



Pontificia Universidad
JAVERIANA
Cali

**DETECCIÓN DE PROBLEMAS CARDÍACOS EN IMÁGENES DIGITALES MEDIANTE
MODELOS DE APRENDIZAJE AUTOMÁTICO Y VISIÓN POR COMPUTADORA**

John Andrés Tierradentro Muñoz – código 9013993
Narly Beatriz Sánchez Caviedes - código 9015132
Luisa Fernanda Torres Beltrán - código 8989289

*Proyecto Aplicado para optar al título de
Magister en Ciencia de Datos*

Director(a)
Julián Gil

FACULTAD DE INGENIERÍA Y CIENCIAS
MAestrÍA EN CIENCIA DE DATOS
SANTIAGO DE CALI
ENERO 2026

TABLA DE CONTENIDO

1. INTRODUCCIÓN
2. CONTEXTUALIZACIÓN DEL PROYECTO
 - 2.1. DEFINICIÓN DEL PROBLEMA
 - 2.1.1. PLANTEAMIENTO DEL PROBLEMA
 - 2.1.2. FORMULACIÓN DEL PROBLEMA
 - 2.2. OBJETIVOS DEL PROYECTO
 - 2.2.1. OBJETIVO GENERAL
 - 2.2.2. OBJETIVOS ESPECÍFICOS
 - 2.3. MARCO DE REFERENCIA
 - 2.3.1. MARCO TEÓRICO
 - 2.3.2. ANTECEDENTES
3. METODOLOGÍA
 - 3.1. DATOS Y PREPROCESAMIENTO DE IMÁGENES
 - 3.1.1. RECOLECCIÓN Y ORGANIZACIÓN DE DATOS
 - 3.1.2. PREPROCESAMIENTO MEDIANTE TÉCNICAS CLÁSICAS DE VISIÓN ARTIFICIAL
 - 3.1.3. EVALUACIÓN DEL PREPROCESAMIENTO
 - 3.2. SEGMENTACIÓN SEMÁNTICA MEDIANTE APRENDIZAJE PROFUNDO
 - 3.2.1. MODELO FULLY CONVOLUTIONAL NETWORK (FCN)
 - 3.2.2. MODELO YOLOV8 PARA VASOS CORONARIOS
 - 3.2.3. IMPLEMENTACIÓN Y OPTIMIZACIÓN DE MODELO SEGNET
 - 3.2.4. IMPLEMENTACIÓN Y OPTIMIZACIÓN DE MODELO U-NET
 - 3.2.5. MEJOR MODELO PARA SEGMENTACIÓN SEMÁNTICA
 - 3.3. MODELO YOLO PARA DETECCIÓN DE OBSTRUCCIONES
 - 3.3.1. EVALUACIÓN DE VARIANTES YOLOV8
 - 3.4. DESARROLLO DE INTERFAZ GRÁFICA
4. ANÁLISIS Y RESULTADOS
 - 4.1. DATOS Y PREPROCESAMIENTO DE IMÁGENES
 - 4.1.1. RECOLECCIÓN Y ORGANIZACIÓN DE DATOS
 - 4.1.2. PREPROCESAMIENTO DE IMÁGENES CON TÉCNICAS CLÁSICAS DE VISIÓN ARTIFICIAL
 - 4.1.3. EVALUACIÓN DE PREPROCESAMIENTO
 - 4.2. SEGMENTACIÓN SEMÁNTICA MEDIANTE APRENDIZAJE PROFUNDO
 - 4.2.1. RESULTADOS DEL MODELO FULLY CONVOLUTIONAL NETWORK (FCN)
 - 4.2.2. RESULTADOS DEL MODELO YOLOV8 PARA VASOS CORONARIOS

- 4.2.3. RESULTADOS DEL MODELO SEGNET
- 4.2.4. RESULTADOS DEL MODELO U-NET
- 4.2.5. RESULTADOS DEL MEJOR MODELO PARA SEGMENTACIÓN SEMÁNTICA
- 4.3. MODELO YOLOV8 PARA DETECCIÓN DE OBSTRUCCIONES CARDIACAS
- 4.4. DESARROLLO DE INTERFAZ GRÁFICA
- 5. CONCLUSIONES Y TRABAJOS FUTUROS
 - 5.1. CONCLUSIONES
 - 5.2. TRABAJOS FUTUROS
- 6. REFERENCIAS BIBLIOGRÁFICAS

1. INTRODUCCIÓN

Las enfermedades cardiovasculares constituyen la principal causa de muerte a nivel mundial, representando el 32% de los decesos globales según la Organización Mundial de la Salud. Su diagnóstico depende de herramientas avanzadas de imagen médica, como ecocardiogramas y angiografías coronarias, que requieren interpretación especializada. Sin embargo, este proceso puede verse afectado por errores y subjetividades, especialmente ante el creciente volumen de datos que deben analizar los profesionales de la salud.

En este contexto, tecnologías como el aprendizaje automático y la visión por computadora se presentan como soluciones clave para apoyar la práctica médica. Estas herramientas permiten analizar imágenes médicas de manera automatizada, reduciendo la posibilidad de errores e incrementando la precisión de los diagnósticos en cardiología.

Para dar respuesta a esta problemática, el proyecto se orientó al desarrollo de un sistema de apoyo diagnóstico basado en dichas tecnologías. Se abordaron etapas críticas como el preprocesamiento de imágenes, la construcción de modelos computacionales confiables y el diseño de una interfaz accesible que facilite su implementación práctica en entornos médicos.

Como resultado, se logró desarrollar un modelo computacional de precisión en la detección de patologías cardíacas, validado con un conjunto de datos de imágenes médicas. Además, se implementó una interfaz intuitiva que promueve su uso práctico en hospitales, centros médicos y entornos académicos. De esta manera, el proyecto contribuye al fortalecimiento de la medicina digital y a la mejora de la atención en salud cardiovascular.

2. CONTEXTUALIZACIÓN DEL PROYECTO

2.1. Definición del problema

Las enfermedades cardiovasculares constituyen una de las principales causas de mortalidad a nivel mundial. Se estima que 17,9 millones de personas fallecieron en 2019 como consecuencia de una enfermedad cardiovascular, lo que representa el 32% de todas las muertes globales. De estos fallecimientos, el 85% se debieron a infartos de miocardio y a accidentes cerebrovasculares [1].

Actualmente, el diagnóstico de estas enfermedades depende de métodos de imagen avanzados, como ecocardiogramas y resonancias magnéticas, cuya interpretación sigue estando en manos de especialistas médicos. Este proceso de interpretación, además de requerir alta experiencia y capacitación, está expuesto a la posibilidad de errores humanos, lo que puede llevar a diagnósticos inconsistentes y, en algunos casos, incorrectos. Además, el creciente volumen de datos generados por los sistemas de salud sobrecarga a los especialistas, quienes a menudo no disponen del tiempo suficiente para realizar un análisis exhaustivo de cada imagen, aumentando así el riesgo de diagnósticos imprecisos [2].

En este contexto, surge la necesidad de explorar la aplicación de técnicas de ciencia de datos, tales como el aprendizaje automático y la visión por computadora, como posibles soluciones para automatizar y mejorar la precisión en el diagnóstico de patologías cardíacas. Sin embargo, se enfrentan múltiples desafíos para el desarrollo de modelos robustos y confiables capaces de detectar de manera efectiva anomalías cardíacas en las imágenes médicas.

2.1.1. Planteamiento del problema

Las enfermedades cardiovasculares son responsables de un tercio de todas las muertes globales, con un alto porcentaje de casos que resultan en infartos y accidentes cerebrovasculares. Este dato no solo destaca su importancia clínica, sino que también pone de relieve la creciente presión sobre los sistemas de salud para mejorar la precisión y la rapidez en el diagnóstico de estas afecciones. En un contexto donde los avances tecnológicos han llevado al desarrollo de herramientas de imagen avanzada, como ecocardiogramas y resonancias magnéticas, la interpretación de estas imágenes sigue dependiendo en gran medida de la experiencia humana, lo cual introduce riesgos de error, subjetividad y variabilidad en los diagnósticos [3].

El problema se agrava con el aumento del volumen de datos de imágenes generadas a diario, lo cual genera una sobrecarga de trabajo para los especialistas en cardiología. Esta situación no solo compromete la calidad del diagnóstico, sino que puede afectar la precisión con la que se detectan y tratan patologías críticas en los pacientes [4]. A medida que se incrementa la demanda de análisis exhaustivos, los especialistas enfrentan el desafío de dedicar un tiempo adecuado a cada caso, aumentando el riesgo de diagnósticos incompletos o inconsistentes y, en consecuencia, el de tratamientos no óptimos.

Por otro lado, tenemos que la angiografía coronaria por rayos X es la herramienta más común para el diagnóstico y tratamiento de la enfermedad arterial coronaria. Consiste en la inyección de agentes de contraste en los vasos coronarios mediante un catéter para resaltar la estructura de los vasos coronarios. Normalmente, se graban múltiples proyecciones de rayos X 2D desde diferentes ángulos para mejorar la visualización. Los recientes avances en el desarrollo de herramientas basadas en el aprendizaje profundo prometen una mejora significativa en el diagnóstico y tratamiento de la enfermedad arterial coronaria [5].

Como alternativa para superar estos desafíos, la ciencia de datos y las técnicas de inteligencia artificial, particularmente el aprendizaje automático y la visión por computadora, ofrecen un enfoque prometedor. Estas tecnologías podrían automatizar parte del análisis de las imágenes, reduciendo la carga de trabajo para los especialistas y minimizando la subjetividad en el diagnóstico. No obstante, el desarrollo de un sistema automatizado confiable requiere de investigaciones para abordar limitaciones actuales en robustez y precisión de los modelos, así como la validación rigurosa de los mismos en entornos clínicos.

La investigación propuesta se orientó a responder cómo pueden diseñarse modelos de aprendizaje automático que, al integrarse en entornos clínicos, mejoren la precisión y consistencia en el diagnóstico de obstrucciones de vasos coronarios en imágenes cardíacas. Con ello, se busca no solo aliviar la carga de los especialistas, sino también establecer un modelo de diagnóstico más preciso, efectivo y accesible en el campo de la cardiología.

2.1.2. Formulación del problema

¿De qué manera se pueden diseñar e implementar modelos de aprendizaje automático y visión por computadora que permitan mejorar la precisión del diagnóstico de obstrucciones en vasos coronarios mediante imágenes digitales, reduciendo así la carga de trabajo y el margen de error humano en la interpretación? Adicionalmente, se tienen las siguientes preguntas de sistematización: ¿Qué pasos se deben seguir para el preprocesamiento y extracción de características en imágenes digitales que permitan la implementación de modelos automáticos de detección?, ¿Qué técnicas de aprendizaje automático y visión por computadora resultan más efectivas para la detección de obstrucciones en vasos coronarios a través de imágenes?, ¿Cómo evaluar el desempeño de los modelos automáticos en términos de precisión y eficiencia? Y ¿Cómo se puede diseñar una interfaz gráfica accesible y funcional que permita a los usuarios interactuar de forma eficiente con la herramienta de detección de obstrucciones en vasos coronarios a través del análisis de imágenes digitales?

2.2. Objetivos del proyecto

2.2.1. Objetivo general

Implementar un sistema de detección de obstrucciones cardíacas en imágenes digitales, basado en aprendizaje automático y visión por computadora, para optimizar la precisión diagnóstica, acortar los tiempos de análisis y reducir el margen de error humano en la interpretación médica.

2.2.2. Objetivos específicos

1. Desarrollar un algoritmo de preprocesamiento de las imágenes digitales que permita la extracción adecuada de características para la posterior introducción en modelos de detección automática.
2. Implementar modelos de aprendizaje automático que permitan la identificación de vasos coronarios y la detección de obstrucciones en las imágenes digitales.
3. Evaluar el desempeño de los modelos automáticos en la detección de obstrucciones en los vasos coronarios, estableciendo métricas que permitan comparar su precisión y eficiencia.
4. Desarrollar una interfaz gráfica accesible y funcional que permita a los usuarios interactuar con la herramienta de detección de obstrucciones en vasos coronarios a través de imágenes digitales.

2.3. Marco de referencia

2.3.1. Marco teórico

El presente marco teórico aborda los fundamentos clínicos, técnicos y conceptuales que sustentan el desarrollo de un sistema automatizado para la detección de obstrucciones coronarias mediante aprendizaje automático y visión por computadora. Se exponen los conceptos médicos relacionados con las enfermedades cardiovasculares, así como las principales teorías y arquitecturas de redes neuronales convolucionales empleadas en tareas de segmentación y detección de objetos en imágenes médicas.

Fundamentos clínicos de las enfermedades coronarias

Las enfermedades cardiovasculares representan la principal causa de mortalidad a nivel mundial, con un 32% de todas las muertes registradas en 2019, según la Organización Mundial de la Salud [1]. Dentro de estas patologías, las obstrucciones en los vasos coronarios son especialmente relevantes por su alta incidencia y su papel en el desarrollo de infartos agudos de miocardio y accidentes cerebrovasculares. Estas obstrucciones, o estenosis coronarias, se producen por la acumulación de placa aterosclerótica que reduce el flujo sanguíneo al corazón, pudiendo comprometer seriamente la función cardíaca.

Para su diagnóstico, se emplean diferentes modalidades de imágenes médicas como ecocardiogramas, resonancias magnéticas y, de manera destacada, la angiografía coronaria, la cual permite visualizar las arterias coronarias mediante contraste radiopaco. No obstante, la interpretación de estas imágenes suele depender de la experiencia del cardiólogo o radiólogo, lo que introduce un componente de subjetividad y variabilidad diagnóstica [8]. Esta dependencia humana puede derivar en diagnósticos inconsistentes o tardíos, por lo cual surge la necesidad de sistemas automatizados que apoyen el análisis objetivo de las imágenes médicas.

Técnicas de preprocesamiento y visión artificial clásica

El preprocesamiento de imágenes angiográficas es una etapa crítica que busca mejorar la calidad del dato de entrada, resaltar estructuras vasculares y reducir artefactos que puedan comprometer el rendimiento de los modelos de aprendizaje profundo. Para ello, se emplean diversas técnicas de visión artificial clásica:

- **Conversión a escala de grises:** Este proceso consiste en transformar una imagen de múltiples canales (RGB) a una representación de un solo canal de intensidad. En el contexto de angiografías, simplifica la complejidad computacional y permite que los algoritmos se centren en las variaciones de luminancia que representan el contraste radiopaco dentro de las arterias [29].

- **Filtro Gaussiano:** Es un operador de suavizado espacial utilizado para reducir el ruido de alta frecuencia en la imagen. Mediante la aplicación de una función gaussiana, se logra difuminar detalles irrelevantes y suavizar transiciones bruscas de intensidad, lo cual es fundamental antes de aplicar algoritmos de detección de bordes o segmentación [30].
- **Filtro de Frangi (Vesselness Filter):** Propuesto por Frangi et al. [31], Es una técnica especializada en resaltar estructuras tubulares o similares a vasos sanguíneos. Este filtro analiza la matriz Hessiana de la imagen para determinar la probabilidad de que un píxel pertenezca a una estructura lineal. Es particularmente eficaz para diferenciar las arterias coronarias del tejido circundante basándose en su morfología geométrica.
- **Umbralización mediante el método de Otsu:** Es un algoritmo de binarización automática diseñado por Nobuyuki Otsu [32], que calcula el umbral óptimo para separar una imagen en dos clases (fondo y objeto). El método de Otsu minimiza la varianza intra-clase, lo que permite segmentar el árbol vascular de manera objetiva sin intervención manual, adaptándose a las variaciones de iluminación de cada toma.
- **Operaciones morfológicas (Cierre morfológico):** El cierre morfológico consiste en la aplicación sucesiva de una dilatación seguida de una erosión [33]. Esta técnica se utiliza para rellenar pequeños huecos o interrupciones dentro de los vasos segmentados y conectar regiones cercanas, preservando la continuidad de la red coronaria.
- **Eliminación de ruido residual y objetos pequeños:** Mediante funciones especializadas como `remove_small_objects` de la librería `skimage` [34], se filtran componentes conectados que poseen un área inferior a un umbral predefinido. Esto permite eliminar artefactos aislados producidos por el ruido que no corresponden a la anatomía vascular real.
- **Eliminación de objetos en los bordes:** Esta técnica suprime aquellas estructuras que tocan los límites del cuadro de la imagen [29]. Su aplicación es vital para evitar interpretaciones erróneas de vasos sanguíneos incompletos o artefactos de captura que podrían sesgar el cálculo de parámetros biométricos en los márgenes de la angiografía.

Fundamentos de aprendizaje automático y redes neuronales convolucionales

El aprendizaje automático (Machine Learning) es una rama de la inteligencia artificial que permite a los sistemas aprender patrones y realizar predicciones a partir de datos sin ser programados explícitamente. Dentro de este campo, el aprendizaje profundo (Deep Learning) ha destacado por su capacidad de procesar grandes volúmenes de información mediante arquitecturas inspiradas en el funcionamiento del cerebro humano, conocidas como redes neuronales artificiales (ANN).

Una de las arquitecturas más relevantes para el procesamiento de imágenes es la red neuronal convolucional (Convolutional Neural Network, CNN). Estas redes aplican filtros convolucionales sobre la imagen para extraer características jerárquicas, desde bordes y texturas hasta estructuras más complejas. Gracias a esta capacidad de aprendizaje de representaciones espaciales, las CNN han demostrado un desempeño sobresaliente en tareas de segmentación, clasificación y detección de patrones médicos [13][14].

Diversos estudios evidencian el potencial de estas redes en el diagnóstico de enfermedades cardiovasculares. Martín-Isla et al. [8], por ejemplo, demostraron que las CNN pueden automatizar la segmentación y clasificación de imágenes cardíacas con una reducción significativa en la variabilidad entre observadores humanos. Rajpurkar et al. [14] reportaron resultados comparables o incluso superiores a los de especialistas humanos en la detección de obstrucciones coronarias mediante angiografías.

Segmentación semántica de imágenes médicas

La segmentación semántica constituye una de las tareas de mayor complejidad y relevancia en el ámbito de la visión por computadora aplicada a la medicina. A diferencia de la clasificación de imágenes, que asigna una categoría global a una toma, o la detección de objetos, que se limita a localizar estructuras mediante cuadros delimitadores, la segmentación semántica opera a nivel de píxel. Este proceso consiste en la asignación de una etiqueta de clase específica a cada unidad mínima de la imagen, permitiendo así una delimitación precisa de la morfología y los contornos de las estructuras anatómicas de interés.

En el contexto del análisis cardiovascular mediante angiografías, la segmentación semántica es un componente crítico por las siguientes razones:

- **Cuantificación de parámetros biométricos:** Esta técnica permite la obtención de mediciones objetivas y precisas sobre la anatomía vascular, tales como el diámetro de la luz arterial, el porcentaje de estenosis y el volumen de las placas ateroscleróticas. Dichos datos son determinantes para la toma de decisiones clínicas y la planificación de intervenciones coronarias.
- **Aislamiento y extracción de características:** Las imágenes angiográficas suelen presentar una alta complejidad visual debido a la superposición de estructuras óseas, tejidos blandos y variaciones en la densidad del contraste radiopaco. La segmentación semántica facilita el aislamiento del árbol coronario, eliminando el ruido de fondo y permitiendo que los algoritmos se concentren exclusivamente en las regiones de relevancia diagnóstica.
- **Mapeo de la topología vascular:** La capacidad de discriminar píxel a píxel permite identificar la continuidad y ramificación de los vasos sanguíneos. Esto es fundamental para

distinguir entre las diferentes arterias (como la descendente anterior o la circunfleja) y localizar con exactitud el sitio de una obstrucción.

Desde una perspectiva técnica, el desarrollo de sistemas de segmentación médica enfrenta el reto de equilibrar la comprensión del contexto global con la precisión en la localización espacial. Para ello, se emplean arquitecturas de aprendizaje profundo que integran etapas de contracción (encoder) para la extracción de características semánticas de alto nivel, y etapas de expansión (decoder) que recuperan la resolución espacial necesaria para generar máscaras de segmentación detalladas. Este enfoque permite que el sistema identifique no solo la presencia de una patología, sino también su extensión geográfica exacta dentro del dominio de la imagen médica.

