En la etapa de generación de video procesado se edita el video original de forma tal de resaltar la información más relevante. A partir del análisis de contenido de información de cada cuadro se sobremuestrean aquellos identificados como informativos y se submuestrean los demás de forma tal que en la reproducción final se ralentiza los fragmentos útiles para el análisis y se aceleran los que sólo brindan contexto respecto a la captura de datos.

En la etapa de conformación de mosaico retinal se submuestrea y ensambla imágenes representativas de diferentes regiones para sintetizar una panorámica de la retina mediante un mosaico final. Para la etapa de detección de características se emplearon las representaciones de SURF (*Speeded Up Robust Features*) y BRISK (*Binary Robust Invariant Scalable Keypoints*). Para la estimación de transformaciones geométricas entre imágenes alternativas se empleó el algoritmo de MLESAC (*Maximum Likelihood Estimation Sample Consensus*). Finalmente, las imágenes se ensamblan mediante un proceso de registración usando el algoritmo "*demons*", que se basa en flujo óptico.

Una vez definido e implementado el pipeline se lo empleó para procesar videos de un subconjunto de casos obtenidos a través del sistema ODI. Con los datos originales y

los procesados en sus diferentes etapas se llevó a cabo una evaluación perceptual de los resultados obtenidos. Para hacerlo se conformó un comité de oftalmólogos expertos en ROP, a quienes se solicitó comparar ambas versiones de datos para establecer empleando escalas graduadas, juicios respecto a las siguientes dimensiones: modificación de la velocidad de reproducción del video en función del contenido (informativo vs. no informativo), realce de vasos (en relación a si mejoran o no la visualización), nivel de aciertos en relación a la marcación de cuadros como informativos, utilidad del video procesado en general, utilidad de las imágenes procesadas para el diagnóstico de ROP, calidad y practicidad del mosaico retinal.

La evaluación respecto a practicidad indicó que resultó al menos aceptable en el 87% de las revisiones de los videos y en el 92,1% de las revisiones del mosaico. Y consideramos que ese estudio es valioso ya que permite conocer, desde el punto de vista de los usuarios, la valoración de los desarrollos efectuados y determinar los aspectos a mejorar.

Sin embargo, merece mencionarse que el nivel de concordancia analizado empleando Kappa de Fleiss para este estudio perceptual resultó ser menor que la esperada por azar. Esto implica la necesidad de revisar el diseño de los formularios de consulta y el nivel de instrucciones dado a los expertos.

Los resultados de ese trabajo también se publicaron en un artículo de SABI en el año 2022.

En lo referente a la actividad de obtención de *parámetros para caracterizar la calidad, portabilidad e interoperabilidad en los videos e imágenes obtenidas*, se realizaron pruebas de envíos al PACS a través del sistema implementado. Las mismas fueron videos y fotografías de distintas resoluciones y duración. Luego se realizó idéntica prueba con los estudios realizados con el ODI.

En cuanto al envío de videos de pruebas al PACS, se comenzó con fotos individuales con gran performance, notando algunos retardos en el visualizador cuando las imágenes son de muy alta resolución. Cuando fueron enviados videos el mismo presentó demoras aceptables, aunque se acentuó el problema en el visualizador. Estas demoras se deben principalmente al sistema de particionado en fotogramas y envío de cada uno como norma DICOM (Dicomización), propio de un video de larga duración con 30 fps.

Para ambos ítems se hizo un estudio bibliográfico del estado del arte en técnicas y algoritmos para la clasificación automática de ROP y segmentación de fóvea, disco óptico y vasos sanguíneos. Se revisaron las características de las bases de datos empleadas y disponibilidad de acceso público, algoritmos y desempeños reportados para diferentes estrategias, así como disponibilidad de acceso a soluciones puestas a disposición por investigadores.

Se pudo experimentar con un desarrollo basado en redes neuronales convolutivas tipo U-Nets[2] cuyo foco es la segmentación de disco óptico y copa óptica en imágenes de fondo de ojo. Se acondicionó el entorno para el entrenamiento de variantes del modelo propuesto sobre datos brindados por los investigadores, y se repitieron varias instancias de entrenamiento en las que se pudo observar los requisitos temporales del proceso y las modificaciones necesarias para su funcionamiento bajo el setup computacional disponible.

2. https://github.com/seva100/optic-nerve-cnn

Por otra parte, como la extensa mayoría de las soluciones encontradas hacen uso de cuerpos de datos basados en retinografías de buena calidad, que difieren del método de adquisición y el estándar observado en los datos recopilados hasta aquí en el proyecto, se estudiaron alternativas de aumento de datos (*data augmentation*) y transferencia de aprendizaje (*transfer learning*) reportados en la literatura para problemas afines.

En relación a la actividad 4, luego de algunos inconvenientes de seguridad informática de los servidores de nuestra facultad, se comenzó con la mudanza del sistema y PACS a un servidor propio del laboratorio, con la asignación de IP y puertos seguros. Recordemos que hasta la fecha habíamos utilizado un servidor pago con recursos propios del laboratorio.

En un primer momento fue instalado el sistema bajo licencia Windows en la unidad de disco principal, donde corre el mismo sistema operativo. Aquí nos encontramos con el primer inconveniente. El mismo consiste en que los estudios ocupan un gran espacio ya que cuentan largamente con más de 300 fotogramas. Esto hace que con pocos estudios se cubra la capacidad de almacenamiento de la unidad en cuestión, provocando el quiebre del sistema operativo.

La problemática planteada no hace más que reafirmar la necesidad de preprocesar los fotogramas previos al comienzo del proceso de dicomización. En tal sentido, en párrafos anteriores se menciona el trabajo realizado en preprocesamiento, en lo que refiere a reconocimiento de la lupa y quedarnos con la información relevante para el diagnóstico. Es decir, disminuir el tamaño de las imágenes y la importancia de descartar imágenes que no poseen dicha información relevante.

El siguiente problema que hemos detectado se encuentra en el visualizador. El visualizador seleccionado a priori fue muy lento en las pruebas, bloqueando cuando el archivo es de gran tamaño. Por tal motivo nos encontramos trabajando en la utilización de otro visualizador de fuentes abiertas como lo es OVIYAM. Oviyam es un visor DICOM basado en la web. Utilizando los protocolos DICOM estándar, se pueden consultar las listas de pacientes, recuperar series o estudios particulares y visualizarlos.

Debemos mencionar que como resultado de la tesis de posgrado del Dr. Guillermo Monteoliva se obtuvieron hasta la fecha alrededor de 300 estudios aproximadamente (insumos de este PID), considerando un buen número para realizar el estudio. Este estudio se encuentra en fase de finalización, comparando los resultados de los expertos entre imágenes obtenidas entre el ODI y el OBI.

El becario PID Tomás Rodriguez se incorporó el 6/07/22. Sus primeras incursiones fueron el estudio de las etapas y la implementación del pipeline desarrollado por el becario EVC-CIN anterior, e involucrarse en aspectos de la codificación del mismo.

Luego, teniendo en cuenta que para el desarrollo de este PID están involucradas implícitamente tareas de envío y almacenamiento de datos entre profesionales que se desenvuelven en entornos clínicos y el laboratorio es que se consideró la posible incorporación de la tecnología de blockchain, en lo que refiere a darle persistencia a los datos y metadatos de imágenes y video de ROP. En este sentido el becario, comenzó a realizar una revisión del estado del arte de la tecnología de blockchain y sus aplicaciones en el ámbito biomédico, actividad 5.

