

**COMPARACIÓN DE ALGORITMOS DE SEGMENTACIÓN HEURÍSTICA DE IMÁGENES
DE ANGIOGRAFÍA CORONARIA UTILIZANDO TÉCNICAS NO SUPERVISADAS.**

Nicolás Ramirez Millán

Tutor:

Ing. Pablo Eduardo Caicedo Rodriguez, PhD



UNIVERSIDAD DEL ROSARIO

ESCUELA COLOMBIANA DE INGENIERÍA JULIO GARAVITO

PROGRAMA DE MAESTRÍA EN INGENIERÍA BIOMÉDICA

BOGOTÁ D.C.

2026

DEDICATORIA

A mi madre, Lucelly, Por enseñarme a soñar en grande y sacrificar tanto para que yo pudiera volar. Todo lo que soy y lo que espero ser, te lo debo a ti.

AGRADECIMIENTOS

Este trabajo es la suma del esfuerzo de una madre por darle a su hijo una vida mejor; de un padre que, sin ser biológico, fue ejemplo y referente; de un hermano menor que siempre creyó en mí; de unos padrinos que nunca me abandonaron; de abuelos, tías y tíos que me acompañaron; de un mejor amigo que nunca me dejó caminar solo; de una novia que fue bastón y refugio; y de unos suegros que me acogieron.

Agradezco profundamente a Lucelly Millán por soñar y empujarme a cumplir algo que un día vimos tan lejano, por ser mi fuente de inspiración, por sus sacrificios, sus esfuerzos y, muy en especial, por su inmensa bondad.

A Alfredo Gonzalias, gracias por ser el padre que la vida me regaló y por tu ejemplo de rectitud. A José Gonzalias, mi hermano menor, por mirarme siempre con orgullo, por creer en mí cuando yo mismo dudaba y por darme la fuerza para no rendirme, pues sé que sigues mis pasos.

A ti, Manuela, por tu amor incondicional, por ser mi bastón y mi refugio en los días más difíciles, y por hacer que la carga fuera más ligera. A tus padres, por abrirme las puertas de su hogar, acogerme con tanto cariño y brindarme su apoyo.

Quiero expresar mi profunda gratitud a mi tutor, Pablo Eduardo Caicedo, por su guía, su paciencia y por compartir su conocimiento para llevar a buen puerto esta investigación. De igual manera, a mi profesor de pregrado, Julián Quintero, por sus enseñanzas y por inspirarme a seguir creciendo profesionalmente desde el primer día.

A la empresa Corazón y Aorta S.A.S., por su respaldo y por brindarme el ambiente propicio para compaginar mi labor profesional con la culminación de este gran reto.

Finalmente, gracias a Dios por la vida, la salud y la fortaleza para no desfallecer en el camino. A todos aquellos que, con una palabra de aliento, hicieron parte de este proceso: este logro también es de ustedes. Gracias.

RESUMEN

La segmentación no supervisada de la red vascular en imágenes de arteriografía coronaria constituye un desafío relevante en el análisis de imágenes médicas. Este problema se ve particularmente afectado por la presencia de ruido, artefactos de adquisición y por la complejidad morfológica de las estructuras vasculares finas. Los métodos tradicionales de segmentación presentan limitaciones para generalizar adecuadamente en escenarios con alto nivel de ruido o variabilidad anatómica, lo que puede afectar su precisión en la identificación de estructuras vasculares finas. Esta investigación presenta el desarrollo, la evaluación y comparación de dos enfoques de segmentación: W-Net y Segment Anything Model (SAM). A partir de estos modelos se propone además un esquema híbrido de aprendizaje profundo que combina las capacidades especializadas de W-Net con el carácter generalista de SAM, en donde el desafío principal consiste en la precisión y eficiencia de los algoritmos para identificar y separar objetos complejos sin requerir anotaciones manuales exhaustivas, un problema habitual en el análisis de imágenes médicas.

La metodología consistió en evaluar los dos modelos de forma independiente y, posteriormente, como un modelo híbrido. Primero, se implementó un modelo W-Net entrenado mediante un pipeline de autoetiquetado basado en el filtro heurístico de Sato + Otsu, adaptándolo como un especialista en ese dominio. Después, se evaluó SAM como un generalista capaz de identificar todas las estructuras segmentables con alta precisión de bordes, pero sin ninguna especificidad de interés.

La contribución principal es el diseño y desarrollo de un modelo híbrido que fusiona ambas arquitecturas. Este modelo implementa una estrategia en la que "W-Net Señala y SAM Dibuja", con W-Net se genera un mapa de calor que identifica las regiones de interés y con mayor probabilidad de contener los vasos, mientras que SAM proporciona un conjunto de segmentaciones de todas las estructuras de la imagen con alta precisión. Estas piezas son filtradas y emparejadas inteligentemente basándose en su superposición con el mapa guía de W-Net, combinando la inteligencia del especialista con la precisión del generalista.

Los resultados demuestran que, aunque W-Net y SAM presentan limitaciones por separado, el modelo híbrido produce segmentaciones de calidad visual superior. Un hallazgo relevante de esta investigación es el comportamiento aparentemente paradójico de las métricas de evaluación en el modelo híbrido. Al compararse contra el benchmark de pseudo-etiquetas utilizado durante el auto-etiquetado, el modelo híbrido obtiene puntuaciones cuantitativas relativamente bajas, como un coeficiente Dice de 0.46. Sin embargo, este resultado no se interpreta como un fallo del modelo, sino como una evidencia de divergencia respecto a un referente que contiene ruido y errores sistemáticos. Dado que las pseudo-etiquetas no representan un ground truth clínico real, una menor similitud puede indicar que el modelo híbrido está corrigiendo dichas imprecisiones, produciendo segmentaciones más coherentes y precisas que las generadas por los modelos individuales y por el propio benchmark de entrenamiento.

Palabras Clave: Segmentación No Supervisada; Arteriografía Coronaria; Deep Learning; Modelo Híbrido; W-Net; Segment Anything Model (SAM); Coeficiente DICE; Benchmark; Pseudo-Etiquetas.

ABSTRACT

Unsupervised segmentation of the vascular network in coronary angiography images is a key challenge in medical diagnosis, particularly hindered by image noise, artifacts, and the structural complexity of fine vascular patterns. Traditional segmentation models tend to fail in generalization or lack sufficient precision. This research presents the development and evaluation of W-Net and SAM (Segment Anything Model), as well as a novel hybrid deep learning model that fuses both architectures, combining the domain-specialized W-Net with the generalist SAM. The main challenge addressed is the accuracy and efficiency of algorithms for identifying and separating complex objects without the need for prior manual annotations, a common limitation in medical image analysis.

The methodology involved evaluating both models independently and subsequently as a hybrid model. First, a W-Net model was implemented and trained using a self-labeling pipeline based on heuristic structural enhancement and statistical thresholding techniques (Sato + Otsu), adapting it as a domain specialist. Then, SAM was evaluated as a generalist model capable of identifying all segmentable structures with high boundary accuracy, but without specificity to the target anatomical structures.

The primary contribution of this work is the design and development of a hybrid model that fuses both architectures. This model implements a strategy referred to as “W-Net points and SAM draws,” where W-Net generates a heatmap identifying regions of interest with a high likelihood of containing vascular structures, while SAM produces a set of high-precision segmentations for all structures present in the image. These candidate segments are then intelligently filtered and matched based on their overlap with the W-Net guidance map, effectively combining the domain intelligence of the specialist with the boundary precision of the generalist.

The results demonstrate that, although W-Net and SAM exhibit limitations when applied independently, the hybrid model produces segmentations of superior visual quality. A relevant

finding of this study is the apparently paradoxical behavior of the evaluation metrics in the hybrid model. When compared against the pseudo-label benchmark used during self-labeling, the hybrid model achieves relatively low quantitative scores, such as a Dice coefficient of 0.46. However, this outcome should not be interpreted as a model failure, but rather as evidence of divergence from a reference that contains noise and systematic errors. Since the pseudo-labels do not represent a true clinical ground truth, lower similarity may indicate that the hybrid model is correcting these inaccuracies, producing more coherent and precise segmentations than those generated by the individual models and by the training benchmark itself.

Keywords: Unsupervised Segmentation; Coronary Angiography; Deep Learning; Hybrid Model; W-Net; Segment Anything Model (SAM); DICE Coefficient; Benchmark; Pseudo-Labels.

TABLA DE CONTENIDO

1	INTRODUCCIÓN	12
2	OBJETIVOS	14
2.1	General	14
2.2	Específicos	14
3	MARCO TEÓRICO	15
4	METODOLOGÍA	22
4.1	Análisis Del Contexto Biomédico.	23
4.1.1	Caracterización de la modalidad de imagen y sus limitaciones	23
4.1.2	Evaluación de las barreras del etiquetado	23
4.1.3	Limitaciones técnicas identificadas frente a la precisión clínica	24
4.2	Entendimiento de los Datos	24
4.2.1	Análisis de propiedades y distribución de píxeles	25
4.2.2	Detección de atípicos y artefactos	25
4.2.3	Análisis de datos faltantes y consistencia	26
4.3	Preparación de los Datos	27
4.3.1	Limpieza y estandarización	27
4.3.2	Generación de máscaras de verdad absolutas	28
4.4	Modelado Individual	30
4.4.1	Implementación del Modelo Especialista: W-Net	31
4.4.2	Implementación del Modelo Generalista: SAM	34
4.4.3	Implementación del Modelo Híbrido	34
4.5	Validación y Explicabilidad	38
5	RESULTADOS	41
5.1	Resultados del Modelo Especialista (W-Net)	41
5.2	Resultados del Modelo SAM	46
5.3	Resultados del Modelo Híbrido	49
5.3.1	La Divergencia Con La Referencia Heurística	51
5.3.2	Resultados Cualitativos e Inspección Visual	52
5.3.3	Interpretabilidad y análisis de activación espacial	54
6	DISCUSIÓN	57

7	TRABAJOS FUTUROS	60
8	CONCLUSIONES	62
9	BIBLIOGRAFIA	64
10	ANEXOS	69
10.1	Anexo 1.....	69

LISTA DE TABLAS

Tabla 1	42
Tabla 2	43
Tabla 3	46
Tabla 4	48
Tabla 5	51

LISTA DE FIGURAS

Figura 1	17
Figura 2	18
Figura 3	26
Figura 4	27
Figura 5	30
Figura 6.	45
Figura 7	50
Figura 8	53
Figura 9	55

1 INTRODUCCIÓN

En el campo del análisis de imágenes médicas, la segmentación de estructuras anatómicas representa un desafío crítico debido a la complejidad y variabilidad de los datos visuales. Específicamente en la cardiología intervencionista y diagnóstica, la segmentación precisa de la red vascular en imágenes de arteriografía coronaria es un paso fundamental, pero por lo general, afectado por la baja calidad, el ruido y la presencia de artefactos en la imagen. El uso de algoritmos de segmentación heurísticos y no supervisados se ha vuelto clave para superar las limitaciones de los métodos tradicionales, que dependen de grandes cantidades de datos etiquetados manualmente, un recurso que es escaso y costoso de obtener en el ámbito clínico [1], [2].

En este sentido, la comparación de modelos avanzados como W-Net y el Segment Anything Model (SAM) se vuelve especialmente relevante, ya que ofrecen enfoques robustos para la segmentación sin el requerimiento de etiquetado manual previo.

W-Net, una arquitectura especialista basada en un doble codificador-decodificador, ha demostrado ser una herramienta potente. Su diseño, que implementa técnicas de optimización basadas en la pérdida de reconstrucción, le permite aprender características contextuales profundas. En esta investigación, W-Net se entrena como un experto del dominio mediante un innovador pipeline de autoetiquetado de naturaleza heurística, que combina realce tubular (Sato) con un método estadístico de umbralización global (Otsu) [3].

Por otro lado, SAM (Segment Anything Model) representa un paradigma completamente diferente: un modelo fundacional o generalista. SAM destaca por su capacidad de segmentar semánticamente cualquier objeto en una imagen en un modo de zero-shot, procesando grandes conjuntos de datos visuales y adaptándose a estructuras que no ha visto previamente. Su fortaleza radica en la precisión de sus bordes, aunque carece de la especificidad contextual para discernir, por sí solo, qué estructuras son de interés [4].