Fully Convolutional Networks (FCN)

Propuesta por Long et al. [16], esta arquitectura marcó un hito en la visión por computadora al demostrar que las redes neuronales convolucionales podían entrenarse de extremo a extremo (end-to-end) para tareas de segmentación densa. La innovación fundamental de la FCN radica en la sustitución de las capas densas o completamente conectadas, situadas tradicionalmente al final de la red, por capas puramente convolucionales. Esta modificación permite que la red acepte imágenes de cualquier dimensión de entrada y produzca mapas de probabilidad espacialmente consistentes con la resolución del dato original.

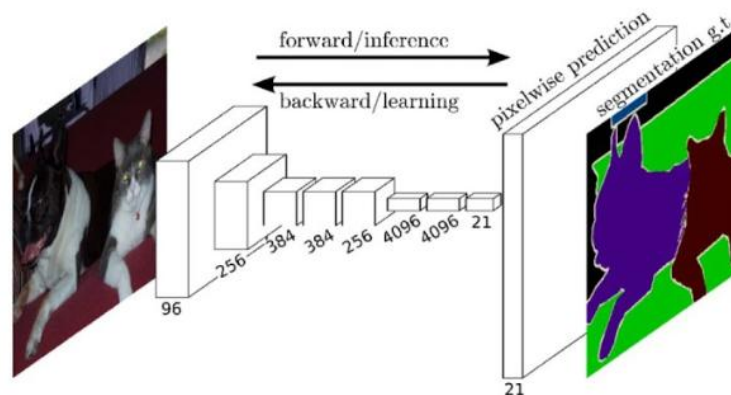


Figura 1. Arquitectura Fully Convolutional Network (FCN).
Fuente: Video *What is a Fully Convolutional Network (FCN)?*

El funcionamiento de la FCN se sustenta en el concepto de convoluciones con filtros aprendibles que actúan como extractores de características jerárquicas. En su arquitectura, el proceso de submuestreo (downsampling) permite capturar información semántica robusta; sin embargo, este proceso reduce la resolución espacial. Para contrarrestar dicha pérdida, Long et al. introdujeron capas de deconvolución o convoluciones transpuestas (upsampling), las cuales expanden las activaciones de las capas profundas hasta igualar las

dimensiones de la imagen de entrada.

En el marco de esta investigación, se ha implementado la variante FCN-8s. Esta versión destaca por el uso de conexiones de salto (skip connections), las cuales combinan la información global y abstracta de las capas más profundas con el detalle fino y local de las capas intermedias (específicamente de las etapas de pool3 y pool4). Al fusionar estas jerarquías de información, la FCN-8s logra una segmentación de los vasos coronarios con una precisión superior en la definición de los bordes, aspecto crítico cuando se busca identificar estenosis de dimensiones reducidas. Asimismo, la integración de ResNet-18 como red base (backbone) potencia esta capacidad mediante el aprovechamiento de bloques residuales que facilitan la convergencia del modelo y mitigan el problema del desvanecimiento del gradiente en arquitecturas profundas.

SegNet

Desarrollada por Badrinarayanan et al. [17], SegNet presenta una arquitectura diseñada específicamente para abordar la segmentación semántica de objetos con una eficiencia computacional optimizada en términos de memoria y precisión de contornos. Su estructura se fundamenta en una topología simétrica de codificador-decodificador (encoder-decoder). El componente codificador es idéntico a las capas convolucionales de la arquitectura VGG-16, lo que le permite extraer representaciones de características de alto nivel mediante una serie progresiva de filtros y operaciones de submuestreo.

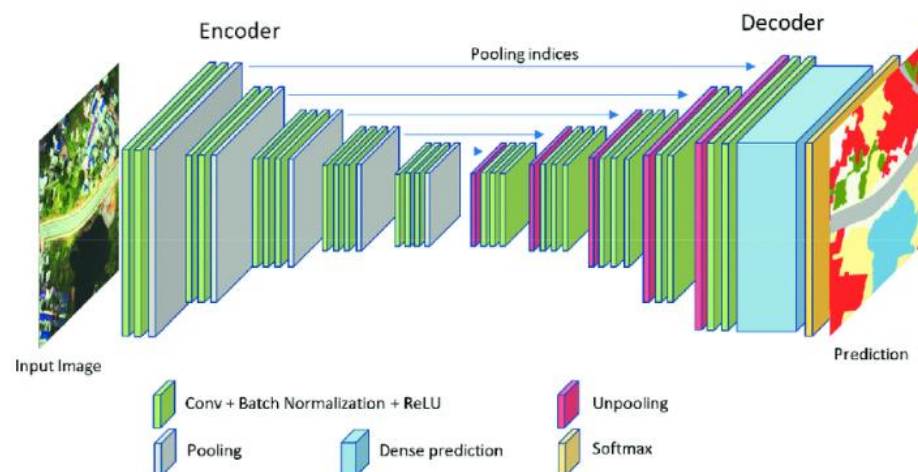


Figura 2. Arquitectura de SegNet.

Fuente: Figura del artículo SegNet architecture, extraída de ResearchGate y basada en la arquitectura de SegNet propuesta por Badrinarayanan, Kendall y Cipolla.

La innovación técnica más relevante de SegNet reside en su mecanismo de max-unpooling durante la fase de decodificación. A diferencia de otras arquitecturas que utilizan el

aprendizaje de filtros para el sobremuestreo (upsampling), SegNet registra y almacena los índices de las posiciones de los valores máximos obtenidos en las capas de max-pooling del codificador. Posteriormente, el decodificador utiliza estos índices almacenados para colocar los valores de activación en sus ubicaciones espaciales originales durante la reconstrucción de la imagen.

Este enfoque ofrece ventajas críticas para el procesamiento de imágenes médicas coronarias:

- **Preservación de la topología:** Al capturar la información de posición exacta de las características, la red es capaz de delinear con mayor nitidez los bordes de los vasos sanguíneos, cuya estructura suele ser delgada y sinuosa.
- **Eficiencia de memoria:** Al eliminar la necesidad de aprender parámetros adicionales para el sobremuestreo, se reduce la carga computacional, facilitando el entrenamiento con resoluciones de entrada estándar.

U-Net

Propuesta por Ronneberger et al. [20], U-Net se ha consolidado como una de las arquitecturas más exitosas y extendidas en el ámbito de la segmentación biomédica. Su diseño se caracteriza por una estructura simétrica de codificador-decodificador que adopta una morfología en forma de "U", diseñada específicamente para trabajar de manera eficiente con conjuntos de datos limitados, una condición común en el entorno clínico donde el etiquetado de imágenes por expertos es costoso y escaso.

La arquitectura se divide en dos trayectorias principales:

- **Trayectoria de contracción (Contracting Path):** Funciona como un codificador convencional que aplica sucesivas convoluciones y operaciones de max-pooling. Este proceso permite reducir la dimensionalidad espacial de la imagen mientras se extraen características semánticas de alto nivel, capturando el contexto global de las estructuras vasculares.
- **Trayectoria de expansión (Extensive Path):** Actúa como un decodificador que utiliza convoluciones transpuestas para aumentar la resolución de los mapas de características. Su objetivo es proyectar la información abstracta aprendida de vuelta al espacio de píxeles original para generar una máscara de segmentación precisa.

El elemento distintivo y fundamental de U-Net son las conexiones de salto (skip connections). Estas conexiones concatenan directamente los mapas de características de la ruta de contracción con los de la ruta de expansión. Esta transferencia de información permite que el decodificador recupere los detalles espaciales y de textura que se pierden durante la

reducción de resolución en el codificador.

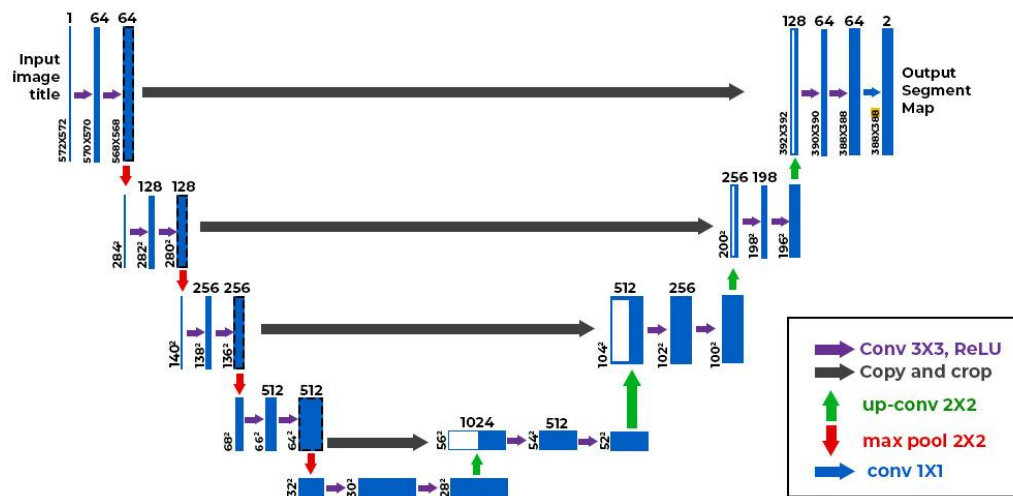


Figura 3. Arquitectura modelo U-Net.

Fuente: *GeeksforGeeks*.

En el contexto de esta investigación, la capacidad de U-Net para preservar la localización precisa es determinante para identificar estenosis coronarias, donde el margen de error entre un vaso sano y uno obstruido puede ser de apenas unos pocos píxeles. La red se adaptó para procesar la morfología sinuosa de las arterias, logrando una delimitación exacta de las áreas obstruidas incluso en angiografías con niveles moderados de ruido o variaciones en la opacidad del contraste.

Encoder (Codificador)

Constituye el segmento inicial de una arquitectura de segmentación, encargado de la extracción de características mediante la aplicación sucesiva de capas convolucionales y de reducción de resolución (pooling). Su función principal es comprimir la información espacial de la imagen de entrada en un espacio latente de menor dimensión, capturando representaciones semánticas de alto nivel [35].

Decoder (Decodificador)

Es la etapa encargada de proyectar las características abstractas aprendidas por el codificador de vuelta al espacio bidimensional original. Emplea operaciones de sobremuestreo (upsampling o convoluciones transpuestas) para reconstruir la resolución de la imagen, lo que permite la generación de una máscara de segmentación con precisión a nivel de píxel [35].

Redes Convolucionales para Clasificación de Imágenes

VGG16

Desarrollada por Simonyan y Zisserman [36], esta arquitectura de red neuronal convolucional se caracteriza por su profundidad y el uso uniforme de filtros de 3X3. Fue pionera en demostrar que la profundidad es un factor determinante en la precisión del reconocimiento visual, utilizándose frecuentemente como base (backbone) para extractores de características en modelos de segmentación médica.

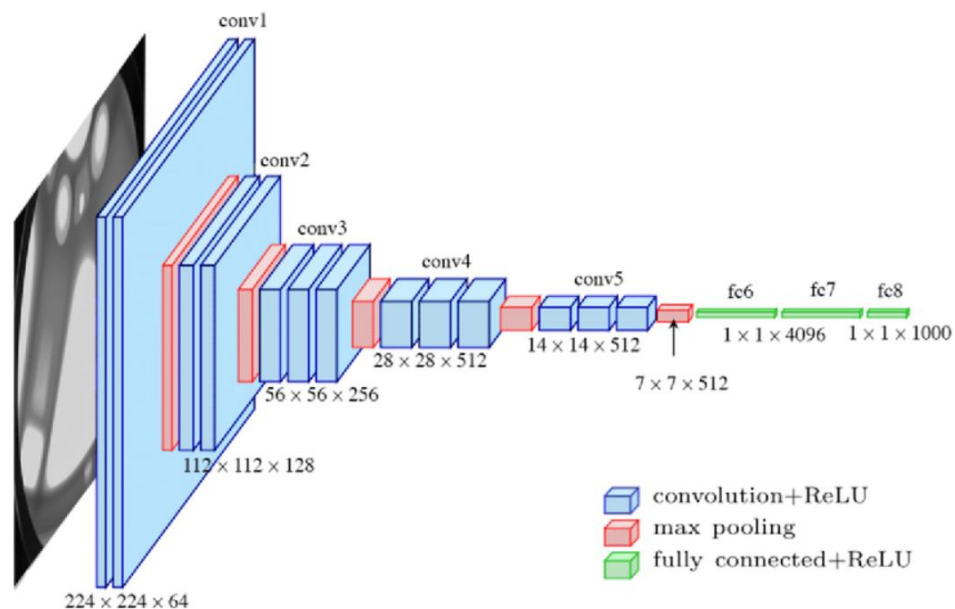


Figura 4. Arquitectura red convolucional VGG16.

Fuente: *ResearchGate*.

ResNet

Propuesta por He et al. [37], esta familia de redes introduce conexiones de salto (shortcut connections) para mitigar el problema del desvanecimiento del gradiente, permitiendo que la información fluya a través de capas extremadamente profundas.

- **ResNet-18 y 34:** Versiones ligeras que utilizan bloques residuales básicos; ideales para entornos con recursos computacionales limitados.
- **ResNet-50:** Variante más profunda que emplea bloques tipo "bottleneck" (cuello de botella), permitiendo aprender representaciones mucho más complejas sin incurrir en el desvanecimiento del gradiente.

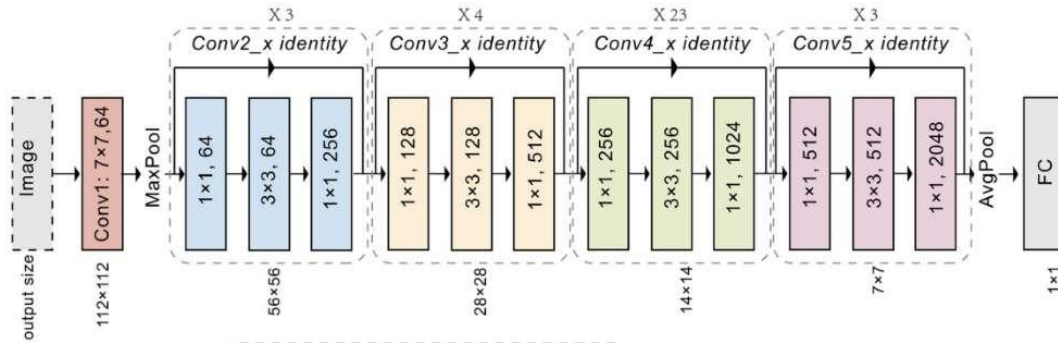


Figura 5. Arquitectura red convolucional ResNet.

Fuente: *ikomia.ai*

EfficientNet B4

Introducida por Tan y Le [38], esta arquitectura optimiza el rendimiento mediante el escalado compuesto de la profundidad, el ancho y la resolución de manera uniforme. El modelo B4 ofrece un equilibrio avanzado entre precisión y eficiencia computacional, superando a modelos convencionales en la clasificación de imágenes médicas complejas.

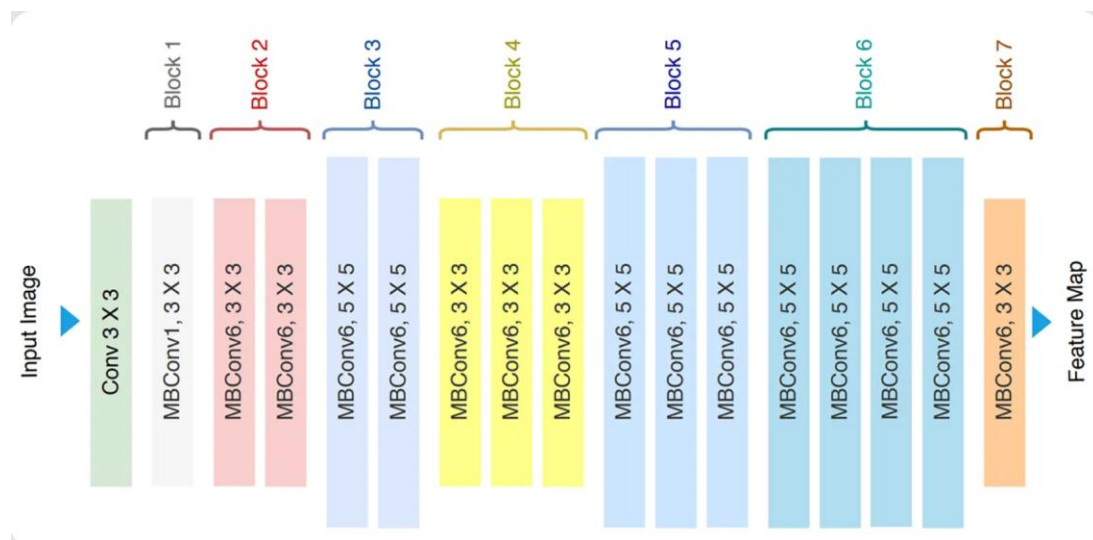


Figura 6. Arquitectura red convolucional EfficientNet.

Fuente: *Ultralytics*.

Estrategias de Entrenamiento y Optimización

- **Transfer Learning (Aprendizaje por Transferencia):** Técnica que consiste en la transferencia de conocimiento de un modelo preentrenado en un conjunto de datos masivo (como ImageNet) hacia una tarea específica de un dominio distinto [39]. En cardiología, esto acelera la convergencia y mejora la robustez del modelo ante la disponibilidad limitada de angiografías etiquetadas.
- **Data Augmentation (Aumento de Datos):** Conjunto de transformaciones geométricas y fotométricas aplicadas a las imágenes de entrenamiento para incrementar artificialmente la diversidad del conjunto de datos [40]. Su implementación es esencial para mejorar la capacidad de generalización del sistema y prevenir el sobreajuste (*overfitting*).
- **Learning Rate (Tasa de Aprendizaje):** Hiperparámetro crítico que regula la magnitud de la actualización de los pesos de la red en cada iteración del optimizador [41]. Una configuración adecuada es vital para asegurar que el modelo alcance un mínimo global en la función de pérdida de forma estable.
- **Early Stopping (Parada Temprana):** Método de regularización que monitorea el desempeño del modelo en un conjunto de validación independiente, deteniendo el proceso de entrenamiento cuando el error de validación deja de disminuir [42]. Esta técnica garantiza que el modelo mantenga su capacidad de generalización y no memorice el ruido presente en los datos de entrenamiento.

Detección de objetos aplicada al diagnóstico coronario

La detección de objetos tiene como objetivo localizar y clasificar instancias específicas dentro de una imagen mediante cuadros delimitadores. A diferencia de la segmentación semántica, que etiqueta cada píxel, la detección busca identificar regiones completas de interés. En medicina, esta técnica permite localizar lesiones o anomalías de manera rápida y eficiente.

En el presente trabajo, se utilizó YOLOv8 (You Only Look Once), un modelo de detección en tiempo real basado en redes convolucionales que combina rapidez y precisión. Su arquitectura de una sola etapa permite predecir simultáneamente las clases y ubicaciones de los objetos, optimizando el tiempo de procesamiento [18].

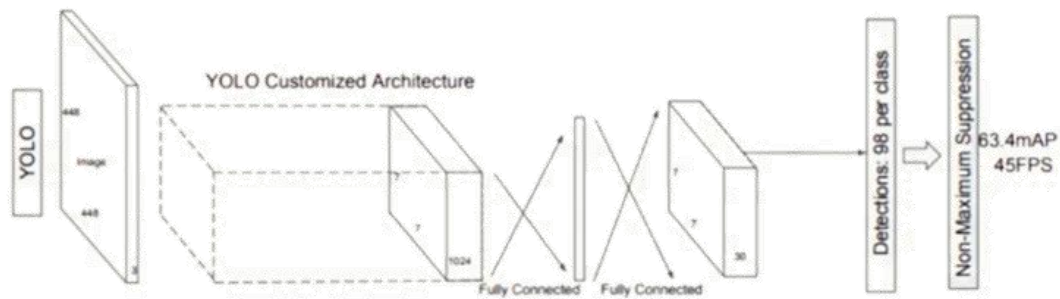


Figura 7. Esquema del modelo YOLO para detección de objetos.