Las aplicaciones biomédicas de la tecnología blockchain incluyen el registro de historias clínicas, dispositivos portátiles y tecnología integrada, salud móvil, investigación y ensayos clínicos, cadenas de suministro médico, bases de datos biomédicas, etc. El

enfoque principal dentro del sector de la salud ha estado en los registros médicos electrónicos (EMR) [1, 2]. Los registros de atención médica basados en texto y los valores de laboratorio son mucho más fáciles de distribuir en una cadena de bloques, ya que el tamaño de los datos es mucho más pequeño que los grandes conjuntos de datos comunes en imágenes médicas.

Específicamente dentro de las imágenes médicas, los casos de uso de blockchain incluyen el intercambio de imágenes (incluida la propiedad de imágenes centradas en el paciente), la telerradiología, la investigación y las aplicaciones de aprendizaje automático/inteligencia artificial. Es más práctico almacenar hashes, metadatos o referencias/enlaces a imágenes dentro de la cadena de bloques en lugar de las propias imágenes, como se ilustra en [3] implementación de cadena de bloques propuesta para compartir imágenes. Esto es especialmente cierto debido a la lentitud y el alto costo de almacenar grandes cantidades de datos en una cadena de bloques pública. Sin embargo, los conjuntos de datos de imágenes completos podrían almacenarse dentro de una cadena de bloques privada, o podría emplearse una combinación de referencias "en la cadena de bloques" a imágenes "fuera de la cadena de bloques", como se discutió anteriormente.

En lo referido a las tareas que involucren compartir imágenes en [4] se propuso un modelo para compartir imágenes facilitado por blockchain, en el que tres transacciones de clave pública/privada en una blockchain permiten la transferencia segura de imágenes definiendo la fuente de la imagen, definiendo los propietarios correspondientes (fuente y paciente) de la imagen, y permitir el acceso a la imagen desde su fuente después de la verificación. En este marco, una imagen se "publica" como un conjunto de claves públicas/privadas al que se accede mediante una clave privada que posee el paciente. La cadena de bloques que lleva estas transacciones se utiliza para verificar que una parte solicitante, como un médico u otro hospital, esté incluida en una lista autorizada para acceder a un estudio de imágenes en particular, y que el estudio en particular corresponda a estos permisos.

Existen algunas plataformas como Cross-enterprise Document Sharing for Imaging (XDS-I) [2, 5,6] y DICOMWeb [7] que permiten compartir estudios de imágenes médicas a través de Internet. Las implementaciones de blockchain para compartir imágenes no reemplazarán dichos estándares, sino que los complementarán. Por ejemplo, una plataforma de intercambio de imágenes médicas disponible comercialmente, Nucleus.io, se implementa en la red Ethereum. Las imágenes médicas no se almacenan dentro de la propia cadena de bloques. En su lugar, se almacenan las URL de DICOMweb, lo que permite a los pacientes controlar el acceso a sus propios datos [8 ,9]. Implementaciones como esta podrían permitir la propiedad centrada en el paciente de sus propios registros médicos [10], que dependen cada vez más de las imágenes. El estándar DICOM [11] facilita el uso del cifrado, pero no lo exige explícitamente. Como tal, DICOM es lo suficientemente dinámico como para ser incorporado en muchas plataformas de cadena de bloques diferentes.

El servidor ha tenido muy buena respuesta de latencia y de acceso de datos desde distintos smartphones casi en simultáneo, sin embargo se ha notado conflictos y demoras en respuesta en la aplicación de visualización. En este punto debemos mencionar que se ensayaron tres aplicaciones de fuentes abiertas resultando OVIYAM 2.0 la que mejor rendimiento ha brindado, tanto de respuesta a la aplicación desarrollada así como también en la facilidad de relacionarse con el servidor DICOM.

Para enviar y recibir los estudios realizados con el ODI, la aplicación web mencionada solicita una validación de acceso mediante cuenta Google. Esta forma de acceso poco segura para una aplicación médica fue pensada en primera instancia para el rápido prototipo. No obstante, se deja abierta la posibilidad de un acceso seguro de doble validación para el futuro desarrollo de la aplicación.

Desafortunadamente no se ha podido medir la respuesta del servidor sobrecargando de estudios el mismo, debido a que la cantidad de estudios realizados por parte de la Asociación ROP Argentina no contaba con la cantidad suficiente de estudios ODI para todos los efectores de salud involucrados.

En cuanto al desempeño de la aplicación web, se ha tenido buena aceptación por parte del profesional médico. En el análisis se destaca la sencillez en la interfaz, fácil configuración para acceder al servidor, buena presentación de los estudios realizados y visualización posterior de los mismo para un mejor diagnóstico.

Como se ha planteado en informes de avance en este prototipo se había pensado posprocesar los videos cargados como archivos DICOM separados como una serie de imágenes individuales. Es aquí que la aplicación desarrollada recibe dichos videos, los separa en fotogramas y conforma cada imagen en un archivo “dcm” con todos sus metadatos. La duración de estos fotogramas en alta resolución hacen que cada estudio deba ser almacenado en crudo, con la demanda excesiva de almacenamiento que ello implica.

Es así que esta serie de imágenes fue el insumo para el pipeline de procesamiento desarrollado en MATLAB y que fuera motivo de una publicación. Lo destacable del mismo es la reducción de fotogramas que no aportan información del fondo ocular, disminuyendo así la extensa cantidad de los mismos por estudio y carga de almacenamiento. Esta aplicación culmina con la registración semiautomática de unos pocos fotogramas de la zona de interés del profesional.

En este punto se ha discutido la posibilidad de separar en un futuro este pipeline en dos etapas. La primera es descartar los fotogramas sin información relevante. Aquí es necesario una integración con la aplicación web desarrollada, de esta manera el servidor sólo debe almacenar aquellos donde realmente se visualice el fondo ocular en estudio, disminuyendo así la cantidad de fotogramas en la serie DICOM.

La segunda etapa se ubicaría como posible menú de la aplicación web haciendo que el médico seleccione los fotogramas a registrar y brinde el mosaico de imágenes de su interés.

En este pipeline, como se informó anteriormente, es relevante la detección de la lupa oftálmica, lugar donde realmente está la información de interés. Detectada la misma se segmenta la misma y recorta la imagen reduciendo aún más la capacidad de almacenamiento.

No obstante, el método de segmentación implementado presenta algunos inconvenientes en fotogramas en los cuales la lupa no se presenta completa o se encuentra obstruida con parte de la mano del profesional. Por otra parte este método utilizando transformada de hough enlentece la aplicación y toma un tiempo excesivo al aplicarlo a todos los fotogramas de un estudio.

Lamentablemente por la falta de la suficiente cantidad de estudios necesarios no se ha podido implementar una segmentación basada en redes neuronales o aplicar transfer learning, estimando que mejoraría notablemente la performance del algoritmo.

Marco teórico y metodológico

El presente proyecto se desarrolló en torno al procesamiento, almacenamiento, análisis y gestión de datos obtenidos en pruebas ROP. La materia prima de todo el proceso son videofilmaciones de exámenes del fondo ocular de bebés prematuros, realizadas con teléfonos inteligentes. En estos estudios el teléfono funciona como un oftalmoscopio indirecto. El médico oftalmólogo manipula con una mano la cabeza y ojo del bebé, mientras que con la otra mano sostiene una lupa que permite observar el fondo del ojo, una vez que logra hacer foco. Por lo tanto, lo que captura la cámara es la imagen indirecta, que se forma en la lupa cuando se enfoca el fondo ocular. Este procedimiento de enfoque y manipulación del bebé no es sencillo y se ha reportado que apenas un cuarto del total de estudios de oftalmología indirecta por video ha podido utilizarse para evaluaciones de ROP [1].