Ambos algoritmos representan avances significativos, pero sus filosofías opuestas plantean la pregunta central de esta investigación: ¿Es mejor un algoritmo especialista entrenado como

W-Net o un algoritmo generalista adaptable como SAM para la compleja tarea de segmentar vasos coronarios en angiografía?

En aplicaciones médicas, donde la calidad de la imagen no es óptima, la robustez al ruido es un factor crítico. Este estudio no solo compara directamente las fortalezas y debilidades de W-Net y SAM en este escenario, sino que va un paso más allá. Explora el desarrollo de una arquitectura híbrida que busca fusionar la inteligencia contextual de W-Net con la precisión de bordes de SAM, con el objetivo de desarrollar una solución superior que mejore la generalización y la eficiencia para el uso en entornos clínicos reales [5], [6].

2 OBJETIVOS

2.1 General

Comparar los modelos de segmentación W-Net y SAM para identificar sus fortalezas y limitaciones en el contexto del desarrollo de un enfoque híbrido que mejore la capacidad de generalización y la eficiencia en entornos clínicos reales.

2.2 Específicos

- Identificar las limitaciones y ventajas de W-Net y SAM en el contexto de la segmentación automática de imágenes médicas, sentando las bases para explorar su aplicabilidad en entornos biomédicos.
- Explorar la posible adaptación híbrida de los modelos W-Net y SAM en aplicaciones heurísticas de segmentación de imágenes médicas, evaluando su contribución al mejoramiento de la precisión diagnóstica.
- Evaluar la capacidad de explicabilidad de los modelos W-Net y SAM en la segmentación de imágenes médicas, analizando criterios de interpretabilidad y mecanismos de visualización de características relevantes.

3 MARCO TEÓRICO

La enfermedad coronaria representa una de las principales causas de muerte a nivel mundial. De acuerdo con los autores Hsia, Peck y Conte [7] esta consiste en una serie de síndromes clínicos causados por un flujo sanguíneo insuficiente al músculo cardíaco en su mayoría por deposición de ateroma cuyas complicaciones son la principal causa de muerte a nivel mundial [8]. El estándar diagnóstico para esta patología es el análisis de angiografías coronarias por rayos X. Su adquisición requiere la inyección de un medio de contraste radiopaco que permita la visualización de las arterias coronarias.

Sin embargo, el análisis e interpretación automática de este tipo de imágenes presenta desafíos a considerar en la visión por computadora como lo son el bajo contraste y ruido, resultado de la disminución de la densidad y heterogeneidad de los fotones incidentes propios de la generación de rayos X [9]. Asimismo, la superposición de estructuras, consecuencia de que la angiografía coronaria es una proyección bidimensional de estructuras anatómicas tridimensionales, en la que los vasos se superponen entre sí, incluso entre otras estructuras como tejido óseo, dificultando el procesamiento eficaz de la información. A estos problemas se suman diversos artefactos, como la presencia de catéteres, clips quirúrgicos y textos informativos quemados en la imagen, los cuales interfieren significativamente con la segmentación automática de los vasos coronarios tal como muestra en la figura 1.

La segmentación automática de imágenes es un componente fundamental para analizar imágenes médicas, se basa en fraccionar una imagen digital en múltiples segmentos o píxeles [10]. El objetivo es lograr que la representación de una imagen sea más fácil de analizar, extrayendo características de interés. Aquí se aborda el paradigma de la segmentación No supervisada y heurística en donde el desarrollo no depende de etiquetas manuales, sino que, utiliza características intrínsecas de la imagen para la agrupación de píxeles. La segmentación significa un problema de optimización complejo. Los algoritmos de segmentación trabajan bajo dos principios:

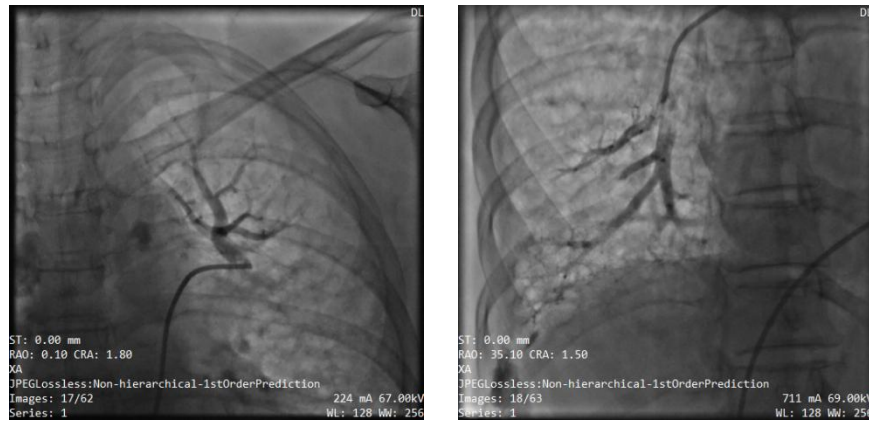
El desajuste o discontinuidad, que busca identificar cambios bruscos en la intensidad para detectar bordes, y la similitud, que agrupa los píxeles con características comunes como color o intensidad [10].

Para que un algoritmo pueda navegar entre las posibles soluciones de segmentación, se necesita una métrica cuantitativa, una función objetivo o también llamada función de aptitud, que evalúe la calidad de las segmentaciones candidatas. La elección de la función es un paso importante, ya que define el tipo de paisaje de búsqueda que el algoritmo debe recorrer. En esta investigación se han empleado algoritmos heurísticos clásicos basados en filtros matemáticos diseñados para resaltar estructuras tubulares. Entre ellos el filtro de Sato, propuesto por Sato [11], el cual se fundamenta en el análisis multiescalar del Hessiano para identificar objetos tubulares a partir de las relaciones entre sus autovalores. Este filtro demostró ser útil para diferenciar entre estructuras alargadas en imágenes ruidosas, convirtiéndose en una herramienta importante para aplicaciones iniciales de segmentación automática en angiografía. Sin embargo, investigaciones posteriores demostraron limitaciones importantes, especialmente en situaciones donde las estructuras presentan bifurcaciones, cambios bruscos de diámetro o morfologías no estrictamente tubulares. En este sentido, el filtro de Sato se integra en esta investigación como una aproximación heurística fundamental para resaltar estructuras tubulares, sirviendo como base comparativa frente a modelos más recientes y complejos como W-Net y SAM [11], [12]

Además del filtro de Sato, otro método utilizado es el método de Otsu, que fue introducido por Nobuyuki Otsu y es una técnica de umbralización automática que selecciona el valor óptimo de corte al maximizar la varianza entre clases. Esto consiste en asumir que la imagen puede dividirse en dos grupos de píxeles con distribuciones de intensidad diferenciables [13], [14], permitiendo separar regiones de interés sin depender de parámetros manuales, en donde la utilizamos para binarizar la salida probabilística de filtros como el de Sato, separando los vasos como región de interés, del fondo o región de no interés.

Figura 1

Angiografías coronarias izquierdas obtenida mediante fluoroscopia con inyección de medio de contraste.



Fuente: Elaboración propia.

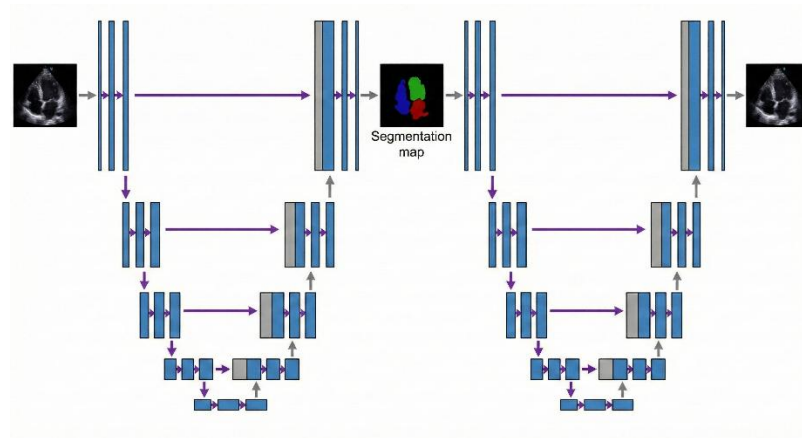
Si bien los métodos heurísticos clásicos como Sato y Otsu establecen una línea funcional, su dependencia limita su capacidad de generalización frente a la alta variabilidad anatómica y los artefactos complejos presentes en la práctica clínica. Para superar esas limitaciones, esta investigación apunta hacia paradigmas del Deep Learning, mediante el uso de Redes Neuronales Convolucionales (CNNs).

En este contexto, la arquitectura U-Net se ha consolidado como el pilar de oro en la segmentación de imágenes médicas. Su diseño de red convolucional, compuesto por una vía de contracción (codificador) para capturar el contexto y una vía de expansión (decodificador) simétrica para permitir una localización precisa, ha demostrado un rendimiento superior en tareas supervisadas. No obstante, la dependencia crítica de U-Net de grandes volúmenes de datos etiquetados manualmente, lo que es un recurso costoso, de difícil acceso y propenso a errores humanos, es lo que impulsa al desarrollo de un modelo no supervisado como W-Net [3], [15].

La arquitectura W-Net representa una evolución significativa al concatenar dos U-Nets en una estructura de codificador-decodificador doble [3].

Figura 2

Representación esquemática de la arquitectura W-Net.



Fuente: Elaboración propia.

Su innovación principal radica en la formulación de un problema de optimización en el que el modelo minimiza una pérdida de reconstrucción, asegurando que la segmentación conserve la información esencial de la imagen original. En el contexto de esta investigación, W-Net se adapta como un modelo de aprendizaje no supervisado, donde las máscaras generadas por los algoritmos heurísticos Sato y Otsu actúan como pseudoetiquetas iniciales. De esta manera, la red aprende a generalizar y refinar las estructuras tubulares presentes en la imagen, mostrando una mayor robustez frente al ruido en comparación con el uso aislado de dichos filtros.

De forma paralela al desarrollo de modelos especialistas como W-Net, el campo de la visión por computador ha experimentado una transformación significativa con la aparición de los llamados modelos fundacionales. Un ejemplo representativo es el Segment Anything Model (SAM), desarrollado por Meta AI. Este modelo fue entrenado con más de mil millones de máscaras correspondientes a aproximadamente 11 millones de imágenes y emplea una arquitectura basada en Vision Transformers, lo que le otorga una notable capacidad de generalización.

Una de sus propiedades más relevantes es la capacidad de realizar segmentación zero-shot, entendida como la habilidad de segmentar objetos sin haber sido entrenado previamente en el dominio específico de aplicación. Gracias a este enfoque, SAM puede delinear objetos con alta precisión geométrica a partir de prompts. No obstante, su naturaleza generalista plantea interrogantes sobre su capacidad para identificar de manera consistente estructuras anatómicas específicas cuando no se dispone de un contexto guiado, lo cual motiva la exploración de estrategias híbridas que integren modelos especialistas con modelos fundacionales [4], [16], [17].

Para evaluar el rendimiento de estos modelos, es necesario emplear métricas cuantitativas que no dependan exclusivamente de la precisión por píxel, la cual puede resultar engañosa en escenarios con desequilibrio de clases, como ocurre en las imágenes de angiografía coronaria, donde el número de píxeles correspondientes a vasos es significativamente menor que el número de píxeles de fondo. En este trabajo se utilizan dos métricas ampliamente aceptadas en segmentación de imágenes médicas: el Coeficiente de similitud de Dice (Dice Similarity Coefficient, DSC), que mide el grado de superposición espacial entre la segmentación predicha A y la segmentación de referencia B , donde A corresponde al conjunto de píxeles clasificados como vaso por el modelo y B al conjunto de píxeles de la máscara de referencia [5],

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \quad (1)$$

y el índice de Intersección sobre la Unión (IoU), una métrica estricta que penaliza tanto los falsos positivos como los falsos negativos, proporcionando una visión integral del rendimiento de los algoritmos propuestos [18]

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

La revisión de la literatura sugiere la existencia de un compromiso entre especificidad y generalidad en los modelos de segmentación. Por un lado, los modelos especialistas tienden a destacar en la localización semántica de estructuras específicas, aunque en algunos casos presentan limitaciones en la precisión de los bordes. Por otro lado, los modelos fundacionales poseen una alta capacidad para capturar patrones geométricos complejos y delinear contornos con alta precisión, pero pueden carecer de la información semántica necesaria para discriminar cuáles estructuras son clínicamente relevantes.