Fuente: *YOLO Object Detection Explained*, Encord.

YOLOv8 fue entrenado con datos del conjunto ARCADE, transformados al formato COCO, para identificar las regiones donde se encuentran los vasos sanguíneos y facilitar así la segmentación posterior.

Este enfoque complementó las tareas de segmentación, ya que permitió localizar de forma preliminar las zonas donde se presentan posibles estenosis coronarias, integrando detección y segmentación en un flujo unificado.

Variantes de Detección YOLOv8

YOLOv8 (You Only Look Once v8) es un estado del arte en detección de objetos que permite la localización y clasificación en una sola etapa. Se presenta en diferentes escalas según la complejidad y el compromiso entre velocidad y precisión:

- **YOLOv8n (Nano):** La variante más ligera y rápida, diseñada para dispositivos móviles o sistemas con baja capacidad de cómputo.
- **YOLOv8s (Small):** Un equilibrio entre velocidad y una precisión superior a la versión nano, adecuado para aplicaciones en tiempo real de complejidad moderada.
- **YOLOv8m (Medium):** Aumenta el número de parámetros para capturar detalles más finos en la imagen, a costa de una mayor carga computacional.
- **YOLOv8l (Large):** La versión más robusta y profunda, capaz de detectar objetos pequeños o lesiones sutiles con la mayor precisión dentro de la familia YOLOv8, ideal para análisis clínicos donde la exactitud es prioritaria sobre el tiempo de respuesta instantáneo.

Métricas de evaluación en segmentación y detección

Para evaluar la efectividad de los modelos de segmentación y detección se utilizaron métricas cuantitativas ampliamente aceptadas en visión computacional y análisis de imágenes médicas: Coeficiente Dice, Índice de Intersección sobre Unión (IoU), Precisión y Sensibilidad (Recall).

El Coeficiente Dice mide la coincidencia entre la segmentación predicha y la real:

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

El IoU (o Jaccard Index) evalúa la proporción entre la intersección y la unión de las máscaras:

$$IoU = \frac{| A \cap B |}{| A \cup B |}$$

Ambas métricas son esenciales para medir la superposición entre predicciones y etiquetas reales, siendo el Dice más tolerante y el IoU más estricto frente a errores de sobre o subsegmentación [21][22].

La precisión (Precision) y la sensibilidad (Recall) se definen como:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

donde TP representa verdaderos positivos, FP falsos positivos y FN falsos negativos. La precisión evalúa la exactitud de las detecciones positivas, mientras que la sensibilidad indica la capacidad del modelo para identificar correctamente las obstrucciones [23].

En entornos clínicos, un balance adecuado entre ambas métricas es crucial, ya que tanto los falsos negativos como los falsos positivos pueden tener implicaciones diagnósticas significativas.

Implementación de la interfaz gráfica de usuario

En coherencia con los principios de accesibilidad y aplicabilidad clínica, la implementación de una interfaz gráfica de usuario (GUI, por sus siglas en inglés) constituye el componente

integrador del sistema. Este módulo cumple una función de mediación técnica y operativa, al permitir que los modelos de aprendizaje profundo desarrollados (orientados a la detección y segmentación de obstrucciones) sean operados de manera intuitiva por el usuario final sin requerir interacción directa con el código fuente. La implementación se fundamenta en la necesidad de transformar un algoritmo de visión por computadora en una herramienta de soporte a la decisión clínica, garantizando que el flujo de datos entre el procesamiento computacional y la visualización de resultados sea eficiente, fluido y comprensible en el entorno médico.

La interfaz gráfica de usuario actúa como un entorno visual dinámico que facilita la interacción entre el profesional de la salud y las arquitecturas de inteligencia artificial. Desde una perspectiva funcional, la GUI permite la abstracción de la complejidad matemática de los modelos, proporcionando un panel de control donde el facultativo puede cargar imágenes angiográficas, gestionar el almacenamiento de datos y ajustar parámetros de visualización en tiempo real.

Más allá de la simple exhibición de datos, la interfaz está diseñada para promover la interpretación asistida. Esto se logra mediante la superposición (overlay) de las máscaras de segmentación generadas por arquitecturas como U-Net o FCN sobre la imagen original, permitiendo una validación visual inmediata de las áreas de estenosis. De este modo, la interfaz no solo actúa como un receptor de información, sino como un sistema de apoyo al diagnóstico que reduce la carga cognitiva del especialista y minimiza la variabilidad interobservador al presentar los hallazgos de forma estandarizada y clara.

Biblioteca Gradio

Para el desarrollo y despliegue de la interfaz de usuario, se seleccionó la biblioteca Gradio, una infraestructura de código abierto diseñada para facilitar la creación de interfaces web interactivas destinadas a modelos de aprendizaje automático. Gradio permite la transición del modelo desde un entorno de desarrollo puramente experimental hacia una aplicación funcional y accesible, facilitando la comunicación entre el backend, desarrollado en Python, y una interfaz de usuario basada en web.

La justificación técnica de su empleo en este sistema se sustenta en los siguientes pilares:

- **Interoperabilidad y Compatibilidad:** Gradio presenta una integración nativa con los principales frameworks de aprendizaje profundo, tales como PyTorch y TensorFlow. Esto resulta fundamental para este proyecto, ya que permite cargar los pesos entrenados de las redes U-Net, SegNet y FCN de manera directa, garantizando que la inferencia de las imágenes angiográficas se realice con una latencia mínima.
- **Prototipado de Inferencia en Tiempo Real:** La biblioteca permite la creación de

componentes de entrada (como cargadores de archivos de imagen o video) y de salida (como visualizadores de máscaras de segmentación o mapas de calor) que se actualizan dinámicamente. Esto facilita que el cardiólogo visualice de forma inmediata la detección de estenosis tras cargar una angiografía, optimizando el flujo de trabajo clínico.

- **Accesibilidad y Despliegue:** Una ventaja estratégica de Gradio es su capacidad para generar enlaces compartibles y alojar la interfaz en servidores locales o en la nube (como Hugging Face Spaces) sin necesidad de desarrollar arquitecturas complejas de frontend o backend. Esto democratiza el acceso a la tecnología de diagnóstico asistido, permitiendo que el sistema sea evaluado en diferentes dispositivos dentro de una red hospitalaria.

Gracias a estas características, la biblioteca Gradio actúa como el puente tecnológico que transforma los algoritmos de segmentación y detección de objetos en una herramienta de software médico práctica, fortaleciendo el propósito del sistema de apoyar la interpretación diagnóstica de imágenes cardíacas mediante inteligencia artificial [28].

Conclusión del marco teórico

El marco teórico expuesto proporciona los fundamentos clínicos y técnicos necesarios para comprender la importancia del diagnóstico automatizado de obstrucciones coronarias. La integración de redes neuronales convolucionales, técnicas de segmentación semántica y detección de objetos demuestra el potencial de la inteligencia artificial para mejorar la precisión y eficiencia del diagnóstico médico. Asimismo, plantea los retos actuales en torno a la calidad de los datos, la interpretabilidad de los modelos y su validación en entornos clínicos reales, aspectos que este proyecto busca abordar mediante la aplicación de arquitecturas avanzadas y metodologías reproducibles.

2.3.2. Antecedentes

Las imágenes médicas son una herramienta fundamental en la medicina moderna, ya que permiten a los profesionales de la salud diagnosticar y tratar enfermedades de manera precisa y efectiva. Las imágenes médicas han revolucionado el campo del diagnóstico clínico, proporcionando herramientas esenciales para la visualización precisa y detallada de estructuras internas del cuerpo humano. En particular, las técnicas de imagen avanzada como la resonancia magnética (RM), la tomografía computarizada (TC), y los ecocardiogramas han permitido detectar y analizar con mayor precisión patologías cardíacas, que son una de las principales causas de morbilidad y mortalidad en el mundo. Estas herramientas permiten la visualización de detalles anatómicos y funcionales del corazón, desde las estructuras vasculares hasta la contractilidad miocárdica, facilitando diagnósticos más rápidos y precisos [6].

A pesar de los avances en la tecnología de imágenes, la interpretación sigue siendo un desafío. La interpretación de las imágenes depende en gran medida de la experiencia del radiólogo o cardiólogo, lo que puede introducir errores humanos debido a la fatiga o la subjetividad. Los modelos de aprendizaje automático han comenzado a abordar estos desafíos al reducir la dependencia de la interpretación humana [7]. Por ejemplo, Martín-Isla et al., en su trabajo denominado “Image-Based Cardiac Diagnosis With Machine Learning: A Review” [8], aborda el uso de redes neuronales profundas y otros algoritmos de aprendizaje automático para mejorar el diagnóstico de enfermedades cardíacas mediante el análisis de imágenes médicas. Se centra en cómo las herramientas avanzadas de inteligencia artificial (IA) pueden automatizar la detección y clasificación de diversas patologías cardíacas, como el diagnóstico de enfermedades arteriales y la evaluación de la función del corazón. La investigación resalta la capacidad de los modelos de visión por computadora para procesar imágenes complejas, como ecocardiogramas, resonancias magnéticas y tomografías computarizadas, con una precisión superior a la de los métodos convencionales. También analiza los desafíos relacionados con la implementación, incluyendo la necesidad de datos de alta calidad, el sesgo en los conjuntos de datos, la validación clínica y las barreras regulatorias para la adopción.

El artículo concluye que, aunque hay desafíos, el aprendizaje automático tiene un gran potencial para mejorar los resultados de diagnóstico y reducir la carga de trabajo de los especialistas. Este antecedente se asemeja a nuestro trabajo en la aplicación de algoritmos de aprendizaje automático a imágenes cardíacas. La diferencia principal es que se plantea de manera general sobre múltiples patologías y modalidades de imagen, mientras que nuestro proyecto se centró en la detección de obstrucciones coronarias mediante angiografías. Esta revisión nos sirvió como base para identificar retos técnicos y clínicos que debíamos considerar, como la calidad de los datos y la validación, lo que orientó decisiones en el diseño de nuestro sistema.

También se tiene el trabajo titulado “Deep Learning for Cardiovascular Medicine: A Practical Primer” [9], en donde ofrece una introducción accesible y práctica al uso del deep learning (aprendizaje profundo) en la medicina cardiovascular, explicando cómo esta tecnología emergente puede ser aplicada para transformar la atención y el diagnóstico de enfermedades cardiovasculares. El artículo tiene como objetivo proporcionar una guía clara para los profesionales médicos, investigadores y desarrolladores interesados en el uso de estas herramientas en la práctica clínica.

Se describe el contexto del aprendizaje profundo dentro del campo más amplio de la inteligencia artificial, señalando cómo los avances en el hardware computacional, el acceso a grandes cantidades de datos y los nuevos enfoques algorítmicos han impulsado su adopción en la medicina. Se destacan las diferencias entre el machine learning clásico y el deep learning, con un enfoque particular en cómo las redes neuronales profundas, especialmente las redes

neuronales convolucionales (CNN), han permitido una mejor interpretación y análisis de imágenes cardíacas.

Los autores presentan ejemplos específicos de cómo las tecnologías de deep learning están siendo utilizadas en la medicina cardiovascular, incluyendo la detección de enfermedades de las arterias coronarias, la segmentación automática de cámaras cardíacas en ecocardiografía y la predicción de eventos clínicos futuros basados en imágenes de resonancia magnética y tomografía computarizada. Se identifican varios desafíos que limitan la adopción de deep learning en la práctica médica, incluyendo la necesidad de interpretabilidad de los modelos, el sesgo en los datos de entrenamiento, y las preocupaciones sobre la transparencia y la ética. Los autores hacen hincapié en la importancia de contar con regulaciones claras y colaboraciones multidisciplinarias para superar estas barreras. Este artículo es similar a nuestro trabajo en la medida en que explora el uso del deep learning en la detección de enfermedades cardiovasculares.

La diferencia radica en que mientras ellos se enfocan en presentar un panorama general y ejemplos variados, nuestro proyecto se concentró en un caso específico: la detección de obstrucciones coronarias en angiografías. Nos fue útil porque nos permitió comprender la importancia de las CNN como técnica principal para el análisis de imágenes, lo que soportó la elección metodológica en nuestra investigación.

Adicionalmente, se presentan dos trabajos, el primero “Machine Meets Biology: A Primer on Artificial Intelligence in Cardiology and Cardiac Imaging” [10], que presenta a la inteligencia artificial (IA) como una herramienta accesible aplicada a la cardiología, enfocándose especialmente en las tecnologías de imagen cardíaca; y el segundo “Artificial Intelligence in Precision Cardiovascular Medicine” [11], que examina el papel de la IA en la medicina cardiovascular de precisión, subrayando cómo estas tecnologías permiten diagnósticos y tratamientos más personalizados para pacientes con enfermedades cardiovasculares. Ambos artículos destacan aplicaciones prácticas de la IA, pero el primero lo hace con un enfoque hacia las imágenes cardíacas y automatización de diagnósticos, mientras que el segundo profundiza en la adaptación personalizada de tratamientos.

Estos trabajos guardan similitud con nuestro proyecto en cuanto al uso de la IA en cardiología, particularmente en el análisis de imágenes y la búsqueda de diagnósticos más precisos. Sin embargo, mientras uno se centra en la generalidad de la IA aplicada a imágenes y el otro en la personalización de tratamientos, nuestro enfoque fue más delimitado: el desarrollo de un sistema para la detección de obstrucciones coronarias. Su revisión fue útil porque permitió contrastar enfoques y resaltar que, aunque hay aplicaciones amplias de la IA en medicina cardiovascular, existía un espacio poco abordado en angiografías que justificaba nuestro proyecto.

Por otro lado, el artículo “Deep Learning for Cardiovascular Imaging: A Review” [12] revisa de manera extensa el impacto del deep learning (aprendizaje profundo) en el campo de la imagenología cardiovascular. Comienza por contextualizar cómo la inteligencia artificial (IA) y, en particular, el aprendizaje profundo ha transformado la práctica médica, con un énfasis especial en la evaluación y el diagnóstico de enfermedades cardiovasculares. Estas enfermedades, debido a su alta prevalencia mundial, han sido el foco de múltiples investigaciones para optimizar los métodos de detección, diagnóstico y tratamiento.

Se destacan las redes neuronales convolucionales (CNN) como una de las herramientas más potentes para procesar imágenes médicas, particularmente en la identificación y segmentación de estructuras cardíacas. Se describen aplicaciones específicas, como la detección automática de obstrucciones coronarias, la evaluación de la función del ventrículo izquierdo, y la identificación de anomalías en estudios de resonancia magnética y tomografía computarizada.

El artículo menciona cómo el aprendizaje profundo ha mejorado la precisión diagnóstica al reducir la variabilidad entre observadores humanos. Esto resulta crucial en situaciones donde la interpretación manual de las imágenes puede estar influenciada por la experiencia o el sesgo del especialista. A pesar de los beneficios, se señala que los algoritmos de aprendizaje profundo requieren grandes cantidades de datos de entrenamiento, lo que puede ser una barrera en situaciones donde el acceso a imágenes etiquetadas es limitado.

Se exploran las posibilidades futuras de incorporar sistemas de deep learning para asistir a los médicos en tiempo real, mejorar la eficiencia de los diagnósticos, e integrar las evaluaciones de imágenes con otros datos clínicos para generar diagnósticos personalizados. En conclusión, el artículo destaca el potencial revolucionario del aprendizaje profundo en la imagenología cardiovascular al ofrecer precisión, eficiencia y nuevas perspectivas para el manejo de enfermedades cardíacas. Sin embargo, también enfatiza la necesidad de enfrentar desafíos técnicos, éticos y de implementación antes de que estas herramientas puedan ser utilizadas de manera rutinaria en la práctica clínica.

Este antecedente es el que más se acerca a nuestro trabajo, pues aborda directamente la aplicación del deep learning en imágenes cardiovasculares, incluyendo la detección de obstrucciones coronarias. La diferencia es que mientras el artículo presenta un panorama amplio de técnicas y aplicaciones, nuestro proyecto se concentró en la implementación práctica de un modelo computacional validado en un conjunto de datos específico. Este antecedente fue fundamental porque confirmó la relevancia de nuestro tema y nos sirvió como soporte académico para justificar la pertinencia del proyecto.

Estado del arte en segmentación y detección coronaria

En el ámbito de la segmentación de arterias coronarias y la detección cuantitativa de

estenosis, diversos modelos basados en aprendizaje profundo han sido propuestos en la literatura reciente. Como se muestra a continuación en la Tabla 1, con el fin de contextualizar y contrastar los resultados obtenidos en este estudio, se consideran como referencia los modelos SAM-VMNet, UNet++ y UNet, reportados en el artículo A Deep Learning Model for Coronary Artery Segmentation and Quantitative Stenosis Detection in Angiographic Images, publicado en Medical Physics (2025) [24]. Adicionalmente, se incluye el modelo Multi-ADS-Net, propuesto en el artículo Multi-attention Dynamic Sampling Network (Multi-ADS-Net): Cross-dataset Pre-trained Model for Generalizable Vessel Segmentation in X-ray Coronary Angiography, publicado en Medical Image Analysis (2025) [25].

Tabla 1. Modelos de segmentación de vasos coronarios

Modelo	F1 Score	IoU	Precisión	Recall
SAM-VMNet	0.7733	0.6303	—	0.7343
UNet++ (benchmark)	0.7387	0.5856	—	0.7143
UNet (benchmark)	0.5523	0.5153	—	0.5325
Multi-ADS-Net	0.8229	0.7045	0.8287	0.8280

Fuente: Elaboración propia.

Asimismo, en la detección de estenosis coronaria, los modelos basados en YOLO han mostrado resultados relevantes en distintos estudios, como se muestra a continuación en la Tabla 2, con el objetivo de evaluar la eficacia de los modelos desarrollados en este estudio, se realizó una comparación directa con resultados previamente reportados en la literatura. En particular, se consideraron los siguientes modelos de referencia: YOLOv8x, propuesto en el artículo publicado en Scientific Data (2024) [26], y YOLOv3, reportado en un estudio disponible en arXiv (2025) [27].

Tabla 2. Modelos de detección de estenosis coronaria

Modelo YOLO	Precisión	Recall	mAP50	mAP50-95	F1-score
YOLOv8	95.27 %	93.89 %	96.92 %	91.03 %	—
DCA-YOLOv8	96.62 %	95.06 %	97.60 %	92.74 %	—
YOLOv3	—	0.18	0.254	0.068	—
YOLOv9-E	—	—	—	—	0.417
YOLOv11-X	—	—	—	—	0.7826

Fuente: Elaboración propia.

Sin embargo, es importante precisar que solo el modelo YOLOv3 reportado en la literatura fue entrenado y evaluado utilizando el dataset ARCADE, lo que permite una comparación directa y válida con los resultados obtenidos en este trabajo. En contraste, YOLOv8 y DCA-YOLOv8

fueron desarrollados y evaluados con conjuntos de datos diferentes, incluyendo datasets generales como COCO o bases de datos clínicas propias de angiografía coronaria, por lo que sus resultados no son directamente comparables con aquellos obtenidos sobre ARCADE.

3. METODOLOGIA

El desarrollo metodológico del proyecto se estructuró en tres fases claramente diferenciadas. En la primera fase se implementaron técnicas clásicas de visión artificial con el propósito de establecer una línea base exploratoria, que permitiera analizar las limitaciones inherentes a este tipo de enfoques en la segmentación de vasos sanguíneos en imágenes angiográficas. En la segunda fase se abordó el problema principal mediante modelos de aprendizaje profundo, orientados tanto a la segmentación semántica de estructuras vasculares como a la detección de obstrucciones coronarias.

Finalmente, en una tercera fase, se desarrolló una interfaz gráfica interactiva que integra los modelos entrenados, permitiendo la carga de imágenes angiográficas, la ejecución de los algoritmos de segmentación y detección, y la visualización de los resultados de forma estructurada y accesible para el usuario final.