Por otro lado, y en referencia a esta práctica, muchos profesionales afirman que al momento de realizar un estudio, pueden observar elementos o características que luego en el soporte de registro (imagen o video) se pierden. En ese sentido, se analizó y propuso una secuencia de procesamiento para mejorar la calidad de los registros, acondicionándolos para su documentación y análisis automático posterior, de forma tal que se minimice el nivel de pérdida referido por los profesionales. Las funcionalidades mencionadas requieren aplicar sobre imágenes y videos crudos una serie de operaciones que se pueden organizar en una secuencia o pipeline de procesamiento. Dado que los médicos oftalmopediatras han manifestado la necesidad de contar tanto con la documentación de imágenes como video de los estudios que realizan se definieron dos pipelines diferentes, y se presentan en las Fig. 1 y Fig. 2.

Figura 1. Pipeline para procesamiento de imágenes

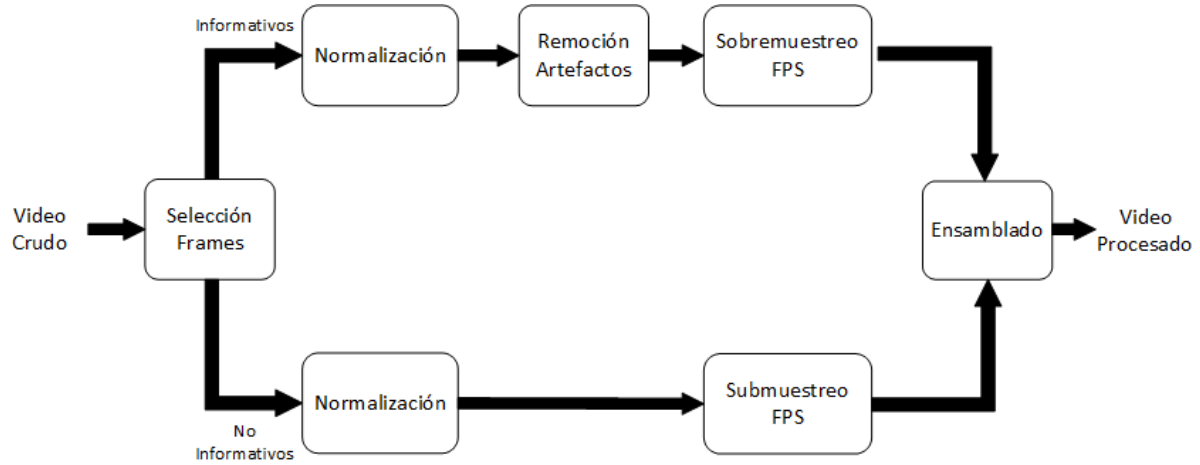

Figura 2. Pipeline prar procesamiento de videos

Para la documentación de imágenes, como se muestra en la Fig. 1 el objetivo es seleccionar las tomas que contengan imágenes claras del fondo ocular en diferentes posiciones (campos) y con ellos, generar un mosaico del fondo en toda su extensión de acuerdo a las capturas realizadas. Esa selección admite dos alternativas, se puede dar a partir de imágenes ya preseleccionadas por el profesional o a partir del video crudo. En el último caso, dado que las secuencias de video, en general, poseen pocos cuadros útiles, es decir conteniendo una imagen adecuada del fondo ocular, la primera tarea a realizar es descartar los cuadros sin información útil, aquellos donde no hay lupa, donde la lupa aparece fuera de foco, o donde aparezcan aberraciones en la imagen de la misma como reflejos u oclusiones. Para determinar los frames donde aparece la lupa obviamente se debe identificar en primer lugar la lupa. Ésta puede segmentarse de manera sencilla mediante procedimientos de detección de curvas y/o por propiedades colorimétricas de los pixeles, o por métodos más actuales como los basados en aprendizaje profundo. Asimismo, una vez efectuada esa segmentación, es necesario determinar si la imagen está enfocada o no.

El módulo de normalización busca llevar las características de las imágenes (y video) de entrada a un rango uniforme de acuerdo a sus atributos fotométricos, geométricos así como de resoluciones. Se busca compensar las diferencias de iluminación en las salas de registro, así como las diferencias en la calidad de las cámaras, y en las características de las tomas. Los dispositivos de captura de imágenes pueden tener propiedades muy heterogéneas: diferentes resoluciones espaciales (asociadas entre otras cosas, con la cantidad de píxeles de los sensores), resolución de contraste (asociada con profundidad en bits por canal), ajustes de brillo y contraste automáticos, etc. El objetivo de esta etapa es proveer a la entrada del flujo de procesamiento, imágenes comparables desde los puntos de vista enunciados previamente. Por ejemplo, definir algo tan sencillo como la relación de aspecto de las imágenes a procesar, el número de pixeles por lado (si es necesario un resampleo con los correspondientes algoritmos de interpolación), si se trabajará con los tres canales RGB o se seleccionará uno de ellos o una combinación de dos canales ponderados, el rango de bits por canal, el nivel de brillo y contraste y características cromáticas. Otro punto no menos importante es conocer el modelo de ruido que afecta a las imágenes y el consecuente filtrado (suavizado Gaussiano, filtrado de mediana y tantos otros disponibles) para mejorar la relación señal ruido.

La tercera etapa, del pipeline de procesamiento de imágenes es la de remoción de artefactos. Tiene el propósito de corregir en la medida de lo posible, rasgos de las imágenes y videos que no tienen que ver con características del fondo ocular sino con el proceso de adquisición. Como el proceso de captura de las imágenes es dinámico en el sentido que por un lado se tiene el movimiento de la cámara (y la lupa) respecto al paciente y a esto se agrega el movimiento del propio paciente, creemos conveniente incluir un paso de compensación. Dado que parte de la información que no resulta útil desde el punto de vista médico puede resultar útil para la estabilización de la imagen, o para la registración, se propone aprovechar además la versión de las imágenes previas al enmascaramiento de la lupa. A su vez en el módulo de estabilización de imagen se considera interesante plantear una operación en dos fases como proponen Klemm et al. [2] para el caso de imágenes de flujo sanguíneo cerebral: una primera etapa en la que se busca eliminar el movimiento rígido de gran escala, causado por el movimiento de la cámara con respecto al paciente, y una segunda etapa de estabilización fina de movimientos causados por el movimiento del ojo y el movimiento residual del primer paso.

El siguiente paso intenta encontrar y segmentar los vasos sanguíneos en la imagen que proyecta la lupa y hacer un seguimiento en frames sucesivos.

El quinto está formulado con el propósito de componer a partir de múltiples proyecciones del fondo ocular una panorámica de mayor resolución con los diferentes campos oculares a fin de permitir una documentación resumida y enriquecida, así como la entrada al módulo de clasificación posterior.

Para el caso del pipeline de video, mostrado en la Fig. 2, el objetivo es almacenar a manera de documentación del estudio todo el video pero maximizando el nivel de información. De esta manera se propone por un lado mejorar la calidad del video crudo y por el otro acelerar la porción del video sin información relevante y hacerlo más lento en las porciones que presenten mayores detalles sobre el fondo ocular. Esto último genera la necesidad de clasificar los frames en informativos y no informativos. A los primeros, someterlos a un proceso de realce que les permita a los profesionales visualizar una secuencia mejorada de la información útil (constituida por los cuadros que contienen la lupa con la imagen del fondo ocular enfocado) y darles la posibilidad de sobremuestrear los cuadros por segundo para observar dicha información en modo "cámara lenta". Por otro lado, los cuadros no informativos no requieren ningún procesamiento adicional al de normalización, y a los efectos de dar continuidad al video, se propone submuestrear los cuadros por segundo no informativos de manera que esta parte no útil del estudio pueda observarse en "cámara rápida". En este pipeline las operaciones del módulo de normalización son compartidas con el pipeline para imágenes.