En este contexto, resulta pertinente contrastar la estrategia propuesta con otros enfoques orientados a enfrentar la escasez de datos etiquetados, como la Adaptación de Dominio No Supervisada (UDA). La generación de anotaciones manuales para tareas de segmentación constituye un proceso costoso y altamente dependiente de expertos. Como respuesta a esta limitación, los métodos UDA han sido desarrollados con el objetivo de transferir conocimiento desde dominios fuente etiquetados como datasets sintéticos o públicos hacia dominios reales sin anotaciones manuales [19].

En ese sentido, los métodos UDA entrenan modelos en dominios donde las etiquetas son más accesibles para después adaptarlos mediante técnicas como alineación de características, aprendizaje adversarial o esquemas de entrenamiento autosupervisado. Sin embargo, los beneficios de esos enfoques dependen del nivel de similitud entre dominios [20], [21]. En situaciones clínicas reales, como en angiografía coronaria, la variabilidad en los protocolos de adquisición, equipos médicos, calidad de imagen y administración de medio de contraste puede generar desplazamientos de dominio significativos, lo que puede intervenir en la estabilidad del proceso de adaptación y ocasionar problemas de transferencia negativa.

A diferencia de enfoques como el de UDA, esta investigación plantea un enfoque no supervisado basado en la extracción de señales directamente desde la estructura de las imágenes de angiografías. Con el uso de filtros heurísticos especializados en estructuras tubulares, como el filtro de Sato, combinados con técnicas de binarización adaptativa como Otsu, se generan pseudoetiquetas o máscaras de verdad iniciales que permiten entrenar el

modelo sin depender de datasets externos. Esto reduce la exposición a sesgos provenientes de dominios externos y permite adaptar el aprendizaje directamente a las características fisiológicas y físicas propias del proceso de adquisición angiográfica.

A partir de la revisión de la literatura, se evidencia una tensión metodológica entre dos enfoques dominantes en la segmentación de imágenes médicas. Por un lado, los modelos especialistas basados en aprendizaje profundo, como W-Net, tienen una elevada capacidad para incorporar conocimiento contextual del dominio, lo que favorece la localización semántica de estructuras anatómicas específicas. No obstante, su rendimiento puede verse condicionado por la calidad de las pseudoetiquetas utilizadas durante el entrenamiento no supervisado. Por otro lado, los modelos fundacionales como Segment Anything Model (SAM) tienen una notable capacidad de generalización y una alta precisión geométrica en la delimitación de contornos, aunque carecen de mecanismos para discriminar qué estructuras poseen relevancia clínica dentro de la imagen.

Esto sugiere que ambos paradigmas no deben entenderse necesariamente como alternativas excluyentes, sino como enfoques integrables. En consecuencia, a partir del análisis comparativo de sus fortalezas y limitaciones, esta investigación propone la exploración de un enfoque híbrido en el que un modelo especialista (W-Net) actúe como mecanismo de atención semántica para localizar regiones de interés vascular, mientras que un modelo fundacional (SAM) contribuya con su capacidad de delineación geométrica de bordes. Bajo esta premisa, se plantea que la integración de ambos enfoques puede mejorar la robustez frente al ruido y la precisión morfológica de la segmentación vascular en angiografía coronaria.

4 METODOLOGÍA

La presente investigación se cataloga como de tipo aplicada, desarrollada bajo un enfoque cuantitativo y un diseño experimental. Su carácter aplicado reside en la resolución de una problemática clínica tangible: la automatización de la segmentación de vasos coronarios para asistir el diagnóstico médico, trasladando conceptos teóricos de visión por computadora a un entorno de salud real. Es cuantitativa y experimental dado que se fundamenta en la manipulación controlada de variables arquitecturas de red, hiperparámetros, técnicas de preprocesamiento y la medición objetiva de su impacto en el rendimiento del modelo mediante métricas estadísticas estandarizadas, tales como el Coeficiente de Dice y la Intersección sobre la Unión (IoU).

En el desarrollo de software, existen metodologías tradicionales de tipo lineal como el modelo en cascada que estructuran el proceso en fases secuenciales, donde cada etapa debe completarse antes de avanzar a la siguiente. Si bien estos enfoques permiten una planificación estructurada, presentan limitaciones en escenarios donde los datos y los modelos requieren ajustes iterativos continuos [22].

Para garantizar el rigor científico, la reproducibilidad de los experimentos y la gestión eficiente del flujo de datos se adoptó el estándar CRISP-DM (Cross-Industry Standard Process for Data Mining). A diferencia de estas, CRISP-DM proporciona una estructura cíclica e iterativa [23], lo cual es crítico en proyectos de Aprendizaje Profundo donde la limpieza de datos y el ajuste de modelos son procesos interdependientes. La elección de este marco metodológico se fundamenta en tres pilares clave para este estudio:

- **Iteración Datos y Modelo:** Permite regresar a etapas anteriores como la preparación de datos cuando los resultados del modelado indican que el ruido o los artefactos están afectando el aprendizaje, tal como ocurrió con la implementación de los filtros Sato y el recorte de metadatos.

- **Trazabilidad:** Asegura que cada decisión técnica esté directamente alineada con el objetivo de comprender la morfología vascular.
- **Generalización:** Facilita la transición desde un modelo especialista hacia una arquitectura híbrida, validando cada fase antes de aumentar la complejidad del sistema.

Para dar respuesta a los objetivos planteados y validar la hipótesis de investigación, el desarrollo del proyecto se estructuró en seis fases, adaptadas del ciclo de vida CRISP-DM al contexto específico del análisis de imágenes angiográficas:

4.1 Análisis Del Contexto Biomédico.

En esta fase, se realizó una caracterización técnica de la angiografía coronaria por rayos X (XRA) para determinar los requisitos funcionales del sistema de visión por computadora. El análisis se centró en tres parámetros críticos que justifican el diseño de esta solución:

4.1.1 Caracterización de la modalidad de imagen y sus limitaciones

Se estableció que la angiografía, al ser una técnica de proyección 3D a 2D, presenta desafíos intrínsecos que los algoritmos de segmentación estándar no logran resolver eficazmente [24]. Se identificó que la superposición de estructuras es el obstáculo principal: las arterias coronarias frecuentemente se cruzan visualmente con la columna vertebral, las costillas y el tejido pulmonar, generando falsos bordes y oclusiones [25]. Además, la naturaleza física de la fluoroscopia introduce ruido de Poisson y un bajo rango dinámico, lo que resulta en un contraste pobre entre el vaso sanguíneo y el fondo, especialmente en vasos distales de menor calibre.

4.1.2 Evaluación de las barreras del etiquetado

Se concluyó que la obtención de etiquetas reales masivo es inviable en este contexto debido a dos factores:

- **Costo y disponibilidad de expertos [24]:** El delineado manual de arterias requiere cardiólogos intervencionistas altamente entrenados, cuyo tiempo es escaso y costoso.
- **Variabilidad Interobservador:** Incluso entre los expertos, existen diferencias sobre los límites exactos de una estenosis o un vaso difuso. Esta restricción definió un requisito técnico clave para la investigación: el modelo debe ser capaz de aprender de forma no supervisada o autosupervisada, reduciendo la dependencia de anotaciones humanas [24].

4.1.3 Limitaciones técnicas identificadas frente a la precisión clínica

Los métodos heurísticos basados en filtros de realce son excelentes para localizar estructuras tubulares, pero fallan al definir bordes precisos o tener sensibilidad a deformaciones puntuales debido al ruido [26] .

Los modelos fundacionales como SAM tienen capacidad superior para segmentar bordes y texturas, pero carecen de contexto médico, segmentando indiscriminadamente costillas, catéteres o diafragma como si fueran objetos de interés.

Basado en este análisis, se estableció que las herramientas individuales no satisfacían los objetivos clínicos. Por tanto, se definió la necesidad de diseñar una arquitectura híbrida que integre W-Net guiado por heurísticas con SAM, abordando así la superposición, el ruido y la falta de etiquetas simultáneamente.

4.2 Entendimiento de los Datos

Los datos utilizados en este estudio corresponden a una base de datos propia y privada, conformada por un total de 1606 imágenes de angiografías coronarias. Estas imágenes fueron facilitadas por la IPS Corazón y Aorta S.A.S. y provienen de estudios clínicos realizados con un angiógrafo GE modelo Optima IGS 320. Los archivos originales fueron anonimizados utilizando un software DICOM y posteriormente exportados en formato PNG con dimensiones de 512 x 512 píxeles.

Con esta base consolidada, se procedió a la revisión del conjunto de datos para garantizar que las entradas del modelo fueran consistentes. Se realizó un análisis exploratorio exhaustivo, abordando tres dimensiones críticas:

4.2.1 Análisis de propiedades y distribución de píxeles

Se examinó la distribución de las intensidades de los píxeles mediante el cálculo de histogramas globales. Se observó un comportamiento bimodal característico.

- **Pico de Fondo:** Una gran concentración de píxeles en valores bajos cercanos a 0, correspondientes al fondo oscuro y tejido blando no contrastado.
- **Regiones de Interés:** Una distribución dispersa en valores medios-altos, correspondientes a las arterias llenas de medio de contraste.
- **Saturación:** Se detectó que frecuentemente presentaba valores saturados en zonas localizadas, lo cual no correspondía a tejido biológico, sino a artefactos artificiales. Este hallazgo fue crucial para determinar que una normalización simple no sería suficiente sin una limpieza espacial previa.

4.2.2 Detección de atípicos y artefactos

Se definieron los valores atípicos en este contexto no como errores numéricos tabulares, sino como estructuras visuales que introducen ruido semántico severo. Se identificaron dos categorías de atípicos que requirieron tratamiento:

- **Atípicos de alta intensidad:** El 100% de las imágenes contenía información alfanumérica como información del estudio quemada en los píxeles de la imagen, predominantemente en la región del 15 a 20 % inferior. Estos caracteres poseen bordes nítidos y alto contraste, lo que los convierte en falsos positivos para el detector de bordes.
- **Atípicos estructurales:** Se documentó la presencia recurrente de catéteres guía, costillas y vértebras. Estos elementos presentan una morfología tubular similar a las arterias, actuando como artefactos distractores

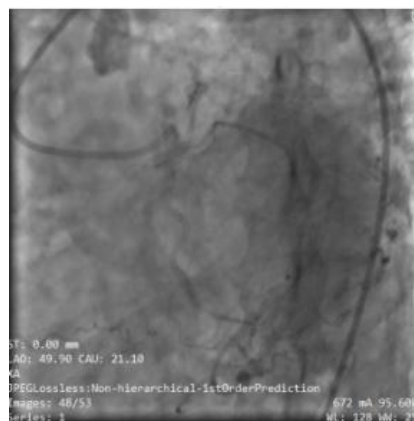
4.2.3 Análisis de datos faltantes y consistencia

Aunque el conjunto de datos no presenta valores nulos en el sentido de una base de datos estructurada, se evaluó la completitud de las imágenes:

- **Imágenes no informativas:** Se identificaron casos donde la inyección de contraste era insuficiente o el momento de captura no correspondía al llenado arterial máximo, resultando en imágenes donde las arterias no eran visibles. Estas imágenes se asumen como ruido de entrenamiento.

Figura 3

Imagen de arteriografía coronaria no informativa.



Fuente: Elaboración propia.

- **Heterogeneidad dimensional:** Con el fin de brindar consistencia al estudio, las imágenes fueron sometidas a un proceso de redimensionamiento estandarizado a $256 * 256$, independientemente de sus resoluciones. Esta etapa permitió garantizar la compatibilidad con el procesamiento por lotes.

De acuerdo con lo anterior, se determinó que, aunque el conjunto de datos es completo en información, también es ruidoso. La presencia de atípicos como el texto y la variabilidad de contraste definen que el éxito del modelado no dependería solo de la arquitectura neuronal,

sino de un preprocesamiento robusto de limpieza espacial y realce selectivo, estrategias que se definieron para la siguiente fase.