Esta separación metodológica permite no solo evaluar el desempeño de cada enfoque de manera independiente, justificar la adopción de arquitecturas profundas como solución principal del sistema propuesto y consolidar los resultados en una herramienta funcional de apoyo al análisis de imágenes médicas.

3.1 Datos y preprocesamiento de imágenes

El primer objetivo específico tiene como finalidad desarrollar un algoritmo de preprocesamiento de imágenes digitales para la extracción de características relevantes. Para esto se diseñó e implementó un conjunto de procedimientos aplicando técnicas clásicas de visión por computadora. Este proceso incluyó la recolección estructurada de datos provenientes de una base de datos especializada, así como la aplicación de técnicas de procesamiento de imágenes que permiten mejorar la calidad visual, reducir el ruido y facilitar la segmentación de las regiones de interés. Dichas técnicas permitieron tener como resultado una línea base para comparar posteriormente con los sistemas de aprendizaje automático destinados a la segmentación semántica.

3.1.1 Recolección y organización de datos

Se tiene como fuente de datos el dataset denominado ARCADE (Automatic Region-based Coronary Artery Disease diagnostics using X-ray angiography images Dataset), publicado en el repositorio científico Zenodo (<https://doi.org/10.5281/zenodo.10390295>). Esta base de datos está compuesta por dos conjuntos de datos independientes, diseñados para abordar tareas complementarias dentro del diagnóstico de enfermedad coronaria, enfocados en la clasificación de vasos coronarios y la detección de estenosis.

El primer conjunto, denominado SYNTAX, está orientado a la segmentación semántica de los

vasos sanguíneos y proporciona imágenes junto con anotaciones a nivel de píxel que delimitan las estructuras vasculares. Este conjunto contiene un total de 1500 imágenes angiográficas, distribuidas en 1000 imágenes para entrenamiento, 200 para validación y 300 para prueba, y fue utilizado tanto para el análisis exploratorio mediante técnicas clásicas de visión artificial como para el entrenamiento y evaluación de los modelos de segmentación profunda.

El segundo conjunto, denominado STENOSIS, está orientado a la detección de obstrucciones coronarias y cuenta igualmente con 1500 imágenes angiográficas, organizadas bajo la misma partición de entrenamiento, validación y prueba. Este conjunto incluye anotaciones a nivel de objeto que permiten localizar regiones con presencia de estenosis y fue empleado exclusivamente para el entrenamiento y evaluación del modelo de detección basado en YOLOv8.

Ambos conjuntos se encuentran organizados en las carpetas train, val y test, acompañadas de archivos de anotación en formato JSON, los cuales describen la información general de cada imagen, las categorías anatómicas de interés y las segmentaciones definidas por listas de coordenadas generadas por especialistas. La información contenida se resume así:

- Información general de cada imagen (ID, nombre del archivo, tamaño).
- Las categorías (regiones anatómicas de interés).
- Las anotaciones específicas, que incluyen segmentaciones definidas por listas de coordenadas que delimitan las regiones marcadas por los especialistas.

En el caso del conjunto SYNTAX, a partir de las anotaciones se generaron máscaras binarias por imagen, que fueron utilizadas como etiquetas de referencia para la validación cuantitativa de los modelos de segmentación. Se generaron máscaras binarias por imagen que representan de forma explícita las zonas anotadas por expertos. Estas máscaras fueron empleadas como referencia para la validación de los modelos automáticos de segmentación. Esta actividad tuvo como finalidad:

- Traducir la información vectorial (coordenadas) de las anotaciones a una representación matricial o de imagen.
- Permitir el uso directo de estas máscaras para comparación con el resultado a partir de técnicas clásicas de visión por computadora como también ser usadas en forma de etiquetas supervisadas en el entrenamiento de modelos de segmentación profunda.
- Facilitar la evaluación visual y cuantitativa del desempeño de los modelos en función de la coincidencia entre predicción y máscara de referencia.

Para lo anterior, se desarrolló un script en Python para recorrer las anotaciones correspondientes a cada imagen. El proceso planteado fue el siguiente:

1. Lectura del archivo JSON de anotaciones, y agrupación de todas las segmentaciones asociadas a cada imagen según su *image_id*.

2. Creación de una máscara binaria vacía (matriz de ceros del mismo tamaño que la imagen original).
3. Para cada región segmentada:
 - a. Se transformaron los puntos de la segmentación (lista de coordenadas XY) en un polígono.
 - b. Se aplicó la función `cv2.fillPoly()` de OpenCV para rellenar el polígono con valor 1 sobre la máscara binaria
4. Finalmente, la máscara fue guardada como imagen en formato PNG (valores 0 y 255) en una carpeta denominada *masks*, correspondiente a cada subconjunto (*train*, *val*, *test*).

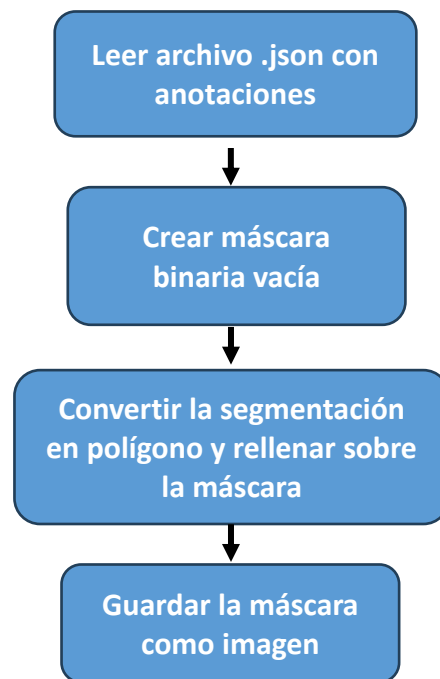


Figura 8. Diagrama de flujo para la generación de máscaras a partir del archivo de anotaciones.
Fuente: Elaboración propia.

Este procedimiento fue aplicado a la totalidad de las imágenes de cada partición del dataset, generando así una máscara por imagen con el sufijo *_mask.png*, que preserva la correspondencia directa con el nombre del archivo original.

Por otro lado, con el fin de preservar la coherencia metodológica del estudio, ambos conjuntos fueron tratados como datasets independientes, evitando la mezcla de datos y garantizando que cada tarea (segmentación y detección) fuera abordada con información específica y adecuada a su naturaleza

3.1.2 Preprocesamiento mediante técnicas clásicas de visión artificial

El objetivo de este apartado fue implementar técnicas clásicas de procesamiento de imágenes que permitan resaltar las estructuras vasculares y reducir el ruido de fondo presente en las angiografías coronarias. Estas técnicas se emplean como una estrategia exploratoria, orientada a analizar las características visuales de las imágenes antes de aplicar modelos de aprendizaje automático más complejos.

Para este propósito, se diseñó e implementó un script de preprocesamiento en Python, el cual aplica de manera secuencial distintas operaciones de mejora de contraste, reducción de ruido y segmentación de regiones de interés.

Las etapas implementadas para el preprocesamiento de las imágenes fueron las siguientes:

1. Lectura de imagen y conversión a escala de grises:

Se realiza inicialmente la carga de cada imagen. Dado que las imágenes originales contienen poca o nula información cromática relevante para el diagnóstico, se realiza una conversión a *escala de grises*, lo que permite simplificar los cálculos y enfocar el procesamiento en la información de intensidad.

2. Reducción de ruido:

Se implementó un *filtro Gaussiano*, el cual suaviza la imagen para reducir el ruido aleatorio presente, sin eliminar las estructuras relevantes como los vasos sanguíneos. Esta técnica se presentó útil para las imágenes utilizadas, donde la preservación de bordes suaves es crucial.

3. Realce de estructuras tubulares:

Para mejorar la visibilidad de los vasos coronarios, se aplicó el *Filtro de Frangi*, permitiendo resaltar estructuras alargadas y ramificadas típicas de las imágenes vasculares. Este filtro está basado en el análisis del tensor de Hessiano y permite detectar patrones que coinciden con la geometría de los vasos sanguíneos.

4. Umbralización:

Luego del realce, se aplicó una técnica de umbralización automática utilizando el *método de Otsu*, el cual calcula un valor de corte óptimo para separar las regiones de interés (vasos) del fondo. El resultado es una imagen binaria donde los píxeles de valor alto corresponden a zonas con alta probabilidad de representar estructuras vasculares.

5. Segmentación morfológica:

Para corregir posibles discontinuidades o fragmentaciones en las regiones segmentadas, se aplicó una operación morfológica de *cierre (closing)*, la cual ayuda a unir regiones cercanas y rellenar pequeños huecos dentro de las estructuras vasculares.

6. Eliminación de ruido residual:

Se utilizó la función *remove_small_objects* de la librería *skimage* para eliminar regiones pequeñas no deseadas que pudieran haber sido identificadas erróneamente como vasos. Esta operación mejora la limpieza del resultado final.

7. Eliminación de bordes:

Finalmente, se aplica una operación de eliminación de objetos ubicados en los bordes de las imágenes, ya que generalmente corresponden a regiones que no son de interés en el análisis, no representando vasos sanguíneos.

Una vez completadas estas etapas, se obtiene una máscara binaria que representa las regiones vasculares detectadas mediante el enfoque clásico.

Extracción de características morfológicas

Después de realizar las etapas de preprocesamiento y obtener los objetos de interés para el análisis, se procedió a extraer características de los vasos sanguíneos que pueden ser de utilidad. El objetivo de esta etapa fue analizar cuantitativamente las estructuras vasculares segmentadas mediante el preprocesamiento clásico, extrayendo características morfológicas básicas que describen las propiedades físicas de los vasos coronarios presentes en las imágenes angiográficas. Estas características no solo permiten entender mejor la morfología de los vasos, sino que también pueden ser utilizadas posteriormente como variables explicativas en análisis estadísticos, entrenamientos supervisados o para evaluar la evolución de la segmentación entre métodos clásicos y automáticos.

Para cada imagen procesada, se obtuvieron las siguientes métricas a partir de las regiones conectadas en la máscara binaria resultante:

- **Número de vasos detectados:** Conteo de regiones independientes resultantes de las etapas de preprocesamiento.
- **Área promedio de los vasos:** Número promedio de píxeles ocupados por cada vaso segmentado.
- **Longitud promedio:** Aproximación de la extensión mayor de cada vaso.
- **Ancho promedio:** Aproximación del grosor medio de los vasos.

Estas características permiten analizar la morfología vascular y comparar el desempeño del preprocesamiento clásico frente a enfoques automáticos posteriores.

3.1.3 Evaluación del preprocesamiento

Con el fin de evaluar cuantitativamente la calidad de las segmentaciones obtenidas mediante técnicas clásicas de procesamiento de imágenes, se definió un proceso sistemático de

evaluación basado en la comparación entre las máscaras binarias generadas automáticamente y las máscaras anotadas por expertos disponibles en la base de datos.

Definición de la grilla de búsqueda

Teniendo en cuenta las etapas del pipeline de preprocesamiento, se evaluaron distintas combinaciones de parámetros con el objetivo de identificar la configuración que maximiza la correcta detección de los vasos sanguíneos. En particular, se evaluaron las siguientes variaciones:

- **Filtro Gaussiano:** tres tamaños de kernel (5 X 5), (7 X 7), (9 X 9), orientados a la reducción del ruido presente en las imágenes.
- **Operación morfológica de cierre:** tres tamaños del elemento estructurante en forma de disco [2, 3, 4], con el objetivo de cerrar discontinuidades en las estructuras vasculares.
- **Eliminación de objetos pequeños:** tres valores mínimos de área [50, 100, 150] píxeles, para suprimir regiones irrelevantes.

La combinación de estos parámetros dio lugar a un total de 27 configuraciones distintas, las cuales fueron evaluadas de manera sistemática sobre el conjunto de validación del dataset SYNTAX.

Tabla 3. Grilla de combinaciones del preprocesamiento.

Kernel Gaussiano	Cierre morfológico	Área mínima
(5×5)	2	50
(5×5)	2	100
...
(9×9)	4	150

Fuente: Elaboración propia.

Métricas de evaluación

Con el fin de evaluar cuantitativamente la calidad de las segmentaciones obtenidas mediante técnicas clásicas de procesamiento de imágenes, se utilizaron métricas estándar de comparación entre máscaras binarias. Estas métricas permiten medir el grado de coincidencia entre las regiones detectadas automáticamente en el preprocesamiento y las máscaras anotadas por expertos disponibles en la base de datos. Se utilizaron las siguientes métricas:

- F1 Score
- IoU (Intersection over Union)
- Precisión
- Recall

Cada una de las configuraciones definidas en la grilla fue evaluada utilizando métricas de desempeño orientadas a la segmentación binaria a nivel de píxel. La selección de la configuración óptima se realizó priorizando el mejor equilibrio entre F1-score e IoU, con énfasis en maximizar la detección correcta de vasos sanguíneos sin incrementar significativamente los falsos positivos. Los resultados correspondientes a las mejores configuraciones se presentan en la sección de Resultados.

3.2 Segmentación semántica mediante aprendizaje profundo

De acuerdo con las limitaciones de los enfoques clásicos de visión artificial en la fase anterior, se procedió a abordar el problema principal del proyecto mediante técnicas de aprendizaje profundo. Esta segunda fase metodológica estuvo orientada a la segmentación semántica de vasos sanguíneos en imágenes angiográficas coronarias, utilizando el conjunto de datos SYNTAX de la base ARCADE.

El objetivo de esta etapa fue identificar la arquitectura de red neuronal más adecuada para extraer y representar de manera precisa las estructuras vasculares, permitiendo una segmentación robusta a nivel de píxel. Para ello, se adoptó una estrategia experimental progresiva, en la cual se evaluaron distintas arquitecturas de segmentación ampliamente utilizadas en visión por computadora y procesamiento de imágenes médicas, incrementando gradualmente el nivel de complejidad del modelo y las técnicas de optimización empleadas.

Esta aproximación permitió no solo comparar el desempeño de diferentes modelos, sino también comprender el impacto de factores como el aumento de datos, la transferencia de aprendizaje, la función de pérdida y la resolución espacial de las imágenes sobre la calidad de la segmentación obtenida.

En una primera instancia, se seleccionaron tres arquitecturas de segmentación semántica ampliamente reconocidas en la literatura: Fully Convolutional Network (FCN), SegNet y U-Net. Estas arquitecturas fueron elegidas por su capacidad para capturar información espacial y contextual en imágenes, así como por su uso extendido en aplicaciones de segmentación médica.

La evaluación se planteó de forma progresiva. Inicialmente se trabajó con arquitecturas base, con el fin de establecer un punto de referencia y analizar sus limitaciones. Posteriormente, se incorporaron estrategias de optimización como aumento de datos y transferencia de aprendizaje, con el objetivo de mejorar la capacidad de generalización de los modelos y su desempeño sobre imágenes no vistas.

Por otro lado, también se probó un modelo YOLO enfocado en la detección de vasos coronarios. Esto con el objetivo de explorar el funcionamiento de una arquitectura diferente y que pueda ser comparable con los modelos de segmentación semántica.

3.2.1 Modelo Fully Convolutional Network (FCN)

En el desarrollo de un sistema automatizado para la detección de obstrucciones cardíacas en imágenes digitales, se implementó un modelo de segmentación semántica basado en redes neuronales convolucionales, específicamente una Fully Convolutional Network (FCN).

Para esta tarea se eligió la arquitectura FCN-8s, una variante mejorada que permite una mayor precisión al combinar información de capas intermedias. Como base del modelo se utilizó la arquitectura ResNet-18, la cual fue adaptada mediante la incorporación de capas de convolución específicas (score_pool3, score_pool4, score_final) y capas de convolución transpuesta (upscore2, upscore4, upscore8) que permiten realizar la operación de upsampling y generar salidas en forma de mapas segmentados. A continuación, se detalla el modelo implementado:

Arquitectura

- **Encoder:** ResNet-18
- **Esquema FCN-8s (skip de medias capas):** proyecciones 1×1 desde layer4: score_final (512:1 canal), layer3: score_pool4 (256:1), layer2: score_pool3 (128:1).
- **Upsampling (decoder):** tres ConvTranspose2d encadenadas: ×2 (kernel=4, stride=2, padding=1), ×2 (mismos parámetros), ×8 (kernel=16, stride=8, padding=4), con sumas intermedias.
- **Salida:** un mapa de un canal (segmentación binaria).

Entrada y preprocesamiento

- Imágenes angiográficas en escala de grises (o 3 canales normalizados).
- Redimensionado a tamaño fijo de trabajo (256×256, múltiplo de 32 para alinear el upsampling).
- Normalización estándar previa al entrenamiento.

Función de pérdida y métricas

- **Pérdida:** BCELoss sobre la salida sigmoide (máscara binaria de referencia en formato float y shape (N,1,H,W)).
- **Métricas de seguimiento durante entrenamiento:** loss de entrenamiento y exactitud por píxel (se registró como referencia operativa). Las métricas principales para evaluación formal (Dice/IoU) se reportan en el apartado de Resultados.

Optimización y protocolo de entrenamiento

- **Optimizador:** Adam
- **Learning rate:** 0.05
- **Épocas:** 20

3.2.2 Modelo YOLOv8 para vasos coronarios

Como alternativa se planteó también evaluar si un detector de objetos (YOLOv8) podía localizar el árbol vascular coronario usando cajas delimitadoras como aproximación ligera a la segmentación.

El modelo fue entrenado utilizando el conjunto de datos ARCADE, el cual fue convertido del formato original JSON al formato COCO, requerido por la arquitectura YOLOv8. Se aplicaron transformaciones esenciales como normalización, verificación de anotaciones, ajuste de tamaño de las imágenes y validación de calidad para preparar los datos de entrada. A partir de las máscaras de vasos, se generaron bounding boxes que encapsulan las regiones vasculares visibles en cada imagen (enfoque de localización, no de delineado por píxel).

Modelo y Entrenamiento

- **Arquitectura:** YOLOv8 (anchor-free) con backbone + neck (FPN/PAN) + head multiescala.
- **Hiperparámetros base:** imgsz fijo (640/768), conf e iou por defecto de Ultralytics, optimizador y pérdidas estándar (obj/clase/caja).
- **Métricas:** mAP@0.50 y mAP@[0.50:0.95]; seguimiento de box loss (train/val).
- **Validación:** conjunto de validación estratificado, early stopping y guardado del mejor checkpoint.
- **Data augmentation (ligero):** flips horizontales, pequeñas rotaciones y ajustes de brillo/contraste (evitando deformaciones que alteren la anatomía).

La tarea se definió explícitamente como detección de vasos (localización por cajas). No se esperó obtener máscaras ni métricas por píxel (Dice/IoU), sino mAP de detección.

3.2.3 Implementación y optimización de modelo SegNet

En el desarrollo del sistema automatizado para la detección de obstrucciones cardíacas en imágenes digitales, se implementó también el modelo SegNet, una arquitectura especializada en segmentación semántica de imágenes que opera mediante un enfoque codificador-

decodificador. Esta alternativa se consideró por su capacidad de generar máscaras binarias precisas a partir de imágenes en escala de grises, lo que resulta ideal para segmentar estructuras médicas relevantes como los vasos coronarios.

Implementación modelo SegNet desde cero

El modelo SegNet fue implementado desde cero utilizando la librería Keras. La arquitectura de SegNet utilizada en este proyecto consistió en tres bloques simétricos de codificación y decodificación. En la fase de codificación, las imágenes se procesaron a través de capas convolucionales seguidas de operaciones de max-pooling, mientras que en la fase de decodificación se aplicó la operación inversa (max-unpooling) para recuperar la resolución espacial original. La entrada al modelo fueron imágenes en escala de grises con tamaño $256 \times 256 \times 1$, y la salida fue una máscara binaria del mismo tamaño que indica la presencia o ausencia de vasos sanguíneos coronarios.

Se entrenó con 1000 imágenes y se utilizaron 200 imágenes para validación. Se utilizó la función de pérdida `binary_crossentropy`, adecuada para segmentación binaria, y el optimizador Adam con una tasa de aprendizaje de $1e-4$. El entrenamiento se llevó a cabo durante 50 épocas, con un batch size de 8, y se aplicó una estrategia de early stopping para evitar el sobreajuste: si la pérdida de validación no mejoraba en 10 épocas consecutivas, el entrenamiento se detenía.