Un detalle no menor a tener en cuenta en la selección de la propuesta está relacionado con la complejidad y requerimiento de memoria/procesamiento de cada uno de los módulos, ya que ello determina si es posible llevar adelante parte del procesamiento del lado del cliente, o por el contrario si todas estas operaciones deberán efectuarse en el servidor.

Las fotografías capturadas por los teléfonos inteligentes para que puedan interoperar en sistemas HIS/RIS deben cumplir con estándares. El estándar reconocido mundialmente para el intercambio de imágenes médicas es DICOM. Fue publicado por primera vez en 1993 por la National Electrical Manufacturers Association (NEMA) y fue desarrollado conjuntamente con el American College of Radiology (ACR). El estándar DICOM original fue denominado el estándar ACR-NEMA en referencia a estas organizaciones. Cada año se publican versiones actualizadas, actualmente consta de 21 partes diferentes, que tratan del protocolo y formatos DICOM, así como la especificación de conformidad DICOM está basada en el entorno de la Programación Orientada a Objetos (OOP), se definen objetos que incluyen imágenes, y siempre se refiere a esos objetos en el contexto de las operaciones aplicables a ellos, tales como almacenarlos, moverlos o buscarlos, crearlos, imprimirlos, etc. Los objetos en DICOM son denominados Objetos de Información (IOD), y las operaciones o servicios son llamados Clases de Servicio (Service Class). Además, se define el Par de Objetos de Servicio (SOP) que consiste en una combinación de un Servicio y un Objeto, formando el Par.

A continuación se describen los criterios establecidos y elementos para la generación de la imagen en formato DICOM. Teniendo definido el lenguaje de programación, se utilizó la librería PyDicom[3] la cual facilitó el proceso de desarrollo. A continuación, se nombran y describen los TAGS utilizados para crear el DataSet de las imágenes DI-

3. https://pydicom.github.io/

COM, los cuales se agruparon en categorías: Sintaxis de Transferencia, Imagen, Paciente, Estudio y Equipo.

Por otro lado, la sintaxis de transferencia define la manera de cómo están escritos los *Data Element*. La imagen se envía en su formato nativo y se eligió la sintaxis de transferencia en formato Big Endian con VR explícito.

Imagen

Columnas (Columns) y Filas (Rows): se obtienen de la imagen cargada.

Interpretación fotométrica (PhotometricInterpretation): especifica la interpretación prevista de los píxeles. Se definen diferentes términos: MONOCHROME, PALETTE COLOR, RGB, YBR, etc. Se estableció por defecto RGB.

Muestras por pixel (SamplesPerPixel): es el número de planos separados en la imagen. Se definen uno y tres planos de imagen. Para la interpretación fotométrica MONOCHROME y PALETTE COLOR el número de planos es 1. Para RGB y otro modelo de 3 vectores es 3.

Paciente

Nombre de Paciente (PatientName), ingresados por el usuario.

Fecha de nacimiento (PatientBirthDate)ingresado por el usuario.

Género (PatientSex): género del paciente por el usuario. Los valores posibles: O, M o F.

Id de Paciente (PatientID): es el DNI del paciente, ingresado por el usuario.

Estudio

Modalidad (Modality): Tipo de equipo que originalmente adquirió los datos utilizados para crear la imagen. Por defecto: OP (Oftalmoscopia) definido por el estándar DICOM.

Lateralidad del objeto fotografiado examinado (ImageLaterality): Ingresado por el usuario puede tomar los valores L, R o B.

Fecha de Estudio (StudyDate): fecha de creación de la imagen.

Hora de Estudio (StudyTime): hora de creación de la imagen.

Equipo

Fabricante (Make): del equipo que generó la imagen.

Modelo (Model): del equipo que generó la imagen.

Pipeline de procesamiento

El desarrollo de la evaluación del pipeline de procesamiento fue del tipo cualitativa-subjetiva. El grupo de expertos fue conformado por un grupo de 4 oftalmólogos, los cuales fueron contestando una serie de preguntas preestablecidas en formularios online. Estas preguntas estaban destinadas a calificar la practicidad de las herramientas brindadas por el algoritmo, es decir, el video procesado y el mosaico retinal. Además, los oftalmólogos contaban con una serie de rúbricas de forma de que puedan contestar las preguntas con un criterio definido.

A continuación se presentan las preguntas presentadas en los formularios y rúbricas:

Velocidad del video (informativo)		
¿Resulta útil la detención del video en puntos importantes?		
Es Impráctico (1)	Resulta indiferente (2)	Si, es útil (3)

Tabla 3.1: Pregunta referida a los cuadros seleccionados como informativos por el video, en los cuáles se detiene momentáneamente.

Velocidad del video (no informativo)		
¿Pasar en cámara rápida los cuadros no relevantes resulta práctico?		
Es Impráctico (1)	Resulta indiferente (2)	Si, es útil (3)

Tabla 3.2: Pregunta referida a los cuadros seleccionados como No informativos por el video, en los cuáles el video se reproduce más rápido.

Resaltado de vasos		
¿Qué tanto mejora la visualización de los vasos el resaltado de las imágenes?		
Empeora la visualización	Es indiferente	Mejora la visualización

Tabla 3.2: Pregunta referida a calificar la visualización de los vasos resaltados.

Características del video
No se detiene en algunos cuadros que son importantes: Esto se refiere a que pasa de largo algunos videos que se pueden identificar importantes para el diagnóstico y que tienen información que no se vió en ningún otro cuadro.
Muestra mucha información redundante: se detiene en cuadros que ya se fueron repitiendo a lo largo del video.
Hay muchos cuadros en donde se detiene el video y muestra información irrelevante: el video se detiene por error de procesamiento en cuadros que no muestran nada importante.

Tabla 3.3: Pregunta enfocada a preguntar las características de los cuadros mostrados en los videos.

Utilidad del Video Procesado		
Califique la practicidad del video procesado con respecto al video original:		
Poco útil (1)	Útil (2)	Excelente Utilidad (3)

Tabla 3.4: pregunta referida a la utilidad que les brinda el video

Diagnóstico de ROP	
¿Le sirve el video procesado para realizar diagnóstico de ROP?	
Si, sirve para diagnóstico(1)	No, no sirve para diagnóstico (2)

Tabla 3.5: pregunta referida a la posibilidad de hacer diagnóstico con el video.

3.4.1.2. Preguntas de Mosaico

Visualización de los vasos	
Calificación	Descripción
Excelente (3)	Los vasos pequeños son claramente visibles y definidos. La neovascularización y/o avascularización puede visualizarse de manera correcta, si existiese.
Buena (2)	Los vasos pequeños son claramente visibles pero no son nítidos
Inadecuada (1)	No hay estructuras anatómicas claramente visibles

Tabla 3.6: pregunta referida a calificar los vasos visualizados.

Nivel de Practicidad del Mosaico en el diagnóstico		
Califique la practicidad del Mosaico presentado:		
Poca utilidad (1)	Aceptable (2)	Excelente Utilidad (3)

Tabla 3.7: pregunta referida a calificar la practicidad del mosaico.

Artefactos en la imagen	
Calificación	Descripción
Excelente (3)	No son apreciables ningún artefacto.
Buena (2)	Los artefactos se distinguen pero no afectan al diagnóstico.
Inadecuada (1)	Los artefactos influyen de manera moderada o significativa en el diagnóstico.

Tabla 3.8: pregunta referida a calificar los artefactos encontrados en la imagen del mosaico.