4.3 Preparación de los Datos.

Con base en el entendimiento de datos, se diseñó e implementó un flujo de trabajo de transformación de datos para reducir el ruido y generar las estructuras necesarias para el entrenamiento no supervisado. Esta fase se dividió en dos procedimientos secuenciales:

4.3.1 Limpieza y estandarización

Para neutralizar los artefactos identificados en la fase anterior, se aplicaron las siguientes operaciones de preprocesamiento a la totalidad del conjunto de datos:

- **Recorte espacial automático:** Se desarrolló un algoritmo de recorte fijo para eliminar el 15% inferior de cada imagen (Figura 4). Esta operación tuvo como objetivo eliminar los metadatos de texto impresos que, al tener bordes nítidos y alto contraste, interferían con la detección de gradientes en las redes convolucionales.

Figura 4

Imagen de arteriografía coronaria original vs. Imagen de arteriografía coronaria recortada.



Fuente: Elaboración propia.

La decisión de aplicar un recorte espacial para remover los metadatos quemados en la angiografía, en lugar de utilizar técnicas modernas de reconstrucción de imagen

como el inpainting basado en aprendizaje profundo, se debe a un principio de seguridad clínica y preservación de la integridad de los datos.

La literatura demuestra un fenómeno de alucinaciones estructurales propios de los modelos generativos. En el contexto de la restauración de imágenes médicas, el inpainting tiende a sintetizar texturas anatómicamente aceptables, pero físicamente inexistentes, o en otros casos, a borrar características que podrían tener relevancia clínica [27], [28]. Dado que los vasos coronarios distales en las angiografías presentan un bajo contraste y se superponen con ruido de fondo, intentar predecir los píxeles detrás de los caracteres introduciría estructuras vasculares inexistentes. En un esquema como el propuesto, W-Net asimilaría los artefactos sintéticos como vasos reales, dañando el modelo de optimización. Por lo anterior, el recorte espacial garantiza que el 100% de los píxeles que ingresan a la red conservan la originalidad de adquisición inicial, evitando la propagación de alucinaciones morfológicas al mapa de segmentación.

- **Redimensionamiento y normalización:** Para asegurar la compatibilidad con la arquitectura W-Net y la estabilidad numérica del descenso de gradiente, todas las imágenes fueron redimensionadas a la misma resolución de píxeles de 256 x 256 píxeles con 3 canales (RGB) seleccionada para equilibrar la carga computacional y la preservación de detalles en los vasos. Posteriormente, antes de ingresar a la red, se normalizaron los valores de intensidad de los píxeles, escalándolos del rango original $[0, 255]$ enteros de 8 bits, al rango $[0, 1]$, este escalamiento es importante para el correcto funcionamiento del optimizador y la estabilidad de las capas de activación.

4.3.2 Generación de máscaras de verdad absolutas

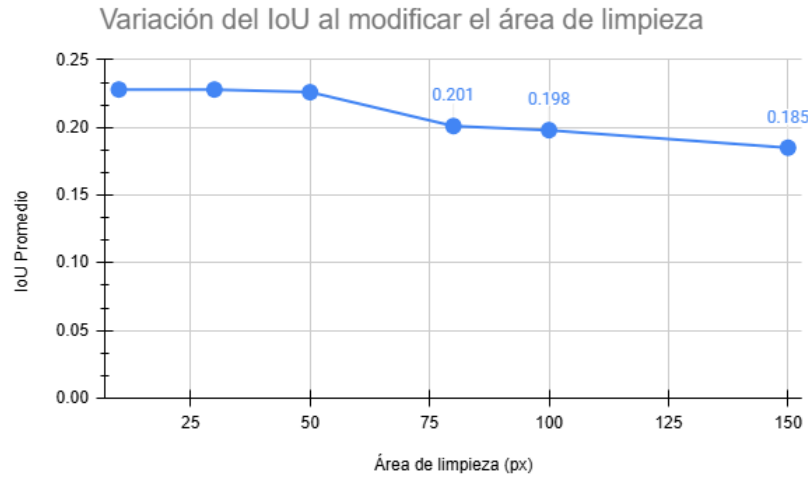
Ante la ausencia de etiquetas manuales, se generaron etiquetas de referencia o máscaras de verdad para guiar el aprendizaje inicial del modelo W-Net. Se implementó un algoritmo secuencial compuesto por tres etapas matemáticas:

- **Filtro de Sato:** Se aplicó el filtro de Sato a las imágenes preprocesadas. Este se basa en el análisis de los valores λ_1, λ_2 de la matriz Hessiana de la imagen para cada píxel [29]. Matemáticamente, el filtro se configuró para maximizar la respuesta en estructuras tubulares donde $\lambda_1 = 0$ y $\lambda_2 > 0$, suprimiendo eficazmente el ruido plano y estructuras esféricas, haciendo uso también de la función black ridges para identificar estructuras oscuras como los vasos contrastados.
- **Método de Otsu:** La salida probabilística del filtro de Sato, un mapa de calor de tubularidad, se convirtió en una máscara binaria mediante el método de Otsu. Este calculó el umbral global óptimo que maximiza la varianza interclase, separando los píxeles candidatos del fondo ruidoso sin intervención de un experto [30] resultando en una máscara binaria inicial.
- **Limpieza de la morfología:** Debido a que el proceso anterior podía generar pequeños artefactos, se aplicó una función de limpieza basada en componentes conectados. Se eliminaron todas las estructuras con un área inferior a 50 píxeles, resultando en máscaras de entrenamiento más limpias.

Esta decisión respecto al área de corte no fue arbitraria, sino que se fundamenta tanto en la literatura como en un análisis propio. Desde el punto de vista teórico, diversos estudios en el procesamiento de imágenes médicas señalan que las estructuras aisladas de áreas muy reducidas corresponden típicamente a ruido, artefactos de los equipos o estructuras no objetivo [31], [32]. Esta premisa teórica se comprobó empíricamente en el presente conjunto de datos, tal como se ilustra en la (Figura 5). Al observar el comportamiento del índice de Intersección sobre la Unión (IoU) en función del área de limpieza, se identifica un comportamiento de rendimiento interesante.

Figura 5

Caracterización del IoU en función del área de limpieza definida.



Fuente: Elaboración propia.

La estabilidad observada en valores bajos como de 10 a 30 píxeles es engañosa porque muestra coincidencia con el ruido de las etiquetas heurísticas más que una limpieza. Se identifica un punto de inflexión en los 50 píxeles, el cual es el límite máximo antes de entrar en una zona de degradación; superar ese umbral provoca una caída muy grande en la métrica lo que nos indica que el filtrado se vuelve muy agresivo, eliminando partes vasculares y afectando la continuidad de los vasos más finos.

4.4 Modelado Individual.

Esta fase se enfocó en dar cumplimiento al primer objetivo específico, Identificando las limitaciones y ventajas de W-Net y SAM; fue necesario entonces implementar ambos modelos para someterlos a una prueba real con el conjunto de datos.

La configuración de estos modelos estableció los cimientos para abordar el tercer objetivo específico sobre capacidad de explicabilidad, integrando mecanismos de visualización de mapas de características.

A continuación, se detalla la implementación algorítmica de cada enfoque:

4.4.1 Implementación del Modelo Especialista: W-Net

Se seleccionó la arquitectura W-Net como modelo especialista. Su implementación individual se hizo con el fin de evaluar si una red neuronal puede aprender la morfología vascular compleja sin supervisión.

- **Descripción Algorítmica:** La red se construyó concatenando dos arquitecturas U-Net secuenciales en una estructura de W compuesta entonces por codificador-decodificador-codificador-decodificador.
 - **Bloque de segmentación U-Net (U_{seg}):** Es un codificador que comprime la imagen y genera una máscara de probabilidad binaria. Se implementaron saltos de conexión entre las capas de contracción y expansión lo cual consiste en preservar la información que se tiende a perder durante el max-pooling [33] o fraccionamiento y reducción de la imagen a través de las capas del modelo [34].
 - **Bloque de reconstrucción U-Net (U_{rec}):** Toma la salida del primer bloque y la procesa para reconstruir la imagen original. Este paso es fundamental, ya que obliga a la red a retener información estructural en la máscara de segmentación para poder recuperar la imagen de entrada con éxito.
- **Función de optimización y pérdida:** Basado en la experimentación empírica y reflejando la implementación en código, se diseñó un esquema de optimización multitarea. La función de pérdida global (L_{total}) se define como una combinación lineal ponderada que busca equilibrar la tarea de segmentación con la fidelidad de la reconstrucción:

$$L_{total} = \lambda_{seg} \cdot L_{seg} + \lambda_{rec} \cdot L_{rec} \quad (3)$$

Los coeficientes λ_{seg} y λ_{rec} actúan como hiperparámetros que priorizan la especialización del modelo. Los valores óptimos para esta ponderación se determinaron mediante un análisis de convergencia y desempeño detallado en la Sección 5.1 (Resultados del Modelo Especialista), seleccionando una configuración que favorece la tarea de segmentación.

Cada componente de esta ecuación se definió con métricas específicas para contrarrestar el ruido de la angiografía:

- **Pérdida de Segmentación (L_{seg}):** Se utilizó la pérdida de Dice ($1 - \text{Dice_Coefficient}$) calculada contra las pseudoetiquetas. Esta métrica mide la sobre posición y evalúa la similitud entre los resultados de segmentación y las etiquetas de referencia comparando la intersección y es insensible al desbalance de clases, lo cual es importante dado que los vasos ocupan alrededor del 10% de la imagen [35].

$$L_{seg} = 1 - \frac{2 \cdot \sum (Y_{ref} \cdot Y_{pred})}{\sum Y_{ref} + Y_{pred}} \quad (4)$$

- **Pérdida de Reconstrucción Perceptual (L_{rec}):** Como diferencia de las implementaciones más comunes que usan el Error Cuadrático Medio (MSE), se utilizó una función de pérdida híbrida que combina el Índice de Similitud Estructural (SSIM) el cual evalúa la luminosidad, el contraste y la estructura y el Error Absoluto Medio (L1) que pondera cada píxel de la imagen original asignándole un valor que luego se compara con la imagen de reconstrucción [36].

$$L_{rec} = 0.85 \cdot (1 - \text{SSIM}(X, X')) + 0.15 \cdot \|X - X'\| \quad (5)$$

Esta ponderación fue establecida como la configuración base del modelo, orientada a priorizar la preservación de la luminancia, el contraste y la estructura por encima de la exactitud del valor de píxel individual. Esta elección de hiperparámetros busca favorecer la continuidad de las estructuras vasculares segmentadas.

Adicional a la función de pérdida, la arquitectura integra decisiones de diseño para estabilizar el aprendizaje no supervisado frente a la variabilidad de las angiografías:

- **Normalización por Lotes (Batch Normalization):** Se implementó una capa de Batch Normalization [37] después de cada convolución y antes de la activación. Dado que las imágenes tienen distribuciones diferentes de intensidad, esta normalización fue importante para reordenar los datos, permitiendo el uso de tasas de aprendizaje más altas.
- **Funciones de Activación:** Se definió la función ReLU (Unidad Lineal Rectificada) para todas las capas ocultas del codificador y decodificador con el objetivo de disminuir el desvanecimiento del gradiente. Para la capa de salida final, se empleó una activación sigmoide, la cual es necesario para comprimir los valores resultantes en el rango de probabilidad $[0,1]$, lo cual es necesario para la generación de la máscara binaria y el cálculo de la pérdida de Dice [38], [39].
- **Estrategia de Entrenamiento:** El modelo se compiló con el optimizador Adam con el objetivo de aprovechar su capacidad de ajuste de las tasas de aprendizaje en función de los gradientes [40]. Debido a que el paisaje de la función de pérdida en el aprendizaje no supervisado no es convexo y tiende a estancamientos en mínimos locales, se implementó un método de optimización dinámico con dos mecanismos de control (callbacks):
 - **Checkpoint de Modelo:** Una función para guardar los pesos que tienen el coeficiente de Dice más alto en el conjunto de validación, asegurando que el

modelo final sea el que mejor desempeño generalizable tiene en lugar del último estado del entrenamiento.

- **Decaimiento de la Tasa de Aprendizaje:** Se utilizó la técnica ReduceLROnPlateau para monitorear la pérdida de validación. Esto reduce la tasa de aprendizaje en un factor de 0.2 cuando el rendimiento del modelo no mejora después de 5 épocas consecutivas (paciencia). Esta estrategia permite al modelo realizar saltos grandes explorando al inicio y, después, disminuir la velocidad de actualización para realizar ajustes finos, asegurando una mejor convergencia hacia los bordes vasculares [41], [42].