Implementación modelo SegNet + Data Augmentation

Con el fin de mejorar la capacidad de generalización del modelo, se incorporaron técnicas de Data Augmentation y se entrena nuevamente el modelo, incluyendo volteo horizontal, rotaciones aleatorias de 90° , transformaciones de desplazamiento, escalado y rotación, así como variaciones de brillo y contraste. Estas transformaciones permitieron incrementar la diversidad del conjunto de entrenamiento sin necesidad de recolectar nuevas imágenes. Los detalles se presentan a continuación:

- **HorizontalFlip**: Voltea horizontalmente (de izquierda a derecha) la imagen con un 50% de probabilidad.
- **RandomRotate90**: Rota la imagen en 90° , 180° o 270° con 50% de probabilidad.
- **ShiftScaleRotate**: Rota aleatoriamente entre -15° y $+15^\circ$, Escala aleatoriamente entre 90% y 110%, y Desplaza la imagen hasta $\pm 5\%$ del tamaño total (arriba, abajo, izquierda, derecha), con una probabilidad del 50%.
- **RandomBrightnessContrast**: Aplica cambios aleatorios de brillo y contraste con una probabilidad del 30%. Este es el único parámetro que sólo aplica para las imágenes y no a la máscara.

Implementación modelo SegNet + Data Augmentation + Transfer Learning

Adicionalmente, se evaluó una estrategia de Transfer Learning, utilizando un encoder VGG16 preentrenado sobre ImageNet, entrenando un nuevo modelo SegNet. En este caso, las capas del encoder se mantuvieron congeladas, mientras que el decodificador fue entrenado para adaptar las características extraídas al dominio de las imágenes angiográficas. Para esta configuración se utilizó la Dice Loss como función de pérdida, manteniendo el optimizador Adam con una tasa de aprendizaje de $1e-4$. Los detalles se presentan en seguida:

- **Encoder:** VGG16 preentrenado (include_top=False, pesos de ImageNet, capas congeladas).
- **Entrada:** Imágenes en escala de grises ((256, 256, 1)), adaptadas a 3 canales internamente.
- **Decoder:** Estructura simétrica tipo SegNet con UpSampling2D + Conv2D.
- **Pérdida:** Dice Loss.
- **Optimizador:** Adam con lr= $1e-4$.

3.2.4 Implementación y optimización de modelo U-Net

Para la etapa de segmentación de vasos coronarios se implementó también la arquitectura U-Net, ampliamente utilizada en el procesamiento de imágenes biomédicas por su capacidad de capturar detalles locales y globales. El modelo fue desarrollado utilizando la librería `segmentation_models_pytorch` (SMP), lo que permitió integrar de forma sencilla diferentes encoders preentrenados bajo el enfoque de transfer learning.

El preprocesamiento de los datos incluyó la normalización de las imágenes y la conversión a tres canales para ser compatibles con los encoders basados en arquitecturas entrenadas sobre ImageNet. Además, se aplicó data augmentation utilizando la librería `Albumentations`, incorporando transformaciones como volteos horizontales y verticales, rotaciones aleatorias, variaciones de brillo/contraste y escalados.

Entrenamiento de modelo inicial U-Net + Transfer Learning

Se entrenó un modelo U-Net con las siguientes características:

- Modelo preentrenado con el conjunto de datos *Imagenet*.
- Se utiliza el encoder ResNet-34 con entrada de imágenes de 3 canales.
- Las capas del encoder no se congelaron, actualizando el modelo completo durante el entrenamiento.
- El decoder es U-Net con series de bloques upsampling con conexiones de salto.

El entrenamiento se fijó con 50 épocas con una estrategia de EarlyStopping, con el optimizador

Adam con un learning rate de $1e-4$. Se utilizó Data Augmentation, aplicando volteo horizontal y vertical, rotación aleatorio de 90° , transformaciones de escala y rotación, y cambios de brillo y contraste aleatorios.

Búsqueda de hiperparámetros para modelo U-Net

Para seleccionar la configuración de U-Net, se realizó una búsqueda secuencial guiada por validación sobre un conjunto hold-out fijo. Se mantuvo constante el encoder preentrenado (ResNet-34) y el esquema de fine-tuning (warm-up con encoder congelado y posterior descongelado progresivo con learning rates diferenciadas), mientras se realizaron variaciones en el tamaño de las imágenes y prueba de funciones de pérdida. Con todo lo anterior, se entrenaron tres modelos U-Net, con las siguientes diferencias:

1. **Modelo 1:** Imágenes de entrada de tamaño 256×256 , con función de pérdida Dice Loss.
2. **Modelo 2:** Imágenes de entrada de tamaño 256×256 , con función de pérdida combinada DiceLoss + BCEWithLogits.
3. **Modelo 3:** Imágenes de entrada con tamaño original 512×512 , con función de pérdida combinada DiceLoss + BCEWithLogits.

Prueba de encoders con modelo U-Net

Posteriormente, con los hiperparámetros fijos, se entrenaron nuevamente tres modelos U-Net con diferentes encoders (ResNet34, ResNet50 y EfficientNet-B4) con pesos preentrenados en ImageNet, manteniendo como salida un único canal para segmentación binaria (vaso vs. fondo). La función de pérdida implementada combinó Dice Loss y Binary Cross Entropy (BCE), con el fin de equilibrar el aprendizaje entre la superposición de áreas y la clasificación píxel a píxel. El optimizador seleccionado fue Adam, con una tasa de aprendizaje inicial de $1e-4$.

3.2.5 Mejor modelo para segmentación semántica

Como se pudo apreciar, para el caso de la segmentación semántica se realizó una tarea progresiva de entrenamiento de diferentes modelos, variando sus configuraciones y evaluando su desempeño (Las respuestas se detallan en la sección de Resultados). Lo anterior, permitió ir obteniendo un aumento en las métricas de desempeño con cada configuración que se iba implementando y reentrenando.

En esta sección se presenta finalmente el modelo que permitió obtener el mejor desempeño a partir de las pruebas realizadas. La configuración del modelo es la siguiente:

- **Arquitectura:** Modelo base U-Net con encoder EfficientNet-B4, con pesos preentrenados con el dataset ImageNet.
- **Imagen de entrada:** Tamaño original de imágenes 512x512 píxeles.
- **Función de pérdida:** Función combinada Dice Loss + Binary Cross Entropy (BCE).
- **Optimizador:** Adam.
- **Learning rate:** tasa de aprendizaje inicial de $1e-4$.

Se aplicó Data Augmentation, bajo la siguiente configuración:

- **HorizontalFlip:** Voltea horizontalmente (de izquierda a derecha) la imagen con un 50% de probabilidad.
- **RandomRotate90:** Rota la imagen en 90° , 180° o 270° con 50% de probabilidad.
- **ShiftScaleRotate:** Rota aleatoriamente entre -15° y $+15^\circ$, Escala aleatoriamente entre 90% y 110%, y Desplaza la imagen hasta $\pm 5\%$ del tamaño total (arriba, abajo, izquierda, derecha), con una probabilidad del 50%.
- **RandomBrightnessContrast:** Aplica cambios aleatorios de brillo y contraste con una probabilidad del 30%. Este es el único parámetro que sólo aplica para las imágenes y no a la máscara.

3.3 Modelo YOLOv8 para detección de obstrucciones cardíacas

Finalizada la etapa de segmentación de vasos coronarios, se plantean los modelos para la detección de obstrucciones cardíacas, correspondiente a la segunda etapa en la implementación de modelos de aprendizaje automático. Se abordó el problema de localización de obstrucciones coronarias mediante un enfoque de detección de objetos, utilizando la arquitectura YOLOv8 y el conjunto de datos STENOSIS de la base ARCADE.

El conjunto de datos original proporcionaba anotaciones en formato *COCO (JSON)*, donde cada obstrucción estaba definida por un identificador de imagen, los puntos del área segmentada de la obstrucción, los parámetros de la caja delimitadora y la categoría del objeto, en este caso *stenosis*. Se convirtieron las anotaciones a formato *TXT*, con las coordenadas normalizadas entre 0 y 1 para las características de las cajas delimitadoras y estableciendo la clase para la detección binaria (obstrucción presente). Se realizó también la creación de un archivo de configuración *YAML* que inicializa el entrenamiento de los modelos, proporcionando la información de la organización de la base de datos y las etiquetas.

3.3.1 Evaluación de variantes YOLOv8

Se implementaron algunas configuraciones del modelo YOLO con el propósito de analizar el impacto del tamaño arquitectónico sobre el desempeño del modelo y su eficiencia

computacional en la detección de obstrucciones coronarias. Los diferentes modelos varían el número de parámetros de su arquitectura, nano con ~3M, small con ~11M, medium con ~25M y large con ~43M parámetros. El entrenamiento de los modelos se realizó con 100 épocas, batch size de 8, parada temprada con 20 épocas, implementando Data Augmentation y Transfer Learning con la librería Ultralytics de Python.

Con el fin de comprobar que el proceso de entrenamiento, validación y predicción con YOLO estaba configurado correctamente, se realizó una prueba con un dataset de referencia más sencillo: el conjunto Rock, Paper, Scissors, disponible en la plataforma Roboflow (<https://universe.roboflow.com/roboflow-58fyf/rock-paper-scissors-sxsw>). Este dataset cuenta con imágenes claras y anotaciones bien estructuradas para tres clases (piedra, papel y tijera), lo que lo convierte en un recurso idóneo para validar la correcta implementación del modelo.

Después de determinar que el modelo funciona con el dataset de prueba de Roboflow, se realiza el entrenamiento con la base de datos STENOSIS. El desempeño de cada configuración fue posteriormente evaluado utilizando métricas estándar propias de la detección de objetos, tales como precisión, recall y mean Average Precision (mAP), calculadas a diferentes umbrales de solapamiento.

3.4 Desarrollo de interfaz gráfica

Con el fin de facilitar la interacción con el sistema propuesto y permitir la visualización integrada de los resultados obtenidos por los modelos de segmentación y detección, se desarrolló una interfaz gráfica de usuario (GUI) como componente final del sistema de apoyo al diagnóstico. Esta interfaz constituye la capa de interacción entre los modelos de aprendizaje profundo implementados y el usuario final, permitiendo una interpretación visual clara de los resultados generados a partir de imágenes angiográficas.

La interfaz fue diseñada como una herramienta de apoyo experimental, permitiendo:

- la carga de imágenes angiográficas.
- la ejecución de los modelos entrenados.
- la visualización de las salidas generadas, incluyendo máscaras de segmentación y detecciones de obstrucciones coronarias.

Para el desarrollo de la interfaz gráfica se emplearon cuatro archivos principales que estructuran el funcionamiento del sistema:

- **main_app.py**: contiene el código principal de la aplicación y la construcción de la interfaz gráfica con Gradio.
- **unet_model.py**: define la arquitectura del modelo U-Net utilizado para la segmentación

de vasos sanguíneos.

- **yolov8_nano.pt:** corresponde al modelo preentrenado de detección basado en YOLOv8, encargado de identificar posibles obstrucciones cardíacas.
- **requirements.txt:** incluye las versiones específicas de las librerías y dependencias necesarias para garantizar la compatibilidad y correcta ejecución del sistema.

Es importante resaltar que la interfaz no interviene en los procesos de entrenamiento, ajuste de hiperparámetros ni optimización de los modelos, sino que integra los modelos finales como mecanismos de inferencia y presentación de resultados.

Para su implementación se utilizó la biblioteca Gradio, la cual permite el desarrollo de interfaces web ligeras y compatibles con modelos de inteligencia artificial en Python. La estructura del sistema integra el modelo U-Net para la segmentación semántica de vasos sanguíneos y el modelo YOLOv8-nano para la detección de obstrucciones, ambos previamente entrenados y validados en etapas anteriores de la metodología.

El flujo de funcionamiento de la interfaz inicia con la carga de una imagen angiográfica por parte del usuario. Posteriormente, el sistema ejecuta de manera secuencial el modelo de segmentación U-Net, generando máscaras que resaltan las estructuras vasculares, y el modelo de detección YOLOv8, encargado de identificar regiones con posibles obstrucciones mediante cuadros delimitadores. Los resultados son presentados de forma simultánea, incluyendo la imagen original, la segmentación obtenida, las detecciones realizadas y un resumen interpretativo de los hallazgos.

De esta manera, la interfaz gráfica permite validar de forma visual la coherencia y funcionalidad del sistema completo, consolidando la integración de los modelos de segmentación y detección en una plataforma accesible, intuitiva y alineada con el objetivo general del proyecto.

4. ANÁLISIS Y RESULTADOS

En esta sección se presentan los resultados obtenidos a lo largo del proyecto, organizados en cuatro frentes complementarios. Primero, se reporta el impacto del preprocesamiento basado en técnicas clásicas de visión por computadora como un punto de partida para el proceso de segmentación. Segundo, se exponen los hallazgos de la segmentación semántica del árbol coronario, con evaluación cuantitativa mediante Dice e IoU y análisis cualitativo de casos representativos, mostrando el proceso que llevó a obtener el mejor resultado, paso a paso aumentando las métricas de desempeño. Tercero, se describen los resultados de la detección de obstrucciones con YOLOv8, destacando métricas de localización y la coherencia con la anatomía vascular. Finalmente, se presentan los logros de la interfaz gráfica, que integra los modelos en un flujo de uso práctico, mostrando visualizaciones y los indicadores clave para el soporte al diagnóstico.

4.1 Datos y preprocesamiento de imágenes

Se logró desarrollar un algoritmo de preprocesamiento de imágenes digitales que permite la extracción de características relevantes en estudios de angiografía coronaria. Este resultado fue posible gracias a la implementación de un conjunto de procedimientos que mejoraron significativamente la calidad visual de las imágenes, redujeron el ruido y facilitaron la segmentación de las regiones de interés. El algoritmo, alimentado por datos provenientes de una base especializada, optimizó las condiciones para el análisis automatizado mediante técnicas de visión por computadora, sentando las bases para su posible integración y comparación con modelos de diagnóstico asistido por computadora.

4.1.1 Recolección y organización de datos

Como resultado inicial del procedimiento de preprocesamiento, se generó una máscara por cada imagen del dataset, preservando la correspondencia directa con el nombre del archivo original mediante el sufijo “_mask.png”. Este proceso se aplicó de forma sistemática a todas las particiones del conjunto de datos, garantizando uniformidad en el tratamiento de las imágenes. Las imágenes originales, todas con dimensiones de 512x512 píxeles, fueron utilizadas como base para la generación de estas máscaras, las cuales permiten identificar con precisión las regiones de interés anotadas por expertos.

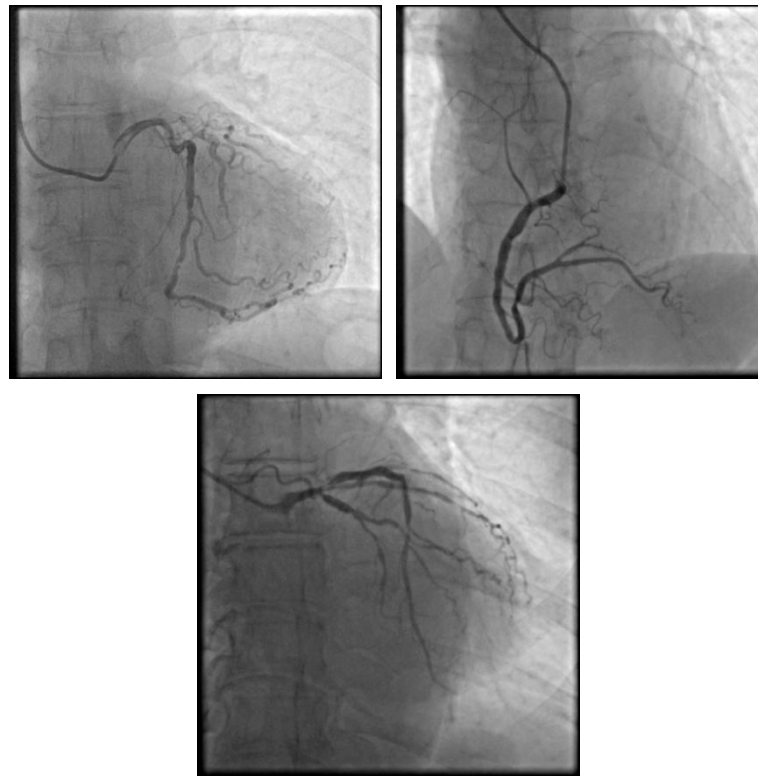


Figura 9. Imágenes de ejemplo originales del dataset.
Fuente: Dataset Arcade.

En la Figura 9 se presentan ejemplos representativos de las imágenes originales contenidas en la base de datos, mientras que la Figura 10 muestra una instancia del resultado obtenido tras aplicar el algoritmo de conversión de máscaras con base en las anotaciones dadas por los especialistas, ya contenidas en el dataset.

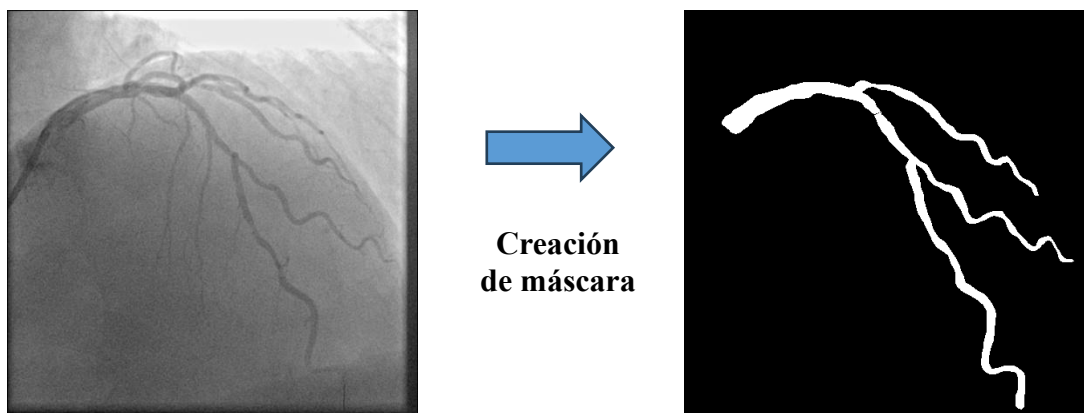


Figura 10. Creación de máscaras a partir de archivo de anotaciones.
Fuente: Elaboración propia con el dataset Arcade en Google Colab.

Estas máscaras son el insumo para los posteriores algoritmos de aprendizaje automático, siendo la entrada supervisada para el proceso de entrenamiento y los puntos de comparación para los casos de validación y prueba.

4.1.2 Preprocesamiento de imágenes con técnicas clásicas de visión artificial

El preprocesamiento basado en técnicas clásicas de visión por computadora permitió obtener una imagen final en la que se destacan los vasos sanguíneos más predominantes, que constituyen el foco principal del presente estudio. Entre las etapas aplicadas, el uso del Filtro de Frangi resultó fundamental para la detección de las estructuras vasculares. Las técnicas posteriores contribuyeron a mejorar el contraste de los vasos y eliminar elementos no relevantes para el análisis. En la Figura 11 se ilustra un ejemplo representativo de las principales etapas, evidenciando cómo el procesamiento progresivo conduce a una imagen depurada y centrada en las regiones de interés.