Un desafío fue investigar la coherencia de puntuaciones establecidas entre oftalmólogos. La estadística Kappa propuesta por Cohen se utiliza para medir la concordancia entre dos evaluadores. Dado que en este caso fueron cuatro los evaluadores, se recurrió a Kappa de Fleiss , una generalización de la kappa de Cohen para más de dos evaluadores que determina el Nivel de concordancia entre calificadores según un

índice κ que toma valores entre 0 y 1. La máxima concordancia posible corresponde a $\kappa = 1$. El valor $\kappa = 0$ se obtiene cuando la concordancia observada es precisamente la que se espera a causa exclusivamente del azar. A la hora de interpretar el valor de κ, se dispone de una escala como la siguiente:

Valor de κ	**Nivel de concordancia**
< 0,20	Pobre
0,21 - 0,40	Débil
0,41 - 0,60	Moderada
0,61 - 0,80	Buena
0,81 - 1,00	Muy Buena

Más allá de los inconvenientes antes mencionados respecto de la conformación de la base de datos, como parte de las actividades de finalización del PID se realizó un relevamiento de datasets públicos relacionados y de los nuevos enfoques en cuanto a tareas de segmentación de imágenes que pueden ser de utilidad en una continuidad del presente proyecto.

Producto de la revisión bibliográfica se identificaron y obtuvieron algunos conjuntos de datos de imágenes de retina [2-13] que se emplearon en distintas fases del proyecto. La Tabla 1 resume alguna de las características de esos conjuntos de datos.

Dataset		Año	Resolución	N° Pacientes	Anotaciones	Condiciones
STARE		2000	605 x 700	20	Segmentación Vasos, Disco Óptico	10 SA, 10 PAT
DRIVE		2004	768 x 584	40	Segmentación Vasos	33 SA, 7 RD
ARIA		2006	576 x 768	161	Segmentación Vasos, Disco Óptico, Fóvea	61 SA, 59 RD,23 DME
REVIEW	HRIS	2008	3584 x 2438	4	Segmentación Vasos	16 RD
	VDIS		1360 x 1024	8		
	CLRIS		2160 x 1440	2		
	KPIS		288 x 119, 170 x 92	2		
CHASE_DB1		2011	990 x 960	28	Segmentación Vasos	28 SA
INSPIRE-AVR		2011	2392 x 2348	40	Segmentación Vasos, Disco Óptico	40 GLA
REFUGE		2018	2124 x 2056 1634 x 1634	1200	Disco Óptico, Fóvea	1080 SA, 120 GA
UoA-DR		2018	2124 x 2056	200	Segmentación Vasos, Disco Óptico, Fóvea	56 SA, 114 RD

IDRiD	2018	4288 X 2848	516	Disco Óptico, Fóvea	SA,DR
FIVES	2021	2048 x 2048	800	Segmentación Vasos	200 SA, 200 RD, 200 DME, 200 GLA
RVD	2022	1800 x 1800 (video)	1270	Segmentación Vasos Temporal	SA,PAT

Tabla 1. Resumen de conjuntos de datos públicos de imágenes de retina. SA: Sanos; PAT: Diferentes patologías; RD: Retinopatía Diabética; DME: Degeneración Macular Vinculada con la Edad; GLA: Glaucoma

Si bien, como se mencionó anteriormente, surgieron inconvenientes en el desarrollo del proyecto para la conformación de un conjunto de datos propios con anotaciones manuales por parte de los médicos oftalmólogos especializados en ROP, se pudo diseñar y desarrollar parcialmente la infraestructura de datos para el almacenamiento y gestión del material que se pudo adquirir en el lapso de ejecución del proyecto. Para ese desarrollo se utilizó un sistema de gestión de base de datos NoSQL. Dentro de los modelos de datos disponibles en este grupo de bases de datos se utilizó el modelo documental implementado a través de MongoDB. Desde el punto de vista del proyecto este modelo de datos jerárquico permite no solo gestionar las imágenes y videos sino también los metadatos que la describen. Asimismo también permite tratar con variaciones en la forma de presentación de las mismas dado que el modelo jerárquico utilizado se implementa en una estructura flexible que permite que los registros de paciente que se reciban puedan tener diferencias en los datos y sus tipos.

Para la implementación de los métodos de visión artificial capaces de identificar estructuras anatómicas en las imágenes de retina, en primer lugar se realizó una revisión bibliográfica, que se fue actualizando conforme se desarrollaba el proyecto.

La segmentación es una tarea fundamental en el análisis de imágenes médicas, e implica identificar y delinear regiones de interés (ROI) dentro de una imagen capturada. Tales áreas de interés pueden ser por ejemplo órganos, lesiones o tejidos específicos. Una segmentación precisa es esencial para muchas aplicaciones clínicas, ya que suele constituir el insumo que emplean posteriormente los algoritmos de diagnóstico automático. En otros casos el tamaño y la forma de la estructura anatómica será un signo que pueda emplear el médico para la planificación del tratamiento y el seguimiento de la progresión de la enfermedad [14]. La segmentación manual ha sido durante mucho tiempo el patrón oro para delinear estructuras anatómicas y regiones patológicas, pero este proceso lleva mucho tiempo, requiere mucho trabajo y a menudo exige un alto grado de experiencia. Los métodos de segmentación semiautomáticos o totalmente automáticos pueden reducir considerablemente el tiempo y el trabajo necesarios, aumentar la coherencia y permitir el análisis de conjuntos de datos a gran escala.

Con respecto a los métodos más recientes para la segmentación de estructuras anatómicas en imágenes de retina, la mayoría están basados en algoritmos de aprendizaje profundo (*deep learning*). Los modelos basados en aprendizaje profundo demostraron ser muy útiles en la segmentación de imágenes médicas debido a su capacidad para aprender características sutiles o intrincadas de la imagen y ofrecer resultados de segmentación precisos en una amplia gama de tareas, desde la segmentación de estructuras anatómicas específicas hasta la identificación de regiones patológicas [15]. Entre las aproximaciones basadas en aprendizaje profundo se puede apreciar una clara tendencia hacia la adopción de redes neuronales completamente convolucionales [16-21]. Estas

redes permiten combinar la información semántica contenida en las capas profundas con información de apariencia, capturada por las capas superficiales, y demostraron mejorar la precisión de las segmentaciones para imágenes de retina [22]. La arquitectura preferida dentro de este grupo es la de U-Nets, sin embargo muchos trabajos se centran en modificaciones de esta estructura de base [23,24]. También se pueden encontrar otros modelos aplicados a la segmentación de vasos retinianos, como las redes generativas adversativas [25], redes convolucionales en grafos [26] y el aprendizaje de ensambles [27].

Finalmente, así como en otras tareas de visión artificial, recientemente se han propuesto varios algoritmos basados en el modelo de Transformers [28], como ViT [29], Swin [30] y Mask2Former [31]. También versiones híbridas entre U-Nets y Transformers [32], con resultados prometedores.

Una limitación importante de los modelos actuales de segmentación de imágenes médicas es su naturaleza específica. Estos modelos suelen diseñarse y entrenarse para estructuras anatómicas particulares, y su rendimiento puede degradarse significativamente cuando se aplican a nuevas tareas o a diferentes tipos de datos de imágenes. Esta limitación supone un obstáculo importante para la aplicación generalizada de estos modelos en la práctica clínica. En cambio, los avances recientes en el campo de la segmentación de imágenes naturales, que se han sucedido durante la ejecución del presente proyecto, trajeron consigo el surgimiento de los denominados *modelos fundacionales* de segmentación [33, 34], que muestran una notable versatilidad y rendimiento en diversas tareas y condiciones de segmentación. Sin embargo, su aplicación a la segmentación de imágenes médicas ha supuesto un reto debido a la importante brecha de dominio [35]. Por lo tanto, existe una demanda creciente de modelos universales en la segmentación de imágenes médicas: modelos que puedan entrenarse, y una vez entrenados, puedan aplicarse a una amplia gama de tareas de segmentación. Tales modelos no sólo mostrarían una mayor versatilidad en términos de capacidad del modelo, sino que también podrían dar lugar a resultados más coherentes en las distintas tareas, al beneficiarse de una misma arquitectura subyacente y un proceso de construcción unificado. En este sentido a continuación se describen brevemente algunas herramientas recientes con las cuales se han realizado algunos ensayos orientados a incorporar este nuevo enfoque en futuros proyectos relacionados con el área de ROP.