4.4.2 Implementación del Modelo Generalista: SAM

Para diferenciarse del modelo especialista, se implementó el modelo Segment Anything Model, con el objetivo de evaluar su aplicación en entornos biomédicos sin un reentrenamiento específico dando entonces un uso aplicado de la técnica zero-shot [43].

- **Configuración del Modelo:** Se utilizó la variante ViT-H (Vision Transformer Huge), la más robusta de SAM [44]. A diferencia de las redes convolucionales como W-Net que procesan píxeles locales, SAM utiliza técnicas de enfoque global para comprender la geometría de la imagen.
- **Mecanismo de Inferencia:** Se implementó la clase SamAutomaticMaskGenerator sin darle al modelo prompts manuales, configurando el algoritmo para generar una rejilla automática de puntos sobre toda la imagen [45]. Esto permitió evaluar la capacidad del modelo para proponer segmentaciones, identificando la limitación principal que resulta en la segmentación de estructuras que no son vasos, debido a la falta de contexto médico.

4.4.3 Implementación del Modelo Híbrido

La contribución central de esta investigación es el algoritmo de fusión que integra la salida de ambos modelos. Basado en los hallazgos observados en los modelos individuales donde W-

Net ofreció localización robusta, y SAM ofreció precisión, se desarrolló una arquitectura secuencial de las generalidades a los detalles minúsculos. La implementación se estructuró en tres componentes de ingeniería: gestión de recursos, generación paralela y fusión lógica.

- **Gestión de Recursos:**

Un desafío identificado durante el desarrollo del modelo híbrido fue la inviabilidad de ejecutar los dos modelos en un entorno de GPU estándar. El modelo SAM ViT-H posee 636 millones de parámetros [46], lo cual satura la memoria RAM al intentar cargar W-Net de forma seguida.

Para solucionar eso, se implementó un protocolo de ejecución secuencial con descarga activa:

- **Inferencia SAM:** Se carga el modelo ViT-H, se realiza la predicción y, mediante comandos de gestión de memoria se purga el modelo de la GPU antes de avanzar.
- **Redimensionamiento:** Se identificó que el consumo de memoria del modelo SAM depende directamente de la resolución espacial de la imagen de entrada, debido al procesamiento basado en Vision Transformers [47]. De acuerdo con eso, se implementó un algoritmo de normalización dimensional. Este proceso utiliza interpolación bilineal para ajustar la imagen a un límite máximo de entrada de 768 píxeles, manteniendo siempre la relación de aspecto original para evitar distorsiones morfológicas en los vasos.

El redimensionamiento consistió en calcular un factor de escala proporcional a las dimensiones originales de la imagen así:

$$\text{Factor de escala} = \frac{768}{\max(H, W)} \quad (6)$$

Donde H y W representan la altura y anchura original de la imagen, respectivamente. El algoritmo ajusta la resolución de entrada para que su dimensión máxima corresponda a 768 píxeles, preservando la relación de aspecto original. En casos donde las imágenes tienen resoluciones reducidas, este procedimiento implica un reescalado espacial ascendente (upsampling) previo a la inferencia en SAM.

Este proceso resulta relevante debido a que el modelo SAM fue preentrenado utilizando resoluciones a 1024 x 1024 píxeles. Por eso, trabajar con resoluciones superiores permite que los mecanismos de atención del Transformer representen con mayor estabilidad estructuras vasculares que podrían perder continuidad en resoluciones bajas.

Posteriormente, una vez generadas las segmentaciones, las máscaras son proyectadas nuevamente a la resolución original mediante interpolación de vecino más cercano (INTER_NEAREST). Este método evita la introducción de valores intermedios durante el reescalado, preservando la naturaleza binaria de la segmentación y disminuyendo alteraciones que podrían afectar métricas de superposición como el Coeficiente Dice [47], [48], [49].

- **Algoritmo de Fusión**

El sistema implementado no modifica ni reentrena los modelos individuales. En su lugar, la interacción entre el modelo especialista W-Net y el modelo generalista SAM se realiza mediante un algoritmo de filtrado espacial, implementando una estrategia de fusión tardía (Late Fusion). Este enfoque permite combinar las fortalezas de ambos modelos sin alterar sus parámetros originales [50], [51].

Primero, el modelo SAM genera automáticamente un conjunto de regiones candidatas al segmentar la imagen en múltiples máscaras. Estas máscaras

presentan contornos precisos, pero incluyen estructuras no deseadas, como costillas y artefactos de la imagen, debido a su naturaleza generalista.

Posteriormente, el modelo W-Net procesa la misma imagen y produce un mapa continuo de probabilidad que indica la presencia de vasos. Este mapa se binariza utilizando un umbral bajo de sensibilidad, con el objetivo de conservar la mayor cantidad posible de información vascular y definir una región de interés amplia que actúa como guía anatómica.

La integración final o fusión se realiza mediante la evaluación del grado de superposición entre ambas fuentes de información, según la siguiente relación de coherencia (ecuación 7):

$$S_i = \frac{\text{Área}(m_i \cap P_{\text{wnet}})}{\text{Área}(m_i)} \quad (7)$$

Donde m_i corresponde a la máscara generada por el modelo SAM y P_{wnet} representa el mapa binarizado obtenido por el modelo W-Net. Si el puntaje S_i supera el umbral de coherencia establecido experimentalmente en 35%, el sistema determina que la máscara m_i corresponde a una estructura vascular validada semánticamente por W-Net y la conserva. De lo contrario, se descarta como ruido de fondo.

Este mecanismo permite eliminar automáticamente segmentaciones irrelevantes generadas por el modelo generalista, conservando únicamente aquellas que son consistentes con el conocimiento anatómico aprendido por el modelo. La selección de este valor de umbral de coherencia se fundamenta en una fase de calibración detallada en la sección de resultados (Sección 5.3.), donde se evaluó el punto de equilibrio entre la recuperación de vasos y la eliminación de artefactos.

4.5 Validación y Explicabilidad

Con el objetivo de evaluar la capacidad de explicabilidad y visualización, así como para medir el rendimiento según el diseño, se implementaron los siguientes mecanismos:

- **Visualización de Mapas de Calor**

Más allá de la segmentación binaria final, el sistema extrae y visualiza el mapa probabilístico de segmentación generado por W-Net antes de la fase de umbralización. Esta representación corresponde a las activaciones de la última capa del modelo, las cuales representan la probabilidad de que cada píxel pertenezca a la estructura vascular coronaria. Según lo propuesto por Ronneberger et al. 2015, esta configuración permite que la capa de salida genere una distribución de probabilidad píxel a píxel [52].

La utilidad de extraer este mapa consiste en su capacidad para mostrar la distribución de confianza del modelo. La visualización de estas activaciones mediante el mapa de calor es una herramienta de interpretación, que permite observar la distribución espacial de los niveles de confianza de la red [53].

En este estudio, la visualización se realiza mediante una escala de colores, donde las zonas cálidas (rojo y amarillo) indican una alta probabilidad de presencia vascular, mientras que las zonas frías (azul) representan el fondo. Esto permite diagnosticar si los errores de segmentación provienen de una falla en la detección de W-Net o una falla en la delimitación.

- **Métricas de Evaluación Cuantitativa**

Para lograr medir el desempeño de los modelos de manera objetiva, no se empleó la métrica de precisión tradicional. Esto es porque, en una angiografía, los vasos sanguíneos ocupan menos del 10% de la imagen, entonces un modelo que predijera todo como un fondo negro obtendría una precisión del 90% siendo inútil.

Por esta razón, se seleccionaron dos métricas basadas en la superposición espacial, estándar en la literatura de segmentación médica:

- **Coeficiente DICE (DSC)**

Es la métrica central de este estudio, la cual mide la similitud entre dos conjuntos de datos o en este caso imágenes, identificando el porcentaje de píxeles de la imagen predicha segmentados correctamente en comparación con la imagen de la máscara de verdad [54] la cual, como se detalló en la sección 4.3.2, se generó mediante un pipeline compuesto por el filtro de Sato, la umbralización de Otsu y una limpieza morfológica de componentes menores a 50 píxeles.

$$DSC = \frac{2 \cdot |Y_{seg} \cap Y_{pred}| + \epsilon}{|Y_{seg}| + |Y_{pred}| + \epsilon} \quad (8)$$

Donde Y_{seg} representa el conjunto de píxeles definidos por el modelo híbrido como vasos, Y_{pred} representa el conjunto de píxeles de la pseudo etiqueta o máscara de verdad generado y ϵ es una constante utilizada para evitar indeterminaciones.

Aquí la métrica DSC otorga el doble de peso a la intersección $2 \cdot |Y_{seg} \cap Y_{pred}|$ lo cual penaliza al modelo si no logra detectar los vasos falsos positivos o si inventa vasos donde no existen, siendo más confiable que la métrica de precisión común para evaluar la continuidad de estructuras delgadas.

- **Intersección sobre la Unión (IoU)**

IoU se implementó como métrica complementaria y más estricta. A diferencia de DSC, esta compara la coincidencia frente al área total combinada de ambas máscaras.

$$IoU = \frac{|Y_{seg} \cap Y_{pred}| + \epsilon}{|Y_{seg} \cup Y_{pred}| + \epsilon} \quad (9)$$

Esta métrica es conocida por ser la prueba absoluta en segmentación. Al tener la unión en el denominador, penaliza con severidad las desviaciones [55]. Si el modelo SAM segmenta el vaso correctamente, pero lo hace, aunque sea un poco más grueso de lo real, el IoU disminuye significativamente, mientras que el DSC puede ser más flexible. Utilizar las dos métricas permite diferenciar si el modelo está acertando en la ubicación con un buen DSC y en la precisión de la forma con un buen IoU.

5 RESULTADOS

En este capítulo se presentan los resultados obtenidos a partir de la implementación y evaluación de los modelos de segmentación propuestos y ya mencionados. El objetivo principal aquí es analizar el desempeño del pipeline en términos de capacidad de discriminación vascular y estabilidad del proceso de segmentación bajo un aprendizaje no supervisado.

Debido a la ausencia de anotaciones manuales realizadas por expertos, no se realiza una validación supervisada convencional. Por esto, se adoptó una estrategia de validación basada en la generación de una referencia base o máscara de verdad, construida a partir de las máscaras binarias generadas por el pipeline heurístico. Estas máscaras son etiquetas verdaderas para el cálculo de las funciones de pérdida y métricas de evaluación, lo que permite cuantificar el comportamiento del sistema y analizar su consistencia interna.

Los resultados presentados, se deben entender, como una evaluación comparativa del desempeño del modelo dentro del marco metodológico propuesto, más que como una medición absoluta frente a una verdad definitiva. Finalmente, se almacenaron los pesos de los distintos modelos y los puntos de control del modelo SAM, lo que permitió optimizar el uso de memoria y garantizar la reproducibilidad de los experimentos durante la ejecución del sistema híbrido.

5.1 Resultados del Modelo Especialista (W-Net)

Durante el entrenamiento del modelo W-Net, se evaluó el impacto de la función de pérdida combinada en la calidad de la segmentación. Para dar respuesta a la configuración de los hiperparámetros, se realizaron experimentos con tres variaciones de pesos para los componentes de segmentación y reconstrucción.

Es fundamental precisar que el cálculo del Dice Score (DSC) presentado en la Tabla 1 no corresponde a una evaluación aislada sobre una imagen o un parche. Por el contrario,

representa el promedio obtenido sobre el conjunto de validación completo, el cual constituye el 20% del dataset total. Asimismo, el modelo fue entrenado y validado utilizando la arquitectura completa de la imagen, garantizando que la métrica refleje la capacidad del modelo para mantener la continuidad de las estructuras vasculares en todo el campo de visión.

Se evaluaron tres configuraciones de pesos (λ) bajo las mismas condiciones de inicialización. Los resultados de convergencia en el conjunto de validación se presentan en la (Tabla 1).

Tabla 1

Rendimiento con diferentes pesos de pérdida.

Configuración	Peso Segmentación (λ_{seg})	Peso Reconstrucción (λ_{rec})	Dice Score	Comportamiento
1	0.50	0.50	0.765	Bordes inconsistentes
2	0.70	0.30	0.801	Convergencia estable.
3	0.85	0.15	0.900	Mejor definición de vasos finos.