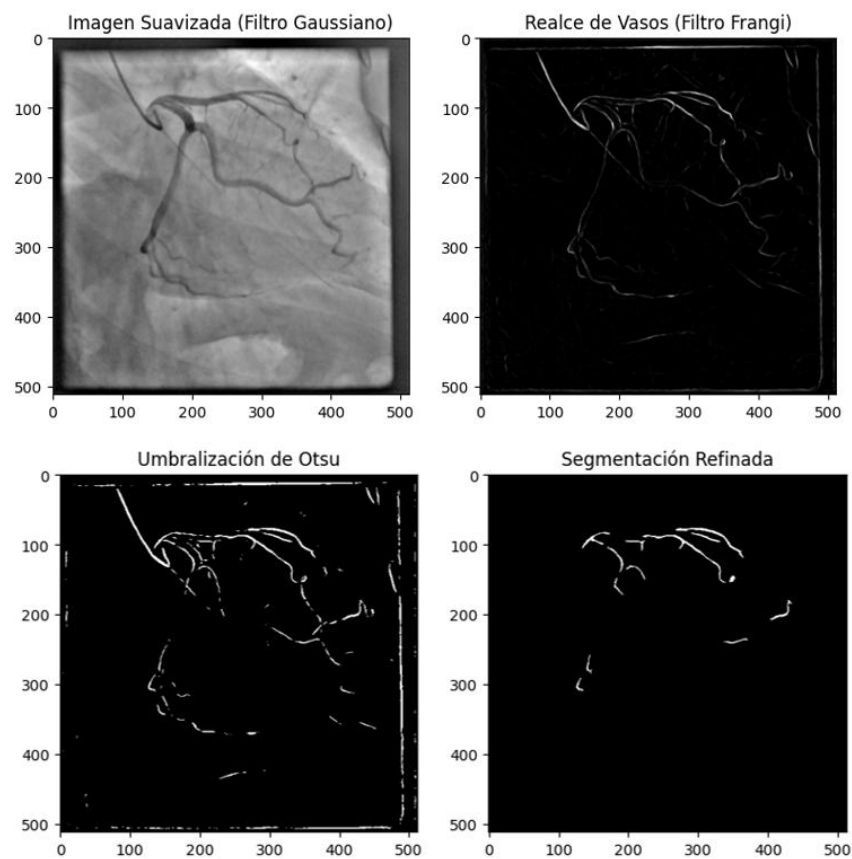


Figura 11. Etapas de preprocesamiento aplicadas a una imagen.
Fuente: Elaboración propia con el dataset Arcade en Google Colab.

Se obtuvo una representación morfológica de las estructuras vasculares presentes en la imagen procesada, lo cual permite caracterizar los patrones anatómicos relevantes para el análisis. Este resultado fue alcanzado mediante la aplicación de técnicas específicas de procesamiento que facilitaron la extracción de características morfológicas como forma, tamaño y orientación de los vasos sanguíneos. En la Figura 12 se muestra un ejemplo representativo de este procedimiento, evidenciando cómo la imagen original evoluciona hasta resaltar las propiedades estructurales de interés.

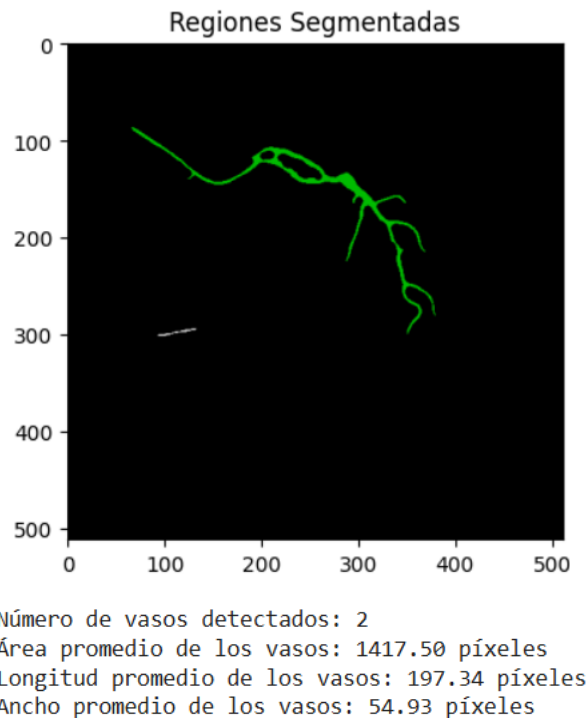


Figura 12. Extracción de características morfológicas de las imágenes preprocesadas.
 Fuente: Elaboración propia con el dataset Arcade en Google Colab.

En la anterior figura se observa que se detecta toda la ramificación como un solo vaso coronario y, por otro lado, se detecta otro vaso que corresponde a la presencia de una partícula de ruido en la imagen. Con esto, se tiene que mediante las técnicas clásicas de procesamiento digital de imágenes resulta difícil llegar a un resultado satisfactorio que se pueda obtener de manera general para todo el banco de imágenes.

4.1.3 Evaluación de preprocesamiento

Se aplicó el algoritmo de preprocesamiento basado en técnicas de visión artificial al conjunto de imágenes de validación de la base de datos “*syntax*” con 200 angiografías. Los resultados obtenidos fueron evaluados mediante las métricas de desempeño propuestas, permitiendo

verificar la efectividad de las técnicas aplicadas. Como resultado, se obtuvo una superposición entre la máscara real de las anotaciones de los expertos y la predicción dada por el algoritmo. Esta superposición facilita la validación visual del desempeño del algoritmo, asegurando que los vasos sanguíneos predominantes fueron parcialmente detectados de manera correcta. La Figura 13 presenta un ejemplo representativo de esta superposición, ilustrando la calidad del resultado alcanzado.

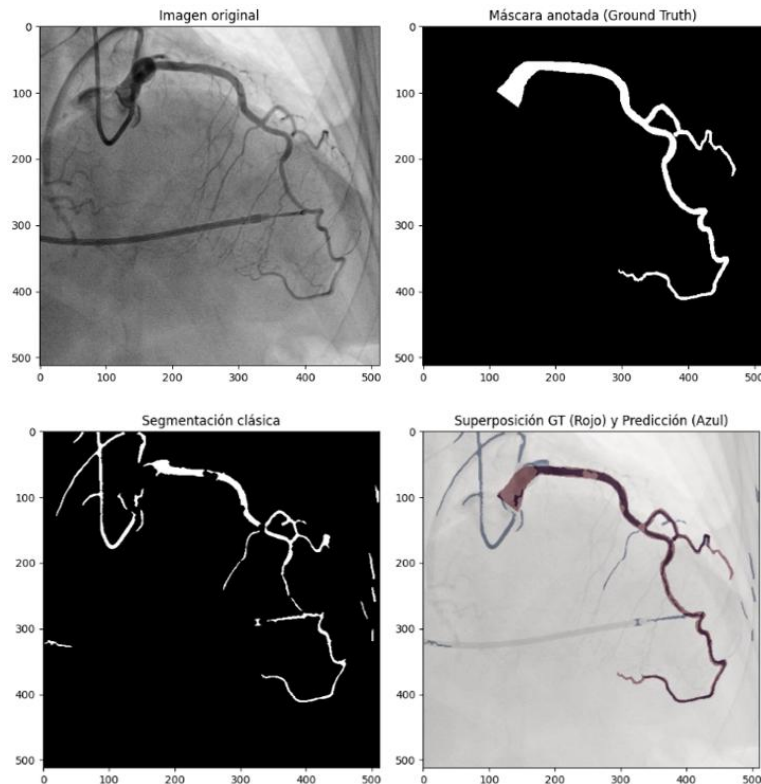


Figura 13. Visualización de superposición en evaluación de preprocesamiento.

Fuente: Elaboración propia con el dataset Arcade en Google Colab.

A partir de la superposición de la máscara real y el resultado del preprocesamiento, se aplicaron las combinaciones en las funciones de filtro gaussiano, tamaño del disco de cierre y el tamaño mínimo de objetos, dados por la grilla definida en la sección de *Metodología*. En la Tabla 4 se presentan las tres combinaciones de parámetros que obtuvieron los mejores resultados en las métricas de evaluación aplicadas al conjunto de validación. Estas combinaciones corresponden a distintas configuraciones del tamaño del kernel del filtro Gaussiano, manteniendo el elemento estructurante utilizado en la operación morfológica y el umbral mínimo aplicado de manera constante.

Tabla 4. Resultados de métricas de evaluación de preprocesamiento.

Combinación	F1 Score	IoU	Precisión	Recall
Kernel: (9, 9), Disco: 4, Min: 50	0.1083	0.0642	0.1461	0.0925
Kernel: (7, 7), Disco: 4, Min: 50	0.1037	0.0609	0.1402	0.0888
Kernel: (5, 5), Disco: 4, Min: 50	0.1033	0.0604	0.1348	0.0944

Fuente: Elaboración propia.

La mejor combinación, que utiliza un filtro Gaussiano más grande (9x9) y una operación morfológica más fuerte (disco(4)), logró el mejor desempeño en términos de superposición (F1 Score). El suavizado más intenso ayudó a reducir el ruido de fondo, y el elemento morfológico más amplio permitió conectar mejor las estructuras vasculares fragmentadas. Sin embargo, la precisión y el recall siguen siendo bajos, reflejando la dificultad del método clásico para captar de manera completa y precisa las estructuras vasculares finas.

A pesar de lo anterior, los valores de las métricas no varían considerablemente cuando se realizan cambios de configuración, por lo que ninguna combinación alcanza métricas altas, lo que refuerza la conclusión de que el método clásico de segmentación no es suficiente para un análisis diagnóstico de calidad, y justifica avanzar hacia modelos de segmentación automática basados en aprendizaje profundo.

4.2 Segmentación semántica mediante aprendizaje profundo

Se logró implementar y evaluar progresivamente distintas arquitecturas de aprendizaje profundo para segmentar y detectar vasos sanguíneos en imágenes médicas. Inicialmente se probó un modelo FCN utilizando la arquitectura ResNet-18. Posteriormente, SegNet fue utilizado como modelo base, mejorado mediante técnicas de aumento de datos y funciones de pérdida específicas. U-Net, con encoders preentrenados y ajustes arquitectónicos, mostró mejoras en el desempeño gracias al enfoque de transfer learning. Para la tarea de detección y localización de obstrucciones coronarias, se adaptó la familia de modelos YOLO, validando también su funcionamiento con un conjunto de prueba externo. Esta evaluación permitió identificar fortalezas y limitaciones de cada enfoque, orientando el desarrollo hacia soluciones más precisas y robustas.

4.2.1 Resultados del modelo Fully Convolutional Network (FCN)

Se logró implementar primeramente un modelo de segmentación semántica basado en la arquitectura FCN-8s, adaptado sobre ResNet-18. A continuación, se presenta el resultado gráfico aplicado a una de las imágenes del dataset.

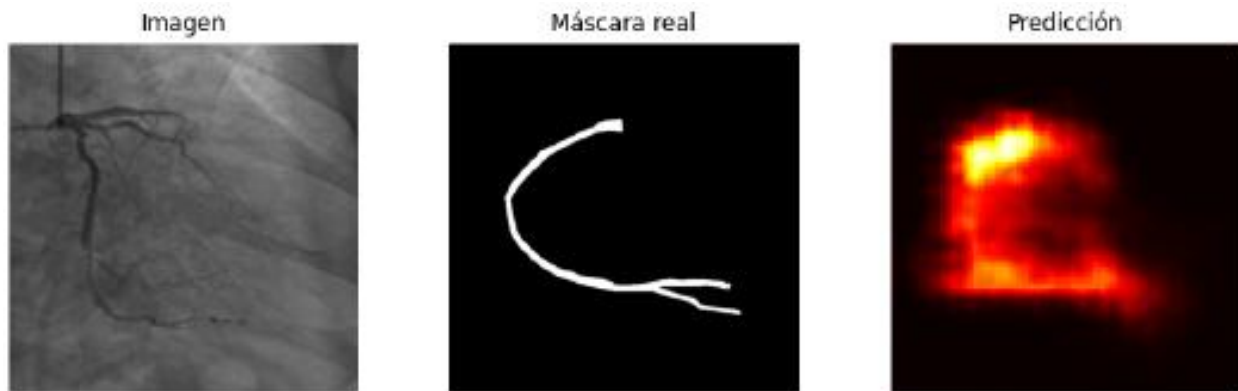


Figura 14. Mapa espacial de etiquetas obtenido con la red FCN.
Fuente: Elaboración propia con el dataset Arcade en Google Colab.

La implementación de la FCN-8s con encoder ResNet-18 produjo salidas en forma de mapas de probabilidad por píxel. En las pruebas cualitativas, la predicción se manifestó como un mapa de activación grueso (tipo heatmap), con bloques y bordes poco definidos respecto a la máscara real. Este comportamiento es consistente con la baja resolución efectiva de la representación (downsampling acumulado $\approx \times 32$) y la ausencia de conexiones de salto de alta resolución hacia el decoder, lo que limita la recuperación de detalle fino en vasos delgados y bifurcaciones.

A nivel visual, la red localiza la región vascular principal, pero presenta:

- Bordes suavizados y discontinuidades en ramas finas.
- Desalineaciones leves al reescalar, evidentes cuando el tamaño de entrada no es múltiplo de 32.
- Predicciones con intensidades continuas (probabilidades), que al umbralizar tienden a perder microramas o a engrosar segmentos.

En términos de entrenamiento, se observó descenso de la pérdida pero las métricas operativas resultaron poco informativas debido al desbalance vaso/fondo. Estos hallazgos justificaron migrar a arquitecturas con decoder más expresivo (SegNet) y, posteriormente, a U-Net con skip connections, que efectivamente mejoraron el delineado y la consistencia espacial de la segmentación.

4.2.2 Resultados del modelo YOLOv8 para vasos coronarios

Se implementó también como alternativa un modelo YOLOv8 enfocado en la detección de los vasos coronarios como una alternativa a la segmentación. A continuación, se presentan los resultados obtenidos durante el entrenamiento del modelo YOLOv8 para esta tarea.

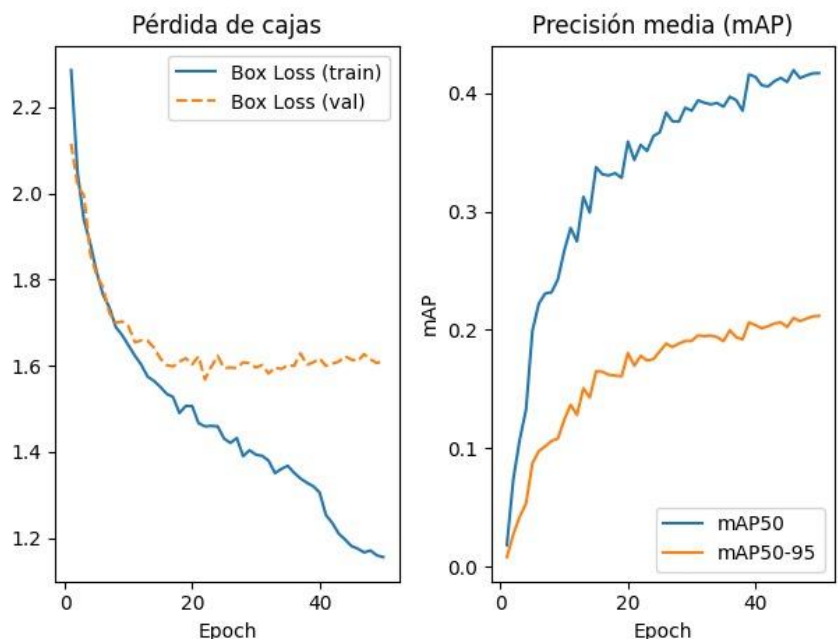


Figura 15. Comportamiento de las métricas de evaluación durante el entrenamiento del modelo YOLOv8.

Fuente: Elaboración propia con el dataset Arcade en Google Colab.

Se observa que la pérdida de cajas en entrenamiento disminuyó de forma sostenida, mientras que la pérdida en validación se estabilizó con ligera divergencia. En paralelo, la mAP creció rápido al inicio y luego se aplanó. Este patrón sugiere un sobreajuste moderado, el modelo sigue ajustándose a patrones del train sin trasladar equivalente mejora a val.

A pesar de que el mAP a un 50% de coincidencia (mAP@0.5) fue de 0.42, lo que muestra que el modelo detecta de manera aceptable los vasos cuando se acepta que la predicción y la realidad coincidan al menos en un 50%, el detector no permite una discriminación exacta de los vasos coronarios, que es lo esperado en tareas de segmentación, además, en imágenes con ramas muy delgadas o contraste bajo, se observaron cajas parciales o fragmentadas.

Se muestra a continuación, una imagen de ejemplo del uso del modelo YOLO para detección de vasos coronarios.

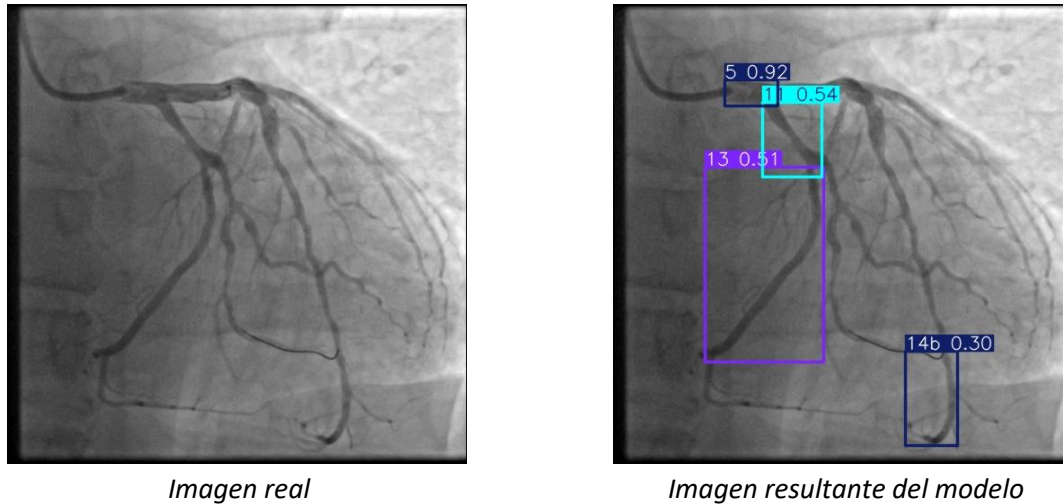


Figura 16. Segmentación de vasos sanguíneos con el modelo YOLOv8 entrenado.

Fuente: Elaboración propia con el dataset Arcade en Google Colab.

A diferencia de otros modelos como SegNet y U-Net, que hacen segmentación detallada píxel por píxel, YOLOv8 solo dibuja cajas alrededor de las obstrucciones con etiquetas y niveles de confianza, su verdadero potencial destacado por la rapidez y precisión, por lo que es ideal para la tarea posterior de ubicación de obstrucciones cardíacas como más adelante se podrá observar, pero no presenta los mejores resultados con tareas de segmentación.

4.2.3 Resultados del modelo SegNet

Resultados modelo SegNet implementado desde cero

Se logró entrenar exitosamente un modelo SegNet implementado desde cero en Keras, orientado a la segmentación binaria de vasos sanguíneos en imágenes médicas. El modelo fue entrenado con 1000 imágenes y validado con 200, utilizando la función de pérdida `binary_crossentropy` y el optimizador Adam con una tasa de aprendizaje de $1e-4$. Durante las 50 épocas de entrenamiento, se aplicó una estrategia de early stopping que permitió detener el proceso de forma eficiente al detectar estancamiento en la mejora de la pérdida de validación.

A partir de este momento, se comienzan a mostrar los resultados al ir aplicando los diferentes métodos que permitieron obtener un crecimiento progresivo en las métricas de evaluación de desempeño. Al aplicar cada método se realizaba un reentrenamiento del modelo y su correspondiente validación. Con el fin de hacer una comparación visual coherente, se toma de ejemplo una misma imagen original con su correspondiente máscara real basada en anotaciones de especialistas, y se va mostrando de manera secuencial la mejora en la

segmentación de vasos coronarios. Estas imágenes y el resultado del modelo SegNet entrenado desde cero se presenta en la siguiente figura.