El objetivo de esta etapa exploratoria fue comparar el desempeño de modelos estándar, que formaran parte del estado del arte en la tarea de segmentación de imágenes, con los modelos de segmentación fundacionales. Para experimentar con los modelos de segmentación estándar se eligió la plataforma OpenMMLab [34]. Entre las razones que motivaron esta elección se pueden mencionar: su versatilidad y eficiencia de ejecución, la cantidad de modelos del estado del arte preentrenados y estrategias de entrenamiento como de evaluación que pone a disposición y las funcionalidades para manipulación de datos.

Esas cualidades surgen de los objetivos planteados por ese proyecto:

- Proporcionar bibliotecas de alta calidad para reducir las dificultades en la reimplementación de algoritmos.
- Crear pipelines de despliegue eficientes dirigidas a una variedad de backends y dispositivos.
- Crear una base sólida para la investigación y el desarrollo en visión por computadora.

- Tender puentes entre la investigación académica y las aplicaciones industriales con cadenas de herramientas completas.

Esta suite, que se lanzó en octubre de 2018, está basada en PyTorch, y proporciona un motor universal de entrenamiento y evaluación, así como operadores de redes neuronales y transformaciones de datos, que sirve como base para el despliegue de sistemas de visión artificial. Está compuesta por más de 30 bibliotecas, 300 algoritmos y 2000 modelos preentrenados. El módulo específico de esta suite que se utilizó en el proyecto es OpenMMSegmentation [37].

Por otro lado, como representante de los modelos fundacionales, se eligió MedSAM [35], el primer modelo de tipo fundacional para la segmentación universal de imágenes médicas. Como se dijo, a diferencia de los modelos previos de propósitos específicos entrenados con un tipo de caso particular y cuyo desempeño cae considerablemente para imágenes fuera de su contexto, los modelos fundacionales se caracterizan por exhibir un buen desempeño para una variedad de casos diferentes. Las características de las imágenes del campo médico suponen un dominio desafiante para los modelos fundacionales debido al gap sustancial respecto a las imágenes usadas para entrenar los modelos de propósitos generales.

MedSAM es una adaptación del modelo SAM [31] entrenados con más de un millón de pares imagen médica y sus correspondientes máscaras. En su desarrollo los autores reportan experimentos exhaustivos, en los que se emplearon más de 70 tareas de validación interna y 40 tareas de validación externa, y abarcan una variedad de estructuras anatómicas, condiciones patológicas y modalidades de imágenes médicas.

Los resultados experimentales demuestran que MedSAM supera sistemáticamente al modelo fundacional de segmentación del estado del arte (SOTA), a la vez que alcanza un rendimiento a la par o incluso superior al de los modelos especializados. Estos resultados sugieren un potencial prometedor de MedSAM como herramienta para la segmentación de imágenes médicas. Su arquitectura utiliza una red basada en transformers [38], que ha demostrado una notable eficacia en diversos ámbitos, como el procesamiento del lenguaje natural y el reconocimiento de imágenes [29]. En concreto, la red incorpora un codificador de imágenes basado en *vision transformers* (ViT) encargado de extraer las características de la imagen, un codificador de prompt para integrar las interacciones del usuario (bounding boxes), y un decodificador de máscaras que genera resultados de segmentación y puntuaciones de verosimilitud a partir de la codificación de la imagen, los bounding boxes y el token de salida.

El modelo ViT que utiliza consta de 12 capas de transformers, cada una de las cuales incluye un bloque de autoatención multicabezal y un bloque de perceptrón multicapa (MLP) que incorpora la normalización de capas. Su preentrenamiento se realizó mediante un modelo de autoencodificador enmascarado, seguido de un entrenamiento totalmente supervisado en el conjunto de datos SAM.

Las imágenes de entrada (1024×1024×3) se transforman en una secuencia de parches 2D aplanados con un tamaño de 16×16×3. Esto produce un tamaño de características en la codificación de la imagen de 64×64. Es decir, que tras pasar por dicho codificador la escala es reducida en un factor de 16. Los codificadores de imagen mapean el punto de esquina de la imagen del cuadro delimitador en codificaciones vectoriales de 256 dimensiones, cada cuadro delimitador se representa mediante un par de incrusta-

ciones del punto de esquina superior izquierdo y el punto de esquina inferior derecho. Para facilitar las interacciones del usuario en tiempo real una vez calculada la codificación de la imagen el algoritmo usa una arquitectura de decodificación de máscaras, conteniendo 2 capas de transformers [9] para fusionar la incrustación de la imagen y la codificación puntual, y dos capas convolucionales transpuestas para mejorar la resolución de la incrustación hasta 256×256. Finalmente, se utiliza una activación sigmoidea, seguida de interpolaciones bilineales para ajustarse al tamaño de entrada.

Durante el preprocesamiento de datos se usaron 1.090.486 pares imagen médica-máscara para el desarrollo del modelo (no se incluyen los conjuntos de validación externa). Para la validación el conjunto de datos se dividió aleatoriamente en 80%, 10% y 10% como entrenamiento, ajuste y validación, respectivamente. Esta configuración les permitió controlar el rendimiento del modelo en el conjunto de ajuste y ajustar sus parámetros durante el entrenamiento para evitar el sobreajuste. Para la validación externa se emplearon conjuntos de datos no vistos por el modelo durante el entrenamiento. Estos conjuntos de datos proporcionan una prueba rigurosa de la capacidad de generalización del modelo, ya que representan nuevos pacientes, condiciones de imagen y, potencialmente, nuevas tareas de segmentación que el modelo no ha encontrado antes. Al evaluar el rendimiento de MedSAM en estos conjuntos de datos desconocidos, es posible esbozar cuál sería el desempeño esperado en entornos clínicos del mundo real, donde se registra una amplia gama de variabilidad en los datos.

Conclusiones

En el presente proyecto, se ha avanzado en el entendimiento y aplicación de técnicas en el análisis de imágenes oculares. Puntualmente se han utilizado herramientas específicas de análisis y procesamiento de imágenes para abordar desafíos particulares en oftalmología y se ha avanzado con aspectos relativos a la integración con Sistemas de Información tipo PACS. Esta integración facilita potencialmente la colaboración entre profesionales de la salud y mejora la calidad de la atención al paciente.

En cuanto al aspecto relativo a la integración como un modalidad oftalmológica para PACS se ha avanzado en una aplicación para la conformación e integración de un Smartphone como una nueva modalidad SCP y SCU de un PACS de imágenes médicas. Específicamente, en un nuevo instrumento para el diagnóstico digital en oftalmología respetando los estándares internacionales. Este diseño y desarrollo se orientó a los objetivos de dar valor agregado al diagnóstico, facilitando su utilización y asegurando la calidad diagnóstica, portabilidad e interoperabilidad. La aplicación Smartphone – Servidor Dicom Web, se implementó y recientemente se trabajó en las etapas de estudios de seguridad de accesos y accesos simultáneos, validación de datos enviados entre otros aspectos.