Fuente: Elaboración propia.

La Configuración 3 demostró tener un mejor comportamiento, lo que indica que priorizar la pérdida de segmentación permite al modelo especializarse en la discriminación de la clase, mientras que el componente de reconstrucción actúa como un regulador estructural para evitar la pérdida de coherencia espacial. Con base en esto, se seleccionó esta configuración para los análisis siguientes y la integración híbrida.

Para demostrar la estabilidad de la Configuración 3 y asegurar la reproducibilidad de los hallazgos, se ejecutó un protocolo de validación utilizando el dataset de 1,606 imágenes. La separación del dataset se realizó mediante una división aleatoria (`random_state = 42`) para asegurar la reproducibilidad de la distribución entre entrenamiento y validación.

La distribución del conjunto de datos y los hiperparámetros generales utilizados durante el proceso de optimización se presentan en la Tabla 2.

Tabla 2

Configuración general del entorno experimental para el modelo especialista

Dimensión Operativa	Especificación	Cantidad o Valor
Volumen de datos	Imágenes de entrenamiento (80%)	1,285
	Imágenes de validación (20%)	321
	Total de dataset	1,606
Parámetros del modelo	Resolución	256 x 256 píxeles
	Tamaño del lote	8
Estrategia para optimización	Algoritmo optimizador	Adam (Tasa de aprendizaje 0,0001)
	Límite de épocas	40

Fuente: Elaboración propia.

El entrenamiento del modelo se estableció con un límite máximo de 40 épocas, definido a después del comportamiento observado en las curvas de pérdida y validación durante las fases preliminares de experimentación. Se identificó que las mayores reducciones de la función de pérdida ocurrían durante las primeras etapas del entrenamiento, mientras que las

épocas finales contribuían principalmente al refinamiento progresivo de la segmentación vascular.

Dado que la arquitectura propuesta corresponde a un modelo multitarea, la dinámica de convergencia presentó diferentes comportamientos entre sus componentes internos. La rama de reconstrucción perceptual, asociada a la pérdida híbrida basada en SSIM y L1, alcanzó estabilidad temprana, mientras que la rama de segmentación continuó refinando gradualmente la representación morfológica de las estructuras vasculares durante las últimas épocas.

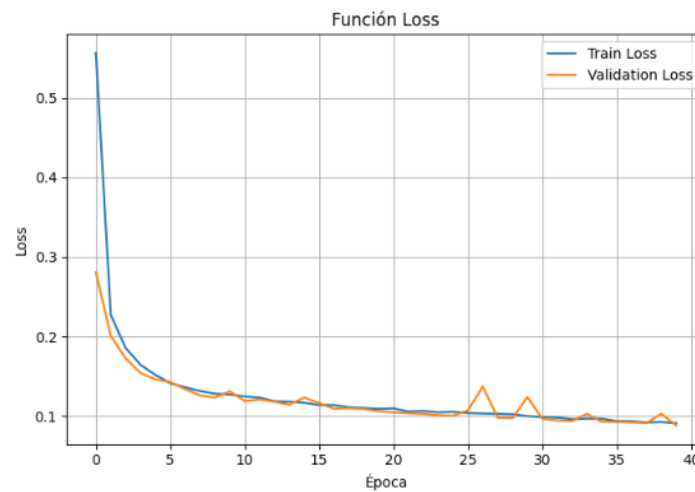
Con el propósito de optimizar el entrenamiento y evitar bajar el rendimiento, se implementaron dos mecanismos automáticos de control:

- **ReduceLROnPlateau:** Fue configurado con una paciencia de 5 épocas y un factor de reducción de 0.2, encargado de disminuir dinámicamente la tasa de aprendizaje al identificar estancamiento en la métrica de validación.
- **ModelCheckpoint:** Fue utilizado para almacenar los pesos correspondientes al mejor desempeño obtenido sobre el conjunto de validación.

La Figura 6 presenta la evolución histórica de la función de pérdida global durante el entrenamiento y la validación del modelo.

Figura 6.

Comportamiento de las curvas de pérdida (Loss) en entrenamiento y validación.



Fuente: Elaboración propia.

La evolución de la pérdida permite ver un descenso durante las primeras épocas, seguido de una etapa de estabilización progresiva. A partir de aproximadamente la época 15, las curvas de entrenamiento y validación mantienen un trayecto cercano y paralelo, indicando estabilidad durante el proceso de optimización y ausencia de divergencias significativas asociadas a sobreajuste.

La fase experimental, incluyó distintas configuraciones de ponderación para la función de pérdida, con el objetivo de analizar el balance entre segmentación vascular y reconstrucción perceptual. No obstante, la Tabla 3 documenta específicamente el comportamiento de la configuración definitiva seleccionada para la implementación del sistema híbrido.

Tabla 3

Evaluación Del Desempeño Modelo W-Net.

Época	Pérdida global (Loss)	Dice Score (Entrenamiento)	Dice Score (Validación)	Estado
38	0.0898	0.8945	0.8965	Estable
39	0.0886	0.8959	0.8878	Oscilación
40	0.0857	0.8995	0.9006	Óptimo global

Fuente: Elaboración propia.

Los resultados finales muestran una convergencia estable en el proceso de optimización. Aunque se observa una ligera fluctuación en la métrica de validación durante la época 39, el modelo recupera la tendencia de mejora en la última iteración, alcanzando el mayor desempeño global del entrenamiento.

Asimismo, la cercanía entre las métricas de entrenamiento y validación evidencia una adecuada capacidad de generalización y bajo sobreajuste. En especial, el DSC de validación (0.9006) superó ligeramente al obtenido durante entrenamiento (0.8995), lo que sugiere que la arquitectura y la incorporación de la reconstrucción perceptual contribuyeron a mantener estabilidad estructural durante el aprendizaje.

La reducción de la pérdida de reconstrucción indica además que el cuello de la W-Net logró conservar información anatómica importante durante el proceso de codificación, favoreciendo segmentaciones con mayor continuidad espacial y coherencia morfológica en vasos coronarios pequeños.

5.2 Resultados del Modelo SAM

Para contextualizar cómo se desempeñó el modelo especialista, se comparó con el modelo generalista SAM en su aplicación de zero-shot. Para garantizar una comparación equitativa y con validez estadística, se realizó sobre el mismo conjunto de validación independiente

derivado de la base de datos original compuesto por las 321 imágenes aisladas, equivalentes al 20% del dataset total. El término aislado se da porque este subconjunto fue protegido y excluido de la fase de entrenamiento del modelo especialista, asegurando que las imágenes evaluadas fueran completamente desconocidas para los gradientes y parámetros de ambos algoritmos permitiendo un análisis comparativo equilibrado.

La evaluación se llevó a cabo utilizando las métricas de Coeficiente de Dice (DSC) e Intersección sobre Unión (IoU). A diferencia de una implementación estándar sin restricciones, se evaluó el desempeño de SAM incorporando un postprocesamiento basado en hiperparámetros de intensidad y solidez de formas. El objetivo fue determinar si un modelo como este, puede ser adaptado mediante heurísticas simples sin necesidad de reentrenamiento.

Los resultados obtenidos revelaron un comportamiento dual caracterizado por precisión geométrica en las estructuras, aunque condicionado por la sensibilidad debido a la dependencia de parámetros heurísticos fijos. A diferencia del modelo especialista W-Net, el modelo SAM fue utilizado bajo la técnica zero-shot, sin actualización de pesos ni retropropagación de gradientes durante esta investigación.

Por esta razón, en esta etapa no se reportan métricas asociadas a pérdida, curvas de convergencia ni evolución por épocas, dado que la arquitectura fundacional fue utilizada exclusivamente como generador automático de regiones candidatas.

Durante el proceso experimental, el módulo SamAutomaticMaskGenerator generó múltiples máscaras independientes sobre las imágenes angiográficas. Posteriormente, las regiones obtenidas fueron expuestas al pipeline heurístico de refinamiento compuesto por filtrado por área, intensidad y sobre imágenes mejoradas con CLAHE y análisis morfológico basado en solidez del contorno.

El valor promedio de 0.9358 reportado en la Tabla 4 corresponde a la media aritmética global del parámetro predicted_iou generado por SAM para las máscaras que superaron las etapas de filtrado y fueron consideradas compatibles con estructuras vasculares.

Tabla 4

Evaluación Del Desempeño Modelo SAM.

Métrica evaluada	Valor promedio	Interpretación
IoU predicho	0.9358	Alto grado de superposición (93%)
Sensibilidad en la detección del vaso	Medio	El modelo es muy selectivo; tiende a ignorar estructuras que no cumplen con los umbrales de contraste determinados.

Fuente: Elaboración propia.

El valor promedio indica que las estructuras conservadas por el sistema presentan una delimitación espacial consistente y bordes bien definidos. Sin embargo, los experimentos evidenciaron limitaciones importantes en la capacidad de generalización del sistema heurístico. Debido a la utilización de umbrales fijos relacionados con intensidad, tamaño y solidez morfológica, pequeñas variaciones en el contraste angiográfico o en el calibre vascular pueden ocasionar la eliminación de ramas distales o vasos de baja visibilidad.

Estos hallazgos soportan el desarrollo del modelo híbrido propuesto, donde las restricciones manuales se reemplazan parcialmente por la representación probabilística aprendida por W-Net, permitiendo una selección anatómica más completa y adaptable frente a variaciones clínicas reales.

5.3 Resultados del Modelo Híbrido

La fase final de la experimentación consistió en integrar la capacidad de localización anatómica del modelo especialista W-Net con la precisión geométrica del modelo fundacional SAM. Esta integración se realizó mediante un mecanismo de coherencia espacial, donde el mapa probabilístico generado por W-Net actuó como una guía anatómica para validar las máscaras candidatas propuestas por SAM.

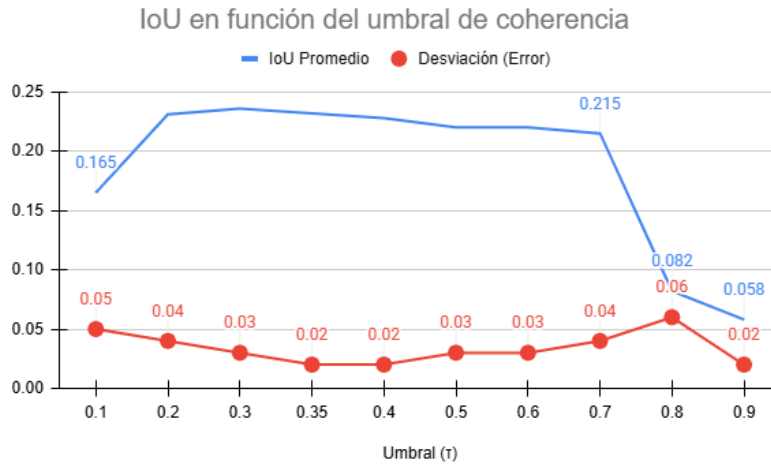
Con el propósito de determinar el umbral óptimo de coherencia espacial, se desarrolló un análisis de sensibilidad sobre el parámetro S_i , evaluando su impacto sobre las métricas de segmentación en el conjunto de validación compuesto por 321 imágenes previamente aisladas del entrenamiento. Para cada valor de umbral evaluado, el sistema ejecutó el proceso completo de inferencia híbrida, calculando posteriormente las métricas DSC e IoU respecto a las máscaras de referencia.

El comportamiento permitió ver tres regiones diferenciadas (Figura 7) Para valores inferiores al 20%, el sistema tendía a aceptar regiones anatómicamente irrelevantes, incluyendo costillas, ruido estructural y artefactos producidos por el medio de contraste. Por el contrario, valores superiores al 50% incrementaban excesivamente el rigor del filtrado, provocando la eliminación de ramas distales y segmentos vasculares de bajo calibre. Entre ambos extremos se identificó una zona de estabilidad comprendida entre el 20% y el 40%, donde las métricas permanecieron relativamente constantes y las segmentaciones conservaron coherencia anatómica visual.

A partir de este comportamiento, se estableció experimentalmente un umbral óptimo de coherencia espacial del 35%, correspondiente al mejor equilibrio entre sensibilidad vascular y reducción de ruido estructural.