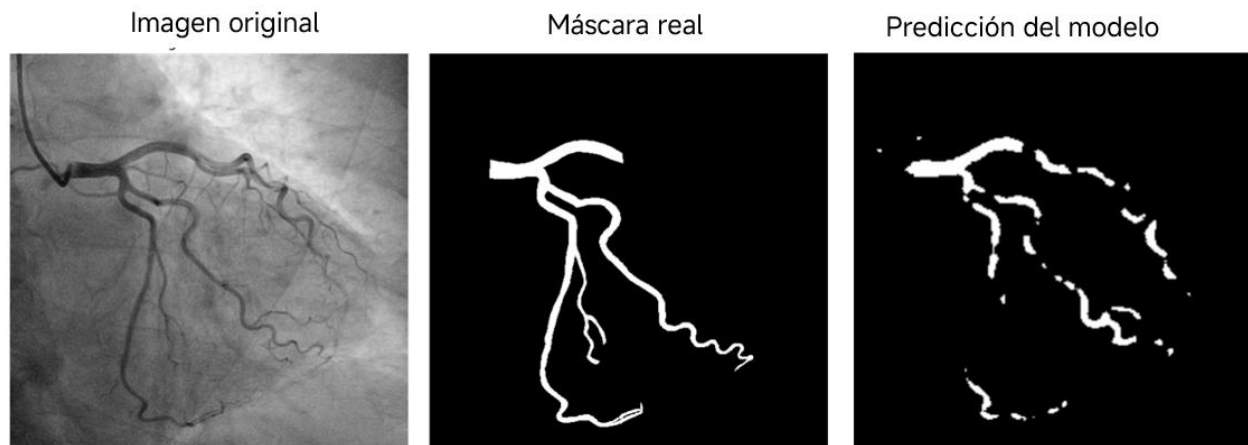


Figura 17. Segmentación de vasos sanguíneos con el modelo SegNet entrenado desde cero.

Fuente: Elaboración propia con el dataset Arcade en Google Colab.

Gracias a esta configuración, el modelo alcanzó una estabilidad en el aprendizaje sin incurrir en sobreajuste, mostrando una capacidad adecuada para segmentar regiones de interés en nuevas imágenes. Este resultado demuestra la viabilidad de SegNet como arquitectura base en tareas de segmentación médica, especialmente cuando se optimiza con técnicas de regularización y control del entrenamiento.

El modelo SegNet entrenado desde cero logró una segmentación efectiva de vasos sanguíneos, alcanzando métricas iniciales de desempeño con un Dice Coefficient de 0.6858, IoU de 0.5322, precisión de 0.7286 y recall de 0.6680. Estos valores reflejan una superposición razonable entre las regiones segmentadas y las reales, validando la capacidad del modelo para identificar estructuras vasculares relevantes.

Resultados de modelo SegNet + Data Augmentation

Con el fin de mejorar el resultado y tras aplicar las técnicas de Data Augmentation mencionadas en el apartado de *Metodología*, se entrenó nuevamente el modelo SegNet y se realizó su correspondiente validación. A continuación, se presenta el resultado de segmentación del modelo, manteniendo la misma imagen original expuesta en la *Figura 17*, para que la respuesta sea comparable.

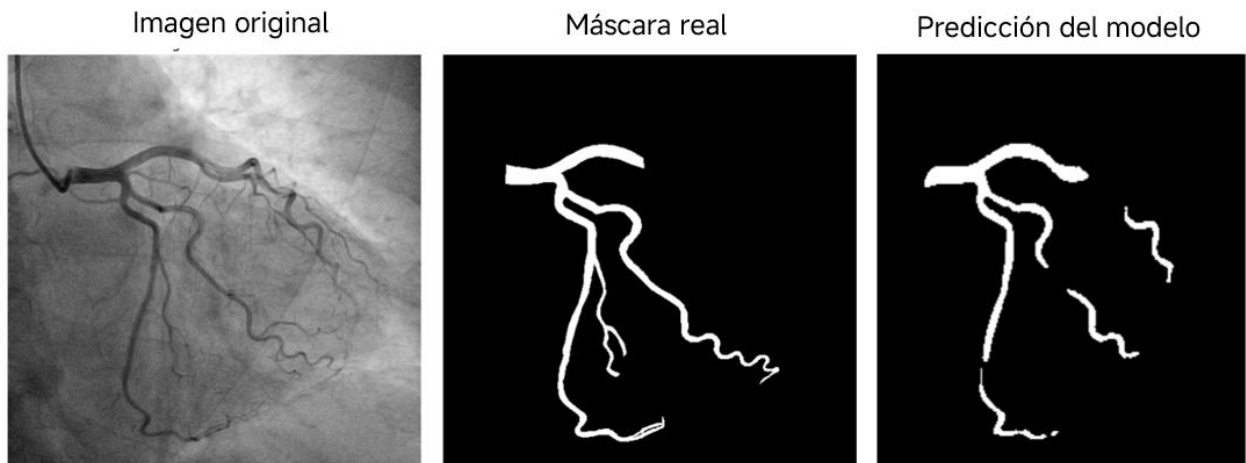


Figura 18. Segmentación de vasos sanguíneos con el modelo SegNet + Data Augmentation.
Fuente: Elaboración propia con el dataset Arcade en Google Colab.

El desempeño del modelo mejoró significativamente, alcanzando un Dice Coefficient de 0.7309, IoU de 0.5865, precisión de 0.7566 y recall de 0.7265. Esta mejora evidencia que la variabilidad introducida en el conjunto de entrenamiento permitió al modelo generalizar mejor y adaptarse a distintas configuraciones anatómicas.

Con respecto al resultado presentado por el modelo SegNet entrenado desde cero, la predicción al incorporar Data Augmentation muestra una mayor continuidad en los vasos sanguíneos segmentados y la eliminación de pequeños objetos que se generan a manera de ruido en la respuesta.

Resultado del modelo SegNet + Data Augmentation + Transfer Learning

La estrategia implementada utilizando Transfer Learning, haciendo uso de un encoder VGG16 preentrenado, empleando también Fine Tuning, no superó el rendimiento alcanzado con el modelo entrenado desde cero y potenciado con Data Augmentation. Este resultado sugiere que, para esta tarea específica, el enfoque clásico con entrenamiento directo y enriquecimiento de datos fue más eficaz que la transferencia de conocimiento desde dominios generales.

Manteniendo la misma imagen en objeto de estudio, se muestra en seguida el resultado visual de la segmentación utilizando Transfer Learning con el modelo SegNet.



Figura 19. Segmentación de vasos sanguíneos con el modelo SegNet + Data Augmentation + Transfer Learning.

Fuente: Elaboración propia con el dataset Arcade en Google Colab.

Se puede apreciar que, visualmente la segmentación al utilizar la estrategia de Transfer Learning no presente una mejoría, aún si se compara con el modelo SegNet entrenado desde cero. Se puede observar una falta de detalle en la segmentación, la figura en forma tiende a parecerse a la máscara real, pero con trazos gruesos que aumentan el número de falsos positivos.

En resumen, se presentan los resultados obtenidos en las configuraciones mencionadas utilizando el modelo SegNet, mediante la siguiente tabla.

Tabla 5. Evaluación de desempeño de los modelos U-Net con ResNet-34.

Métrica	SegNet desde cero	SegNet + Data Augmentation	SegNet + Data Augmentation + Transfer Learn
Dice F1	0.6858	0.7309	0.6815
IoU	0.5322	0.5865	0.5259
Precisión	0.7286	0.7566	0.7086
Recall	0.6680	0.7265	0.6703

Fuente: Elaboración propia.

Este hallazgo orienta los siguientes pasos hacia la exploración de arquitecturas más especializadas en Transfer Learning que puedan adaptarse mejor al dominio de imágenes médicas en escala de grises, como es el caso del modelo U-Net.

4.2.4 Resultados del modelo U-Net

En este apartado se presentan los resultados obtenidos con el modelo U-Net, permitiendo mayores variaciones en las configuraciones y opciones a explorar, siguiendo el enfoque de Transfer Learning junto con Data Augmentation. Asimismo, se almacenó automáticamente el mejor modelo alcanzado durante los procesos de entrenamiento, lo que permitió sus posteriores evaluaciones.

Resultado del modelo inicial U-Net + Transfer Learning

A continuación, se presenta la respuesta obtenida con el modelo U-Net inicial, aplicado a la imagen original de ejemplo que se mantuvo durante la sección del modelo SegNet, esto con el objetivo de mantener una comparación coherente en los resultados.



Figura 20. Segmentación de vasos sanguíneos con modelo U-Net con transfer learning.

Fuente: Elaboración propia con el dataset Arcade en Google Colab.

Se puede observar que en comparación con el modelo SegNet, el presente modelo U-Net tiene una mayor captación del detalle, con una mejor preservación de los bordes y texturas. Sin embargo, se comienza a evidenciar que la segmentación resalta algunas estructuras vasculares que no fueron anotadas por los expertos, lo cual se mejoró más adelante al realizar algunas variaciones en los hiperparámetros y posteriormente, probando otros encoders.

Resultados de la búsqueda de hiperparámetros para el modelo U-Net

De acuerdo a la estrategia de búsqueda de hiperparámetros y las configuraciones de los tres modelos mencionados en el apartado de *Metodología*, se presentan los resultados obtenidos en las métricas de desempeño sobre el conjunto de validación, manteniendo el encoder ResNet-34.

Tabla 6. Evaluación de desempeño de los modelos U-Net con ResNet-34.

Métrica	Modelo 1	Modelo 2	Modelo 3
Dice F1	0.8095	0.8101	0.8238
IoU	0.6859	0.6871	0.7076
Precisión	0.8538	0.8526	0.8565
Recall	0.7799	0.7819	0.8044

Fuente: Elaboración propia.

Al entrenar el modelo U-Net con imágenes en resolución 512×512, se logró una mejora notable en el desempeño general, alcanzando un Dice Coefficient de 0.8238 y una IoU de 0.7076. Este aumento se atribuye a la mayor riqueza espacial en las imágenes, lo que permite al modelo capturar mejor los detalles finos de los vasos sanguíneos. También se observó un incremento en el recall, evidenciando una mejor capacidad de detección de las estructuras vasculares.

Resultados de la prueba de encoders con el modeloU-Net

Teniendo en cuenta la mejor configuración de hiperparámetros obtenida anteriormente y con el objetivo de mejorar el desempeño del modelo de segmentación semántica sobre el conjunto de datos SYNTAX, se llevaron a cabo distintas pruebas utilizando la arquitectura U-Net como base, pero variando el encoder preentrenado que actúa como extractor de características. Cabe resaltar que, durante cada variación, el proceso fue realizar nuevamente el entrenamiento y validación de cada modelo. Esta etapa fue clave para explorar el impacto de encoders más profundos o eficientes sobre la calidad de las predicciones, manteniendo constantes el tamaño de entrada (512×512), la función de pérdida compuesta (*Dice Loss + BCEWithLogitsLoss*), y la estrategia de entrenamiento con Data Augmentation. Los resultados obtenidos se presentan la siguiente tabla:

Tabla 7. Evaluación de desempeño de los modelos U-Net con diferentes encoders.

Encoder	Dice Coefficient (F1)	IoU	Precisión	Recall
ResNet34	0.8238	0.7076	0.8565	0.8044
ResNet50	0.8101	0.6888	0.8550	0.7839
EfficientNet-B4	0.8294	0.7149	0.8608	0.8103

Fuente: Elaboración propia.

Del análisis comparativo se concluye que el modelo basado en *EfficientNet-B4* como encoder logró el mejor desempeño general en todas las métricas evaluadas. Esto evidencia el impacto

positivo de utilizar arquitecturas más modernas y eficientes para la extracción de características, particularmente en tareas de segmentación médica. Por tanto, este modelo será tomado como base para los siguientes experimentos y ajustes dentro del presente trabajo.

Además, en comparación con los modelos encontrados en el estado del arte (SAM-VMNet, UNet++ y Multi-ADS-Net), el modelo propuesto en este trabajo para segmentación de vasos coronarios mostró resultados competitivos que refuerzan su elección.

4.2.5 Resultados del mejor modelo para segmentación semántica

De acuerdo con los resultados obtenidos durante todos los procesos mencionados anteriormente, la intención de esta sección es concluir con el mejor modelo que presentó los valores más altos en las métricas de evaluación de desempeño. Como se pudo evidenciar anteriormente, para llegar a este modelo se realizaron pasos progresivos, cambiando configuraciones, reentrando y validando nuevamente. La arquitectura del mejor modelo para segmentación semántica corresponde a U-Net con el encoder EfficienNet-B4, manteniendo las estrategias de Data Augmentation y Transfer Learning que se detallan en el apartado de *Metodología*.

A continuación, y manteniendo la misma imagen original de referencia, se presenta el resultado del modelo U-Net con el encoder EfficienNet-B4.



Figura 21. Segmentación de vasos sanguíneos con modelo U-Net y encoder EfficienNet-B4.

Fuente: Elaboración propia con el dataset Arcade en Google Colab.

Se puede observar que al utilizar el encoder más moderno EfficienNet-B4, la segmentación mantiene un gran nivel de detalle aún con los vasos sanguíneos más pequeños, siguiendo adecuadamente su estructura y distribución.

Este modelo alcanzó los mejores resultados: Dice 0.829, IoU 0.715, Precisión 0.861 y Recall 0.810, lo que indica un alto solapamiento entre las máscaras predichas y las reales y un delineado consistente del árbol vascular. La Precisión de 0.861 sugiere bajo nivel de falsos positivos, mientras que el Recall de 0.810 evidencia una cobertura sólida de la estructura, con pérdidas puntuales en ramas muy finas o de bajo contraste, propias del dominio. En conjunto, estas métricas confirman que la combinación de skip connections y Transfer Learning aporta detalle fino y robustez superior a las alternativas evaluadas, por lo que se concluye que este modelo es la mejor respuesta para la tarea de segmentación en el presente escenario.

4.3 Modelo YOLOv8 para detección de obstrucciones cardiacas

En esta etapa, se implementó el modelo *YOLOv8* para abordar el problema desde un enfoque de detección de objetos, orientado específicamente a la identificación de obstrucciones coronarias, utilizando la base de datos STENOSIS.

Gracias a la incorporación de técnicas de Data Augmentation y Transfer Learning mediante la librería Ultralytics en Python, los modelos mostraron comportamientos progresivos en su capacidad de detección. Esta estrategia permitió evaluar el impacto del tamaño arquitectónico sobre el rendimiento. Estos resultados aportan criterios valiosos para seleccionar el modelo más adecuado según los requerimientos clínicos y computacionales del sistema de diagnóstico.

Se construyó el algoritmo de procesamiento, basado inicialmente en la variante YOLOv8 nano. Antes de comenzar a utilizar el dataset ARCADE, se probó el funcionamiento del pipeline con una base de datos de prueba, que resultara más fácil de procesar y validar. Se utilizó el dataset de prueba Rock, Paper, Scissors, disponible en la plataforma Roboflow. El entrenamiento se desarrolló sin inconvenientes, confirmando que la estructura de carpetas, la preparación de los labels en formato YOLO y el archivo de configuración data.yaml estaban correctamente definidos.

Posteriormente, al evaluar el modelo sobre imágenes de prueba, se obtuvieron predicciones visualmente correctas, donde el modelo fue capaz de identificar las manos y clasificarlas en la categoría correspondiente. A continuación, se presenta un ejemplo de la clasificación del modelo entrenado.

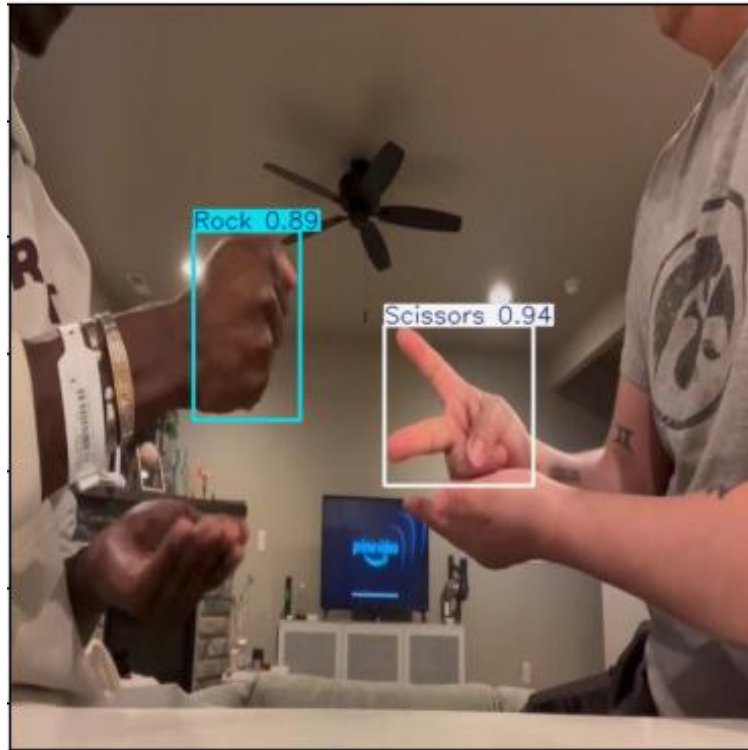


Figura 22. Prueba modelo YOLOv8 con dataset Rock, Paper, Scissors.
Fuente: Elaboración propia.

También se presentan las métricas de evaluación de desempeño dadas por el modelo entrenado con el dataset de prueba.

Tabla 8. Resultado del modelo YOLOv8 sobre el dataset Rock, Paper, Scissors.

Modelo	Precisión	Recall	mAP50	mAP50-95
YOLOv8n (Nano)	0.893	0.942	0.934	0.721

Fuente: Elaboración propia.

Como se puede observar, los valores permiten determinar que el modelo funciona correctamente y permite una buena identificación y clasificación en los diferentes grupos otorgados por la base de datos. Con esto, se procede a realizar el entrenamiento del modelo implementado, pero esta vez con el dataset ARCADE.

A continuación, se presenta una imagen de resultado del modelo YOLOv8 nano, utilizado para la detección de obstrucciones cardiacas.

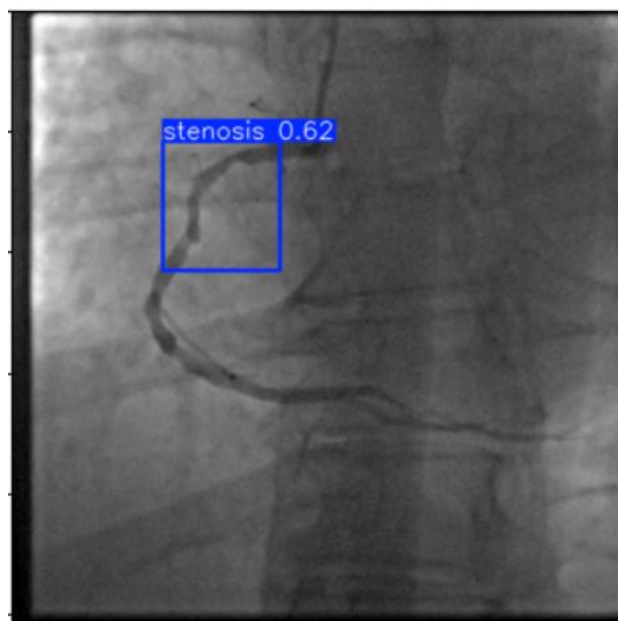


Figura 23. Detección de obstrucciones con modelo YOLOv8n.

Fuente: Elaboración propia con el dataset Arcade en Google Colab.

Con el objetivo de evaluar el rendimiento de distintas configuraciones del modelo YOLOv8 en la detección de obstrucciones coronarias, se analizaron cuatro variantes con diferente complejidad arquitectónica. La Tabla 9 resume los resultados obtenidos en términos de precisión, recall y mAP, permitiendo identificar el impacto del tamaño del modelo sobre su desempeño.

Tabla 9. Evaluación de desempeño de los modelos YOLO para detección de obstrucciones.

Modelo	Precisión	Recall	mAP50	mAP50-95
YOLOv8n (Nano)	0.368	0.416	0.321	0.12
YOLOv8s (Small)	0.322	0.35	0.283	0.112
YOLOv8m (Medium)	0.384	0.36	0.298	0.114
YOLOv8l (Large)	0.317	0.31	0.256	0.104

Fuente: Elaboración propia.