En la aplicación para móviles se han propuesto cambios para facilitar el uso por parte del experto médico. Uno de los más importantes es la incorporación de un visualizador DICOM de los estudios con la posibilidad de sincronización automática con el servidor. Esto permite que el profesional pueda elegir cuáles, y cuándo serán enviados los estudios definitivos, así como también, observar estudios anteriores. Otro aspecto importante es seguir trabajando en la autenticación de accesos y poder adicionar un prediagnóstico.

En lo referente al análisis de imágenes y videos ODI, se observa que la cantidad de píxeles útiles sólo alcanzan 1 Mpixel en promedio, esto no solo afecta a la calidad de imágenes, sino que se encuentra por debajo de lo necesario para trabajar con sistemas de procesamientos digitales. El mayor perjuicio se observa en la futura detección automática de vasos sanguíneos pequeños, recordando que esta enfermedad se caracteriza por una anormal angiogénesis.

Por los motivos antes descriptos se ha propuesto trabajar fuertemente en el sistema de captura tratando de mejorar dos áreas principales. La primera es el mejoramiento en la parte física (hard) tratando de minimizar los reflejos sobre la lupa. Una de las posibles medidas es dotar a la lupa de una fuente de luz y no utilizar la del propio smartphone, incorporando sistemas difusos o iluminación de campo oscuro. Por otra parte, trabajar con los médicos expertos sobre la posibilidad de acondicionar el ambiente de registro, por ejemplo oscureciendo el ambiente de trabajo y cambiando la relación zoom de la cámara del smartphone, versus distancia de este equipo y la lupa. El cambio de esta relación se debe orientar a que la imagen de los píxeles útiles tenga la resolución digital más próxima a la de la cámara. La segunda mejora se podría lograr aplicando técnicas de procesamiento y restauración. El objetivo de esta mejora es facilitar la detección automática de los vasos y su tortuosidad, ya que describen la progresión de la enfermedad. Es bien sabido que utilizando el canal verde del modelo RGB, se enfatiza el contraste de dichos vasos para facilitar luego la aplicación de detectores o clasificadores.

En relación al procesamiento digital de imágenes y video se desarolló un flujo de procesamiento para caracterizar y mejorar el aspecto visual de las grabaciones retinianas obtenidas en el contexto de la ROP. Los resultados perceptuales de los oftalmólogos sobre la utilidad y características de las imágenes y videos procesados sugieren que potencialmente son útiles para apoyar el proceso de toma de decisiones, así como su idoneidad para la documentación de casos. Además, tienen el potencial de convertirse en una herramienta útil para acelerar los tiempos de análisis de cada paciente y obtener más información con los datos obtenidos a través del prototipo ODI. Sin embargo, los bajos niveles de acuerdo en las evaluaciones entre los evaluadores sugieren que es necesario mejorar este procedimiento de evaluación y capacitar a los usuarios en las herramientas. En este sentido, aumentar el número de evaluadores, de videos o realizar una preselección de videos a procesar pueden ser alternativas para mejoras. Este desarrollo buscó generar una primera versión de procesamiento que sirva como referencia comparativa contra futuros desarrollos. Se encontró que especialmente las etapas de detección de enfoque y eliminación de artefactos son claramente perfectibles en futuras iteraciones del algoritmo.

Indicadores de producción

ARTÍCULOS PUBLICADOS EN REVISTAS DE DIFUSIÓN CIENTÍFICA

"Análisis de un Sistema de Información para Retinopatías del Prematuro (ROP)", A.Salvatelli, A.Hadad, D. Evin, G. Bizai, B. Franseschini, B. Drozdowicz, REVISTA ARGENTINA DE BIOINGENIERÍA, VOL. 24 (3), 2020, pag. 25 – 30. http://revistasabi.fi.mdp.edu.ar/index.php/revista/article/view/317/354, Published: 2020-11-09. ISSN 2591-376X. https://ri.conicet.gov.ar/handle/11336/133552

Sistema de Información basado en Norma Dicom para aplicaciones oftalmológicas orientadas a Retinopatías del Prematuro. Adrián Salvatelli, Alejandro Hadad, Gustavo Bizai, Diego Evin XXIV Workshop de Investigadores en Ciencias de la Computación, 28 y 29 de Abril de 2022. https://wicc2022.uch.edu.ar/wp-content/uploads/2022/09/Libro-de-Actas-WICC-2022-1.pdf ó http://sedici.unlp.edu.ar/handle/10915/143931

Application of image and video processing techniques to eye fundus with retinopathy of prematurity acquired from a smartphone. Hernán J. Rodríguez Ruiz Díaz, Gustavo Bizai, Adrián Carlos Salvatelli, Diego Evin and Alejandro Hadad, presentado en el congreso SABI 23° Congreso Argentino de Bioingeniería y 12° Jornadas de Ingeniería Clínica (SABI2022) del 13-16 Septiembre 2022 organizado por la regional San Juan, Argentina. https://sabi2022.unsj.edu.ar/wp-content/uploads/2022/09/TrabajosPorSesion.pdf

Rodríguez Ruiz Díaz, H.J., Bizai, G., Salvatelli, A., Evin, D.A., Hadad, A. (2024). Application of Image and Video Processing Techniques to Eye Fundus with Retinopathy of Prematurity Acquired from a Smartphone. In: Lopez, N.M., Tello, E. (eds). IFMBE Proceedings, vol 105. Springer, Cham. https://doi.org/10.1007/978-3-031-51723-5_52

WORKSHOP

Adrián Salvatelli, Alejandro Hadad, Gustavo Bizai, Diego Evin. "Sistema de Información basado en Norma Dicom para aplicaciones oftalmológicas orientadas a Retinopatías del Prematuro". XXIV Edición del Workshop de investigadores en Ciencias de la Computación, realizado en Mendoza, abril de 2022.

Bibliografía

[1] Ahmad, S., Wallace, D. K., Freedman, S. F., & Zhao, Z. (2008). Computer-assisted assessment of plus disease in retinopathy of prematurity using video indirect ophthalmoscopy images. Retina, 28(10), 1458-1462.

[2] Klemm, S., Rexeisen, R., Stummer, W., Jiang, X., & Holling, M. (2020). A video processing pipeline for intraoperative analysis of cerebral blood flow. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 8(4), 356-366.

[3] Hoover A, Kouznetsova V, Goldbaum M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Trans Med Imaging. 2000;19(3):203–10. https://cecas.clemson.edu/~ahoover/stare/

[4] Staal J, Abramoff MD, Niemeijer M, Viergever MA, Van Ginneken B. Ridge-based vessel segmentation in color images of the retina. IEEE Trans Med Imaging. 2004; 23(4): 501–9. https://drive.grand-challenge.org/

[5] Farnell, D. J., Hatfield, F. N., Knox, P., Reakes, M., Spencer, S., Parry, D., & Harding, S. P. (2008). Enhancement of blood vessels in digital fundus photographs via the application of multiscale line operators. *Journal of the Franklin Institute*, *345*(7), 748-765. https://petebankhead.github.io/ARIA/

[6] Al-Diri, B., Hunter, A., Steel, D., Habib, M., Hudaib, T., & Berry, S. (2008, August). A reference data set for retinal vessel profiles. In 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 2262-2265). IEEE. http://www.aldiri.info/Image%20Datasets/Review.aspx