Figura 7

IoU en función del umbral de coherencia.



Fuente: Elaboración propia.

Una vez fijado el umbral, el modelo híbrido fue evaluado mediante ejecuciones iterativas sobre imágenes representativas del conjunto de validación independiente. En cada ejecución, el sistema desarrolló el proceso completo de inferencia híbrida, integrando las máscaras propuestas por SAM con el mapa probabilístico generado por W-Net. Posteriormente, las segmentaciones obtenidas fueron comparadas frente a las máscaras de referencia generadas mediante el pipeline heurístico Sato + Otsu, calculando las métricas DSC e IoU para cada caso analizado.

Los valores reportados en la Tabla 5 corresponden al rango de variación observado en las métricas promedio de validación durante las ejecuciones finales del sistema híbrido bajo el umbral óptimo de coherencia espacial del 35%. Este comportamiento permitió evaluar no solo el rendimiento global, sino también la estabilidad morfológica del sistema frente a diferentes configuraciones anatómicas y variaciones de contraste presentes en las angiografías.

Tabla 5

Evaluación De Desempeño En El Modelo Híbrido.

Métrica	Rango (Validación)	Interpretación
Coeficiente DICE (DSC)	0.4303 - 0.4601	Disminución cuantitativa respecto al modelo W-Net.
Intersección sobre la unión (IoU)	0.2741 - 0.2987	Divergencia morfológica frente a la máscara.

Fuente: Elaboración propia.

Para interpretar correctamente los resultados de la Tabla 5 y resolver la aparente contradicción entre la mejora visual observada y la disminución de las métricas cuantitativas, es necesario analizar la naturaleza de las máscaras de referencia utilizadas durante la evaluación y la sensibilidad inherente de las métricas basadas en superposición espacial.

5.3.1 La Divergencia Con La Referencia Heurística

El hallazgo científico de la implementación del modelo híbrido se hace evidente al analizar los resultados cuantitativos. Aquí se observa una contradicción donde el modelo híbrido presenta métricas de superposición $\text{IoU} = 0.28$ que es muy inferior a las del modelo W-Net calculado $\text{IoU} = 0.81$. Sin embargo, esto lejos de indicar una falla en la segmentación, constituye el hallazgo principal de la fase de validación, que llamaremos sesgo o divergencia con la referencia heurística.

Esta diferencia se fundamenta en la calidad de las máscaras de verdad utilizadas. Dado que las etiquetas de entrenamiento no fueron generadas por expertos humanos, sino por algoritmos de filtrado Sato + Otsu, estas representan un estándar de referencia secundario o referencia ruidosa, caracterizada por segmentaciones fragmentadas y subestimación del tamaño y longitud de los vasos.

El análisis comparativo permite establecer dos comportamientos diferentes:

1. **Sobreajuste al Ruido (W-Net):** Este modelo, al ser entrenado para disminuir la pérdida contra las etiquetas que no son perfectas, aprendió a replicar los errores del algoritmo de Otsu de trazos rotos y delgados. Su alto puntaje refleja una alta similitud con una referencia que no es perfecta, no necesariamente una correcta representación anatómica.
2. **Generalización Estructural (Híbrido/SAM):** El componente SAM, operando bajo el enfoque zero-shot, posee patrones visuales aprendidos que le permiten reconstruir la morfología del vaso. Al generar una máscara continua y de representación real, esta deja de coincidir píxel a píxel con la referencia delgada y rota, resultando en una penalización del índice IoU.

Como se menciona en la literatura sobre segmentación con supervisión débil [56], las métricas basadas en superposición como Dice e IoU, representan las herramientas principales para la evaluación cuantitativa de modelos de segmentación. Sin embargo, la interpretación de sus valores debe contextualizarse respecto a la calidad y naturaleza del Ground Truth utilizado durante la validación [57]. Estas métricas miden la coincidencia píxel a píxel frente a la máscara de referencia disponible, y no precisamente frente a la anatomía clínica real.

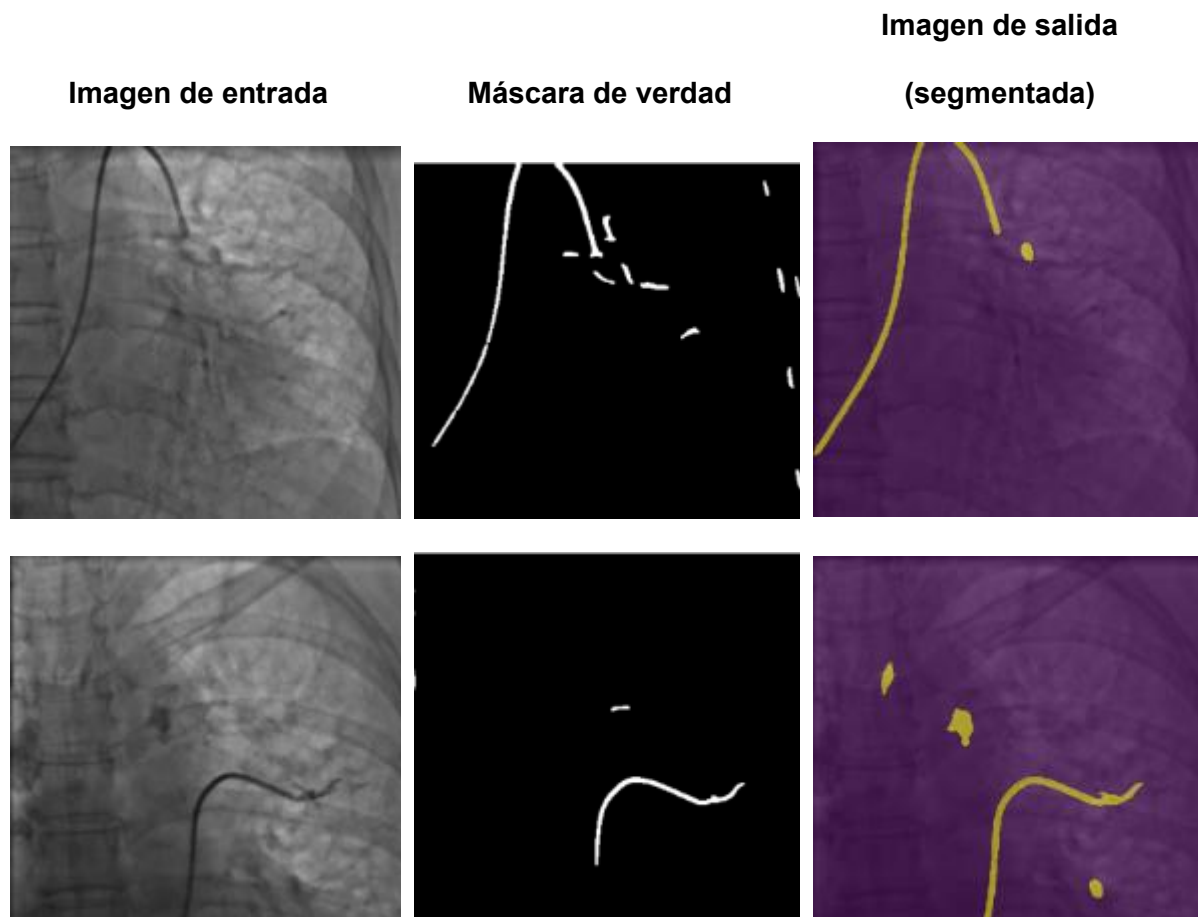
5.3.2 Resultados Cualitativos e Inspección Visual

Para verificar que la divergencia explicada anteriormente corresponde a una mejora en la calidad clínica de la imagen y no a una respuesta falsa del modelo, se procedió a realizar una inspección visual comparativa de los resultados.

En la Figura 8 se presenta el análisis comparativo entre la imagen de entrada, la pseudo-etiqueta de referencia (Sato + Otsu) y la salida final producida por el modelo híbrido. Los casos expuestos en esta ilustración fueron seleccionados de manera deliberada debido a que representan escenarios clínicos posibles. Esta tiene como objetivo evaluar el comportamiento del sistema, demostrando visualmente el porqué de la caída en las métricas perimétricas.

Figura 8

Comparación de la entrada original, la referencia o máscara de verdad y la salida del modelo híbrido.



Fuente: Elaboración propia.

El análisis morfológico permite destacar dos características en la propuesta híbrida:

1. **Restauración Topológica:** Mientras que la máscara muestra discontinuidades, el modelo híbrido logra mejorar la continuidad del flujo vascular. La red W-Net localiza la región, y SAM conecta los píxeles adyacentes basándose en la similitud visual, generando estructuras de vasos coherentes.
2. **Limpieza de Artefactos:** Se observa que el mecanismo de filtrado por superposición fue eficaz para suprimir el ruido de fondo. Elementos como vértebras, costillas y

diafragma, que comúnmente aparecen como falsos positivos, fueron descartados al no coincidir con el mapa de atención de W-Net.

Esta comparación entre la baja métrica cuantitativa y la alta calidad cualitativa respalda que el modelo entrega máscaras más limpias que las herramientas clásicas de procesamiento de imágenes utilizadas para entrenarlo.

5.3.3 Interpretabilidad y análisis de activación espacial

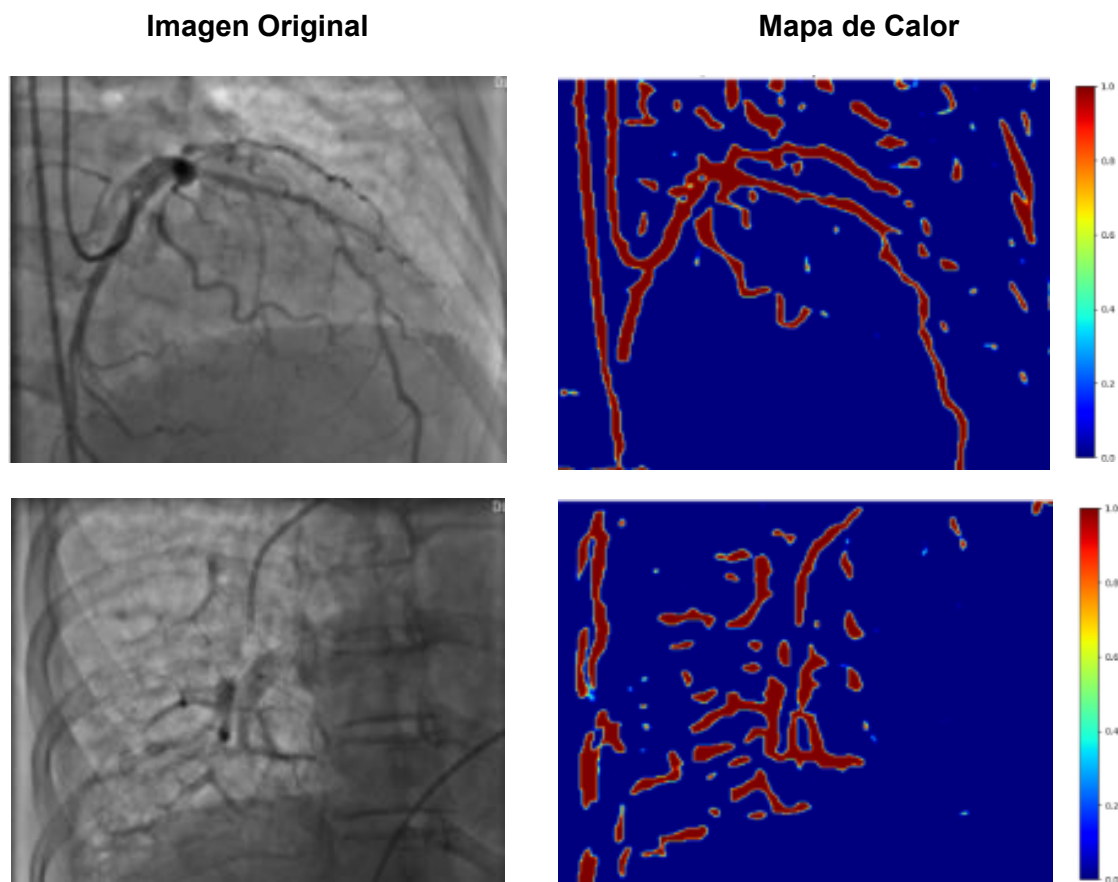
Para asegurar que el modelo W-Net actúa como la guía del sistema híbrido y basa sus decisiones en características anatómicas reales y no en sesgos del conjunto de datos se analizó el mapa de probabilidad espacial generado directamente por la salida del decoder de la red.