Como se observa en la Tabla 9, YOLOv8n (Nano) alcanzó el mejor desempeño global, mientras que los modelos más grandes no lograron mejorar el rendimiento, esto probablemente por

la cantidad limitada de datos que puede llevar al sobreajuste de modelos con mayor cantidad de parámetros. Sin embargo, las métricas en general presentan un desempeño aceptable atribuido a la complejidad de la base de datos STENOSIS, esto se explica principalmente por la complejidad intrínseca del conjunto, alta variabilidad de apariencia entre ejemplos, oclusiones y solapamientos, bajo contraste en varios casos, y la presencia de objetos pequeños respecto al tamaño de la imagen. En síntesis, el modelo funciona correctamente y capta patrones relevantes, pero la complejidad y heterogeneidad del dataset establecen un techo práctico al rendimiento.

Por su parte, en comparación con los resultados obtenidos mediante la revisión del estado del arte presentados en la Tabla 2, el modelo YOLOv8 presentó superioridad debido a su arquitectura más moderna que permite una detección más eficiente.

Además, el experimento sobre el dataset de prueba Rock, Paper, Scissors cumplió un rol fundamental, permitiendo confirmar que la baja precisión observada en el dataset STENOSIS no era consecuencia de errores en la implementación del modelo o en el flujo de trabajo, sino más bien de la complejidad intrínseca de ese conjunto de datos. En otras palabras, la prueba con el dataset de piedra, papel o tijera confirmó que la metodología era sólida, y que los desafíos encontrados en la base médica se deben principalmente a la naturaleza más difícil de las imágenes y las anotaciones.

Se completó la evaluación comparativa del desempeño de los modelos implementados para el análisis de imágenes médicas, abarcando arquitecturas de segmentación (SegNet y U-Net) y detección de objetos (YOLO). Esta etapa permitió contrastar los enfoques metodológicos aplicados, identificando fortalezas y limitaciones en cada caso. Los modelos de segmentación fueron evaluados mediante métricas estándar como el Coeficiente de Dice (F1), el Índice de Intersección sobre Unión (IoU), la precisión y el recall, mientras que los modelos de detección basados en YOLO fueron valorados con métricas específicas como precisión, recall y mAP a distintos umbrales (mAP@50 y mAP@50–95).

Los resultados obtenidos evidenciaron que las arquitecturas de segmentación ofrecen mayor detalle a nivel de píxel, siendo útiles para el análisis estructural de las imágenes, mientras que los modelos YOLO destacan por su rapidez y capacidad para localizar obstrucciones de forma eficiente. La validación adicional con un conjunto externo más sencillo (Rock-Paper-Scissors) confirmó la correcta implementación del flujo de entrenamiento, descartando errores técnicos y atribuyendo la baja performance inicial en el conjunto STENOSIS a la complejidad inherente de los datos.

En síntesis, el proceso de implementación permitió establecer un flujo de trabajo sólido para la experimentación con distintas arquitecturas de aprendizaje profundo aplicadas a imágenes angiográficas. La construcción inicial con SegNet brindó un punto de partida para comprender los retos de la segmentación, mientras que la incorporación de U-Net con Transfer Learning

demonstró el potencial de aprovechar encoders preentrenados para mejorar la extracción de características. Finalmente, la adaptación de los modelos YOLO al conjunto de datos de obstrucciones coronarias confirmó la viabilidad de integrar técnicas de detección de objetos dentro del proyecto.

4.4 Desarrollo de interfaz gráfica

Como parte del desarrollo de una herramienta automatizada para el análisis de imágenes cardíacas, se diseñó una interfaz gráfica interactiva que integra los modelos de segmentación y detección implementados. Esta interfaz permite al usuario cargar imágenes angiográficas, ajustar parámetros de análisis y visualizar los resultados generados por los modelos U-Net y YOLOv8 de forma intuitiva y eficiente.

El diseño de la interfaz se implementó mediante la clase `gr.Blocks`, que organiza los componentes visuales en filas, columnas y secciones. En su estructura principal se presentan tres columnas para visualizar:

1. La imagen original cargada por el usuario.
2. El resultado de segmentación obtenido con U-Net.
3. El resultado de detección producido por YOLO.

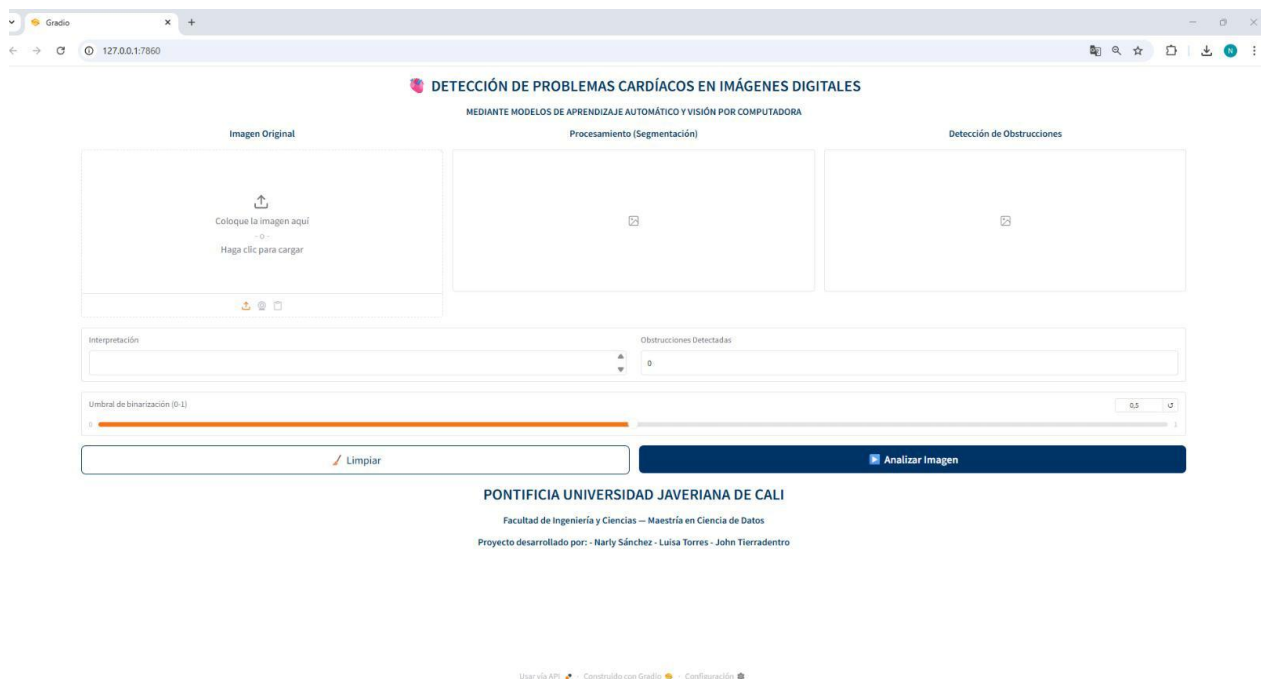


Figura 24. Diseño de interfaz gráfica.
Fuente: Elaboración propia.

Asimismo, se incorporaron elementos interactivos como un control deslizante (slider) para ajustar el umbral de binarización, lo que permite modificar la sensibilidad del modelo, y botones de “Analizar Imagen” y “Limpiar”, que facilitan la interacción con la aplicación. Para asegurar una apariencia limpia y coherente, se integró una hoja de estilos CSS personalizada, en la que se definieron los colores institucionales, tipografía, tamaño de fuente, bordes redondeados y disposición de los elementos. También se incluyó el logotipo de la Pontificia Universidad Javeriana como elemento representativo de identidad visual.

En conjunto, la interfaz desarrollada materializa la integración de los modelos de segmentación y detección en una plataforma práctica, accesible e intuitiva, que permite al usuario cargar, analizar y visualizar imágenes médicas, mostrando de manera automática la segmentación de vasos, la detección de obstrucciones y el nivel de confianza de cada hallazgo, contribuyendo así a la interpretación asistida de imágenes cardíacas.

El flujo de funcionamiento de la interfaz inicia cuando el usuario carga una imagen angiográfica en el panel principal. Posteriormente, puede ajustar el valor del gradiente o umbral según el nivel de detalle que desee observar y presionar el botón “Analizar Imagen”. A partir de esta acción, el sistema ejecuta los modelos integrados: el U-Net genera la segmentación de los vasos sanguíneos y el YOLOv8-nano realiza la detección de posibles obstrucciones cardíacas. Los resultados se presentan en pantalla, mostrando la máscara segmentada, la imagen con las detecciones resaltadas, el número total de obstrucciones encontradas y el porcentaje de confiabilidad asociado a cada detección. En caso de no hallarse anomalías, la interfaz notifica al usuario que no se detectaron obstrucciones. Este proceso resume el propósito del desarrollo: ofrecer una herramienta visual, automatizada y de fácil uso que respalde el análisis y diagnóstico de imágenes cardíacas.



PONTIFICIA UNIVERSIDAD JAVERIANA DE CALI

Facultad de Ingeniería y Ciencias – Maestría en Ciencia de Datos
Proyecto desarrollado por: Nariy Sánchez - Luisa Torres - John Tierradentro

Usar vía API - Construido con Gradio - Configuración

Figura 25. visualización del uso de los modelos sobre la interfaz gráfica.
Fuente: Elaboración propia.

5. CONCLUSIONES Y TRABAJOS FUTUROS

5.1 Conclusiones

Los resultados obtenidos en este proyecto demostraron el potencial de la visión por computadora y el aprendizaje profundo para detectar obstrucciones coronarias de manera asistida. Esta evidencia permitió comprender que, aunque los modelos desarrollados no sustituyen el criterio clínico, sí representan un apoyo valioso para reducir la subjetividad y aumentar la consistencia en la interpretación de imágenes angiográficas.

Uno de los aprendizajes más significativos fue identificar la importancia del preprocesamiento como etapa clave para mejorar la calidad de las imágenes y optimizar el desempeño de los modelos posteriores. Esta experiencia reforzó la comprensión de que la precisión del análisis automático depende no solo de arquitecturas avanzadas, sino también de la calidad y preparación adecuada de los datos.

Se implementaron distintos modelos de aprendizaje profundo, orientados tanto a la segmentación de vasos coronarios (SegNet y U-Net) como a la detección de obstrucciones (YOLOv8). Los modelos basados en segmentación demostraron una alta capacidad para identificar las estructuras coronarias principales, alcanzando métricas destacables con la arquitectura U-Net y el uso de transfer learning con EfficientNet-B4 como codificador, obteniendo valores de desempeño superiores al 0.82 en el coeficiente Dice y 0.70 en IoU. En cuanto a la detección de obstrucciones, el modelo YOLOv8 permitió validar la efectividad de los métodos basados en bounding boxes para localizar lesiones, logrando resultados prometedores pese a las limitaciones del conjunto de datos y la complejidad del problema.

El diseño de la interfaz gráfica también representó un aporte significativo, ya que permitió transformar una solución técnica en una herramienta práctica y accesible para el usuario final. Este proceso destacó la importancia de la usabilidad, la claridad visual y la experiencia del usuario en la adopción de tecnologías médicas.

En conjunto, el proyecto permitió desarrollar una visión más crítica y profesional sobre la aplicación de la inteligencia artificial en cardiología, reconociendo tanto su contribución como sus desafíos. Se puede considerar que el trabajo realizado constituye un paso importante hacia la integración de herramientas digitales de apoyo diagnóstico y abre la puerta a futuros desarrollos con bases de datos más amplias, modelos más robustos y validaciones clínicas más completas.

5.2 Trabajos futuros

A futuro, se propone validar el sistema en entornos clínicos reales mediante la colaboración con instituciones de salud, con el propósito de comparar su desempeño con diagnósticos de especialistas y ajustar sus parámetros a las condiciones prácticas. Asimismo, se plantea

evolucionar el prototipo hacia una plataforma digital web o móvil integrada con sistemas hospitalarios, que permita el análisis automático de imágenes médicas en tiempo real. Estas mejoras, junto con la posibilidad de implementar soluciones en la nube para ampliar su alcance, fortalecerían la utilidad y escalabilidad del modelo como herramienta de apoyo diagnóstico en cardiología.

También se plantea explorar nuevas arquitecturas de detección de objetos que puedan ofrecer un mejor desempeño que el alcanzado con el modelo YOLOv8. Aunque este modelo permitió validar la metodología y obtener resultados satisfactorios en la identificación de obstrucciones coronarias, las métricas de desempeño sugieren la posibilidad de mejora mediante el uso de modelos más recientes o especializados en contextos médicos. En este sentido, se propone evaluar alternativas como DETR (DEtection TRansformer), Faster R-CNN, o variantes basadas en Vision Transformers (ViT) y SAM (Segment Anything Model), las cuales podrían aprovechar mejor la información espacial y contextual de las imágenes angiográficas. La comparación sistemática de estos enfoques permitirá identificar la arquitectura más adecuada para optimizar la detección automática de patologías cardíacas en entornos clínicos reales.

6 REFERENCIAS BIBLIOGRÁFICAS

- [1] Organización Mundial de la Salud, «Enfermedades cardiovasculares,» 11 junio 2021. [En línea]. Available: [https://www.who.int/es/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)#:~:text=Se%20estima%20que%2017%2C9%20millones%20de%20personas%20fallecieron,a%20infartos%20de%20miocardio%20y%20a%20accidentes%20cerebrovasculares..](https://www.who.int/es/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)#:~:text=Se%20estima%20que%2017%2C9%20millones%20de%20personas%20fallecieron,a%20infartos%20de%20miocardio%20y%20a%20accidentes%20cerebrovasculares..)
- [2] B. Popescu, A. Stefanidis, K. Fox y B. Cosyns, «Training, competence, and quality improvement in echocardiography: the European Association of Cardiovascular Imaging Recommendations: update 2020,» *European Heart Journal - Cardiovascular Imaging*, vol. 21, nº 12, pp. 1305-1319, 2020.
- [3] World Heart Federation, «Cardiovascular disease: Fact sheet,» 2022. [En línea]. Available: <https://world-heart-federation.org/wp-content/uploads/World-Heart-Report-2023.pdf>. [Último acceso: 2 Noviembre 2024].
- [4] N. Kh, V. Sudarshan, S. Chattopadhyay y O. Faust, «Data Overloading in Medical Imaging: Emerging Issues, Challenges and Opportunities in Efficient Data Management,» *Journal of Medical Imaging and Health Informatics* 5, 2015.
- [5] M. Popov, A. Amanturdieva, N. Zhaksylyk, A. Alkanov y A. Saniyazbekov, «Dataset for Automatic Region-based Coronary Artery Disease Diagnostics Using X-Ray Angiography Images,» *Scientific Data*, 2024.
- [6] J. Rong y Y. Liu, «Advances in medical imaging techniques,» *BMC Methods*, 2024.
- [7] J. Gao, Q. Jian, B. Zhou y D. Chen, «Convolutional neural networks for computer-aided detection or diagnosis in medical image analysis: An overview,» *Mathematical Biosciences and Engineering*, 2019.
- [8] C. Martín-Isla, V. M. Campello, C. Izquierdo, Z. Raisi-Estabragh, B. Baebler, S. E. Peterson y K. Lekadir, «Image-Based Cardiac Diagnosis With Machine Learning: A Review,» *Frontiers in Cardiovascular Medicine*, vol. 7, pp. 1-19, 2020.
- [9] C. Krittanawong, K. W. Johnson, R. S. Rosenson, Z. Wang, M. Aydar, U. Baber, J. K. Min, W. H. W. Tang, J. L. Halperin y S. M. Narayan, «Deep Learning for Cardiovascular Medicine: A Practical Primer,» *Eur Heart J.*, pp. 2058-2073, 2019.
- [10] M. E. Dilsizian y E. L. Siegel, «Machine Meets Biology: A Primer on Artificial Intelligence in Cardiology and Cardiac Imaging,» *Nuclear Cardiology (V Dilsizian, Section Editor)*, vol. 20, nº 139, 2018.
- [11] C. Krittanawong, H. Zhang, Z. Wang, M. Aydar y T. Kitai, «Artificial Intelligence in Precision Cardiovascular Medicine,» *Journal of the American College of Cardiology*, vol. 69, nº 21, pp. 2657-2664, 2017.

- [12] R. M. Wehbe, A. Katsaggelos y K. J. Hammond, «Deep Learning for Cardiovascular Imaging: A Review,» *JAMMA Cardiology*, vol. 8, pp. 1089-1098, 2023.
- [13] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, 1st ed. Cambridge, MA, USA: MIT Press, 2016.
- [14] Rajpurkar, P., Hannun, A. Y., & Haghpanahi, M. (2018). Cardiologist-level arrhythmia detection with convolutional neural networks. *Nature Medicine*, 24(3), 438-444. <https://doi.org/10.1038/s41591-018-0016-7>.
- [15] G. E. Hinton, N. Srivastava, and K. Swersky, "Autoencoders, unsupervised learning, and deep learning," *CoRR*, vol. abs/1805.03709, 2018. [Online]. disponible en: <https://arxiv.org/abs/1805.03709>.
- [16] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3431–3440). <https://doi.org/10.1109/CVPR.2015.7298965>
- [17] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
- [18] Jocher, G., et al. (2023). YOLOv8: Ultralytics. GitHub Repository. <https://github.com/ultralytics/ultralytics>
- [19] S. Deepak and P. Ameer, "Cardiovascular disease prediction using deep learning techniques: A comprehensive review," *Journal of Healthcare Engineering*, vol. 2021, pp. 1-15, 2021. [Online]. Available: <https://doi.org/10.1155/2021/8894523>.
- [20] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Springer, Cham, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.
- [21] L. R. Dice, "Measures of the Amount of Ecologic Association Between Species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
- [22] P. Jaccard, "The distribution of the flora in the alpine zone," *New Phytologist*, vol. 11, no. 2, pp. 37–50, 1912.
- [23] S. Powers, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.
- [24] Huang B, Luo Y, Wei G, He S, Shao Y, Zeng X, Zhang Q. Deep learning model for coronary artery segmentation and quantitative stenosis detection in angiographic images. *Med Phys*. 2025 Jul;52(7):e17970. doi: 10.1002/mp.17970. PMID: 40665573.

- [25] Z. Wu, H. Fang, and Y. Li, "Multi-attention Dynamic Sampling Network (Multi-ADS-Net): Cross-dataset Pre-trained Model for Generalizable Vessel Segmentation in X-ray Coronary Angiography," *Medical Image Analysis*, vol. 86, 102794, 2025. [Online]. doi: <https://doi.org/10.1016/j.eswa.2025.127589>
- [26] Duan, H., Yi, S., & Ren, Y. (2024). DCA-YOLOv8: A Novel Framework Combined with AICI Loss Function for Coronary Artery Stenosis Detection. *Sensors*, 24(24), 8134. Doi: <https://doi.org/10.3390/s24248134>
- [27] Ansari M M. Evaluating Stenosis Detection with Grounding DINO, YOLO, and DINO-DETR. arXiv preprint (2025). arXiv:2503.01601. doi:10.48550/arXiv.2503.01601
- [28] Abid A., Abdalla A., Abid A., Khan D., Zou J. Gradio: Hassle-Free Sharing and Testing of ML Models in the Wild. arXiv preprint (2019). arXiv:1906.02569. doi:10.48550/arXiv.1906.02569.
- [29] Gonzalez, R. C., & Woods, R. E. (2018). *Digital Image Processing*. Pearson.
- [30] Nixon, M., & Aguado, A. (2019). *Feature Extraction & Image Processing for Computer Vision*. Academic Press.
- [31] Frangi, A. F., Niessen, W. J., Vincken, K. L., & Viergever, M. A. (1998). Multiscale vessel enhancement filtering. In *Medical Image Computing and Computer-Assisted Intervention — MICCAI'98*.
- [32] Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*.
- [33] Serra, J. (1982). *Image Analysis and Mathematical Morphology*. Academic Press.
- [34] Van der Walt, S., et al. (2014). scikit-image: image processing in Python. *PeerJ*.
- [35] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE TPAMI*.
- [36] Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv.
- [37] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE CVPR*.
- [38] Tan, M., & Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *ICML*.
- [39] Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE TKDE*.
- [40] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*.
- [41] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- [42] Prechelt, G. (1998). Early Stopping - But When? In *Neural Networks: Tricks of the Trade*. Springer.