[7] Owen CG, Rudnicka AR, Mullen R, Barman SA, Monekosso D, Whincup PH, Ng J, Paterson C. Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (CAIAR) program. Investig. Ophthalmol. Vis Sci. 2009; 50(5) : 2004–10. https://blogs.kingston.ac.uk/retinal/chasedb1/
[8] Niemeijer M, Xu X, Dumitrescu A, Gupta P, van Ginneken B, Folk J, Abramoff M. Automated Measurement of the Arteriolar-To-Venular Width Ratio in Digital Color Fundus Photographs. IEEE Trans Med Imaging. 2011 Jun 16. https://medicine.uiowa.edu/eye/inspire-datasets
[9] Fang, H., Li, F., Wu, J., Fu, H., Sun, X., Son, J., ... & Xu, Y. (2022). REFUGE2 Challenge: A Treasure Trove for Multi-Dimension Analysis and Evaluation in Glaucoma Screening. *arXiv preprint arXiv:2202.08994*. https://refuge.grand-challenge.org/Home2020/
[10] University of Auckland Diabetic Retinopathy (UoA0DR) Database The University of Auckland. https://auckland.figshare.com/articles/journal_contribution/UoA-DR_Database_Info/5985208
[11] Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., & Meriaudeau, F. (2018). Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research. Data, 3(3), 25. https://idrid.grand-challenge.org/
[12] Jin, K., Huang, X., Zhou, J., Li, Y., Yan, Y., Sun, Y., ... & Ye, J. (2022). Fives: A fundus image dataset for artificial Intelligence based vessel segmentation. Scientific Data, 9(1), 475. https://figshare.com/articles/figure/FIVES_A_Fundus_Image_Dataset_for_AI-based_Vessel_Segmentation/19688169/1
[13] Khan, M. W., Sheng, H., Zhang, H., Du, H., Wang, S., Coroneo, M., ... & Yu, X. (2024). RVD: a handheld device-based fundus video dataset for retinal vessel segmentation. Advances in Neural Information Processing Systems, 36. https://uq-cvlab.github.io/Retinal-Video-Dataset/
[14] Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods, 18(2), 203-211.
[15] Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B. A., ... & Cardoso, M. J. (2022). The medical segmentation decathlon. *Nature communications*, *13*(1), 4128.
[16] Ibrahim Atli and Osman Serdar Gedik. Sine-net: A fully convolutional deep learning architecture for retinal blood vessel segmentation. EngineeringScience and Technology, an International Journal, 24(2):271–283, 2021.
[17] Avijit Dasgupta and Sonam Singh. A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation. In 2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017), pages 248–251. IEEE, 2017.
[18] Zhexin Jiang, Hao Zhang, Yi Wang, and Seok-Bum Ko. Retinal blood vessel segmentation using fully convolutional network with transfer learning. Computerized Medical Imaging and Graphics, 68:1–15, 2018.
[19] Wei Li, Mingquan Zhang, and Dali Chen. Fundus retinal blood vessel segmentation based on active learning. In 2020 International conference on computer information and big data applications (CIBDA), pages 264–268. IEEE, 2020.

[20] Oliveira Américo, Pereira Sergio, and Silva Carlos A. Retinal vessel segmentation based on fully convolutional neural networks. Expert Systems with Applications, 112:229–242, 2018.
[21] Toufique Ahmed Soomro, Ahmed J Afifi, Junbin Gao, Olaf Hellwich, Lihong Zheng, and Manoranjan Paul. Strided fully convolutional neural network for boosting the sensitivity of retinal blood vessels segmentation. Expert Systems with Applications, 134:36–52, 2019.
[22] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, 3431–3440, 2015.
[23] Yishuo Zhang and Albert CS Chung. Deep supervision with additional labels for retinal vessel segmentation task. In Medical Image Computing and ComputerAssistedIntervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11, pages 83–91. Springer, 2018.
[24] Rui Xu, Jiaxin Zhao, Xinchen Ye, Pengcheng Wu, Zhihui Wang, Haojie Li, and Yen-Wei Chen. Local-region and cross-dataset contrastive learning for retinal vessel segmentation. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part II, 571–581. Springer, 2022.
[25] Jaemin Son, Sang Jun Park, and Kyu-Hwan Jung. Towards accurate segmentation of retinal vessels and the optic disc in fundoscopic images with generative adversarial networks. Journal of digital imaging, 32(3):499–512, 2019.
[26] Seung Yeon Shin, Soochahn Lee, Il Dong Yun, and Kyoung Mu Lee. Deep vessel segmentation by learning graphical connectivity. Medical image analysis, 58:101556, 2019.
[27] Debapriya Maji, Anirban Santara, Pabitra Mitra, and Debdoot Sheet. Ensemble of deep convolutional neural networks for learning to detect retinal vessels in fundus images. arXivpreprint arXiv:1603.04833, 2016.
[28] He, K., Gan, C., Li, Z., Rekik, I., Yin, Z., Ji, W., & Shen, D. (2023). Transformers in medical image analysis. Intelligent Medicine, 3(1), 59-78.
[29] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. Animageisworth 16x16 words: Transformers for image recognition at scale. arXivpreprint arXiv:2010.11929, 2020.
[30] ZeLiu, YutongLin, Yue Cao, Han Hu, YixuanWei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. arXivpreprint arXiv:2103.14030, 2021.
[31] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 1290–1299, 2022.
[32] Chen, B., Liu, Y., Zhang, Z., Lu, G., & Kong, A. W. K. (2023). Transattunet: Multi-level attention-guided u-net with transformer for medical image segmentation. IEEE Transactions on Emerging Topics in Computational Intelligence.
[33] Kirillov, A. et al. Segment anything. In *IEEE International Conference on Computer Vision.* 4015–4026 (IEEE, 2023).

[34] Zou, X. et al. Segment everything everywhere all at once. In *Advances in Neural Information Processing Systems* (MIT Press, 2023).
[35] Ma, J., He, Y., Li, F., Han, L., You, C., & Wang, B. (2024). Segment anything in medical images. *Nature Communications*, *15*(1), 654.
[36] MMSegmentation Contributors. MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark. https://github.com/open-mmlab
[37] Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, Zheng Zhang, Dazhi Cheng, Chenchen Zhu, Tianheng Cheng, Qijie Zhao, Buyu Li, Xin Lu, Rui Zhu, Yue Wu, Jifeng Dai, Jingdong Wang, Jianping Shi, Wanli Ouyang, Chen Change Loy, Dahua Lin (2019). MMDetection: Open MMLab Detection Toolbox and Benchmark. *arXiv preprint arXiv:1906.07155*.
[38] Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," CoRR, vol. abs/1706.03762, 2017.

PID 6205

Denominación del Proyecto
Sistema de Información basado en norma DICOM para aplicaciones Oftalmológicas orientadas a retinopatías del prematuro (ROP)

Director
Salvatelli Adrián Carlos

Codirector
Hadad Alejandro Javier

Unidad de Ejecución
Universidad Nacional de Entre Ríos

Dependencia
Facultad de Ingeniería

Contacto
adrian.salvatelli@uner.edu.ar

Cátedra/s, área o disciplina científica
Laboratorio de Sistemas de Información (ex Grupo de Investigación en Inteligencia Artificial Aplicada). Cátedras involucradas: Equipamiento para Diagnóstico por Imágenes / Inteligencia Artificial / Bases de Datos

Integrantes del proyecto
Docentes Uner: Bizai Gustavo Horacio; Evin Diego Alexis. Colaborador Interno: Drozdowicz Bartolomé. Colaborador Externo: Torres Rodrigo. Colaborador. Becario Formacion Vinculado A Pid: Gruber Maribel Maria Y Rodriguez Tomas Dario

Fechas de iniciación y de finalización efectivas
15/08/2019 y 07/12/2023
Aprobación del Informe Final por Resolución C.S. N° 192/24 (28-06-2024)

<<< Volver al inicio