En el aprendizaje profundo aplicado a imágenes médicas, la inspección de mapas continuos de activación es una estrategia ampliamente utilizada para interpretar el comportamiento espacial de los modelos de segmentación. Este enfoque permite analizar la distribución de confianza píxel a píxel antes de la etapa de binarización final, lo que facilita la identificación de las regiones anatómicas en las cuales la red concentra su respuesta [53], [58].

De acuerdo con esto, se visualizaron los mapas de calor producidos por W-Net, que representan la probabilidad de pertenencia a la clase vascular. En la Figura 9 se presenta la comparación entre la angiografía original recortada y el mapa probabilístico generado por la red.

Figura 9

Comparación de la imagen original recortada y el mapa de calor generado.



Fuente: Elaboración propia.

• **Respuesta espacial:** El análisis visual evidencia una respuesta coherente con la morfología coronaria observada. Las regiones de mayor probabilidad, representadas en tonalidades cálidas, siguen principalmente la trayectoria de las arterias coronarias realzadas mediante el medio de contraste. Sin embargo, también se observan activaciones parciales sobre algunas estructuras de alta intensidad radiográfica, como tejido óseo y bordes anatómicos, lo cual es esperado debido a la similitud local de gradientes y contrastes presentes en las angiografías.

A pesar de estas activaciones secundarias, la mayor concentración probabilística permanece localizada sobre las estructuras tubulares vasculares, lo que indica que el modelo no depende exclusivamente de artefactos de borde o variaciones globales de intensidad, sino de patrones morfológicos relacionados con la continuidad y geometría vascular.

Esta evidencia respalda el uso de W-Net como un mecanismo semántico capaz de guiar la selección de regiones candidatas generadas por SAM dentro del enfoque híbrido propuesto.

6 DISCUSIÓN

La presente investigación enfrentó el desafío de la segmentación automática de la red vascular en imágenes de angiografía coronaria, un entorno que se caracteriza por la complejidad de las estructuras, por el ruido y la falta de anotaciones manuales hechas por expertos. Los resultados obtenidos demuestran que los enfoques evaluados como W-Net, SAM y el modelo híbrido presentan comportamientos bien diferenciados dependiendo del criterio de evaluación utilizado, lo que abre un debate sobre la validación en entornos no supervisados.

En términos cuantitativos, al realizar un análisis de ablación se identificó que W-Net presentó una mayor fidelidad de píxel frente a las pseudoetiquetas generadas por el enfoque heurístico, alcanzando un índice IoU de 0.81 frente al 0.28 del modelo híbrido. Sin embargo, la revisión cualitativa estructural basada en criterios anatómicos reveló que el modelo híbrido produjo segmentaciones con una continuidad superior y una eliminación consistente de posibles falsos positivos estructurales. En lugar de indicar una falla en la convergencia del algoritmo, esta contradicción numérica constituye el hallazgo principal de la fase de validación, evidenciando el fenómeno de divergencia con la referencia heurística. Dado que las etiquetas de entrenamiento no fueron generadas por expertos, representan un estándar secundario, que se caracterizaron por discontinuidades y subestimación del tamaño de los vasos.

Este comportamiento se entiende debido a que W-Net, al optimizar su pérdida directamente contra estas pseudomáscaras imperfectas, experimentó un sobreajuste al ruido, aprendiendo y replicando las interrupciones y trazos rotos del algoritmo de Otsu. Por el contrario, el modelo SAM funcionó bajo un enfoque zero-shot, utilizando características visuales que le permitieron reconstruir la morfología continua del vaso. Al generar una máscara continua y anatómicamente más coherente, el modelo híbrido dejó de coincidir geométricamente píxel a píxel con la referencia. Como se discute en la literatura, la interpretación de los índices basados en superposición espacial como DSC e IoU debe contextualizarse según la naturaleza de la anotación de referencia. Estas métricas cuantifican coincidencia posicional,

no verdad clínica absoluta [56], [57]. Por lo cual, el descenso cuantitativo del modelo híbrido no implica una peor segmentación; sino que es la confirmación de que el sistema logró desprenderse del sesgo del preprocesamiento.

La comparación entre la literatura [4], [47] y los resultados empíricos confirma que los modelos fundacionales como SAM presentan una alta capacidad de delimitación geométrica de bordes y texturas, pero carecen de la capacidad semántica para discriminar objetos de interés médico por sí solos. En las pruebas individuales, SAM segmentó estructuras irrelevantes debido a la falta de contexto médico.

La arquitectura híbrida resolvió esta limitación mediante un esquema de fusión donde "W-Net señala y SAM dibuja". La extracción de los mapas de calor muestra que W-Net enfoca sus mapas de calor probabilísticos sobre la trayectoria de las arterias coronarias perfundidas con medio de contraste. Al utilizar este mapa semántico como guía para filtrar por coherencia espacial las precisas propuestas geométricas de SAM, el modelo híbrido logró integrar la inteligencia del especialista con la precisión del generalista.

Desde la perspectiva metodológica, la adopción de CRISP-DM permitió organizar las fases de comprensión de datos, preparación, modelado y evaluación del sistema propuesto. Esto facilitó el desarrollo iterativo de la investigación, permitiendo identificar limitaciones asociadas a las pseudoetiquetas heurísticas y ajustar progresivamente la estrategia híbrida planteada. Asimismo, la calibración experimental de los parámetros y umbrales utilizados contribuyó a mejorar la estabilidad del pipeline de segmentación y a reducir la dependencia de decisiones arbitrarias durante el proceso de optimización.

A pesar de lo robusto que es la arquitectura propuesta, la investigación presenta limitaciones relevantes. La falta de anotaciones manuales realizadas por especialistas en cardiología intervencionista impidió realizar una validación supervisada. Esto obligó a depender de métricas de superposición frente a referencias heurísticas sesgadas.

Sin embargo, el modelo híbrido representa una alternativa eficiente para el diagnóstico asistido en entornos clínicos con pocos recursos y ausencia de bases de datos etiquetadas. Al resolver las fallas del entrenamiento con etiquetas imperfectas combinando la capacidad de localización de W-Net con la precisión en los bordes que ofrece SAM, se demuestra que es posible extraer información anatómica morfológicamente consistente a partir de datos ruidosos.

En conclusión, la divergencia observada entre las bajas métricas cuantitativas y la mejora de coherencia morfológica observada no resta valor al sistema, sino que demuestra la necesidad crítica de proponer nuevos paradigmas y métodos de validación adaptados al aprendizaje no supervisado en el diagnóstico por imágenes médicas.

7 TRABAJOS FUTUROS

A partir de los resultados obtenidos y de las limitaciones identificadas a lo largo de este trabajo, se abren múltiples oportunidades de investigación que permitirán fortalecer y proyectar el sistema propuesto.

En primer lugar, aunque los resultados cuantitativos evidencian un desempeño competitivo, es importante avanzar hacia una validación clínica más detallada y de muy alto nivel. Para eso, se propone como prioridad la evaluación del sistema híbrido frente a una máscara de verdad manual, elaborada por un grupo de especialistas en cardiología intervencionista. La construcción de este estándar de referencia, validado bajo criterios clínicos, permitirá comparar las segmentaciones generadas por la arquitectura con el verdadero estándar de oro médico. Este paso no solo aporta mayor solidez metodológica, sino que también facilitará una interpretación realista en la práctica clínica, donde la precisión anatómica y la coherencia diagnóstica son obligatorias.

Para mejorar el rendimiento del algoritmo híbrido, se plantea la línea de desarrollo con Adaptación de Bajo Rango (LoRA). En este trabajo, el modelo SAM opera bajo un paradigma zero-shot, lo que limita su precisión. La integración de LoRA permitiría implementarlo bajo la técnica de fine-tuning. Esta busca especializar las capacidades de delineación geométrica del modelo generalista en las propiedades físicas del dominio de la angiografía coronaria, reduciendo significativamente la dependencia de umbrales heurísticos y mejorando la recuperación de vasos distales finos [59].

De igual manera, se considera necesario la realización de estudios en diferentes entornos clínicos que incorporen imágenes provenientes de diferentes instituciones y equipos de angiografía. La ampliación de los datos permitirá evaluar la capacidad de generalización del modelo ante variaciones del entorno real, tales como diferencias en protocolos de adquisición, niveles de medios de contraste, calidad de imagen, resoluciones espaciales y características propias de cada fabricante. Esta validación externa es clave para garantizar la

reproducibilidad de los resultados y para analizar la sensibilidad del sistema frente a diferentes condiciones.

Adicionalmente, sería correcto explorar estrategias de adaptación que permitan optimizar el desempeño del modelo cuando se enfrente a datos con distribuciones distintas a las del conjunto de entrenamiento original. En entornos clínicos reales, la variabilidad es una constante, por lo que integrar mecanismos que favorezcan la estabilidad del sistema ante cambios en la calidad o morfología de los vasos contribuirá significativamente a su aplicabilidad práctica.

En definitiva, estas líneas de trabajo futuro buscan elevar el rendimiento del sistema, también garantizar su validez, reproducibilidad y utilidad en escenarios reales, características necesarias para su posible entrada en la práctica médica.

8 CONCLUSIONES

Esta investigación abordó uno de los desafíos más complejos en el análisis de imágenes médicas, el cual es la segmentación automática de estructuras anatómicas como lo es la red vascular coronaria en estudios de angiografías generados por rayos X, que está muy condicionado por la superposición de estructuras, la presencia de ruido y la falta crítica de bases de datos con anotaciones de expertos en la materia. La principal conclusión de este trabajo es que las limitaciones propias del aprendizaje no supervisado pueden superarse mediante el diseño de arquitecturas híbridas que integren el conocimiento semántico específico del dominio con la capacidad de generalización morfológica de los modelos fundacionales. Desde el diseño de la investigación, la implementación de la metodología CRISP-DM y el ajuste experimental de los hiperparámetros como el umbral de limpieza morfológica de 50 píxeles y el criterio de coherencia espacial del 35%, demostraron que es posible construir un flujo de trabajo confiable. El sistema desarrollado logró generar su propio marco de entrenamiento a partir de heurísticas, reduciendo la dependencia de un amplio etiquetado manual durante la fase de desarrollo, aunque reconociendo que la anotación experta sigue siendo indispensable para protocolos de validación clínica convencional.

Los resultados cuantitativos evidencian un desempeño diferenciado, en donde el modelo W-Net validó su funcionamiento como un especialista, alcanzando un Coeficiente Dice (DSC) de 0.9006 en validación. Este rendimiento se logró estableciendo la tarea de segmentación con una ponderación de pérdida del 0.85 frente al 0.15 de reconstrucción. El modelo Segment Anything Model (SAM), trabajando en modo zero-shot, mostró una elevada precisión geométrica de bordes, reflejada en un valor promedio de predicted IoU de 0.9358 sobre las máscaras aceptadas tras el filtrado heurístico, aun sin tener la especificidad necesaria para discriminar estructuras de forma autónoma.

El descubrimiento científico más importante de esta investigación consiste en la caracterización de la divergencia con la referencia heurística. El estudio demuestra empíricamente que, en entornos no supervisados, la evaluación del rendimiento no debe

recaer de manera exclusiva en métricas cuantitativas de superposición estricta. Se comprobó que el modelo W-Net, al ser entrenado contra pseudoetiquetas ruidosas, logró altos valores en métricas de superposición, comprometiendo la capacidad de generalización debido al sobreajuste ante errores y discontinuidades. Por otro lado, el modelo híbrido, a pesar de tener un descenso en las métricas (DSC entre 0.4303 y 0.4601; IoU entre 0.2741 y 0.2987) logró salir del sesgo del preprocesamiento. Esta contradicción evidencia que el sistema híbrido dejó de replicar estrictamente las limitaciones morfológicas presentes en las pseudoetiquetas heurísticas utilizadas durante la evaluación. A nivel de arquitectura, los resultados evidenciaron que la integración entre ambos paradigmas permitió complementar las limitaciones individuales de cada modelo. El algoritmo híbrido logró integrar ambas características, corrigiendo la imprecisión de bordes del codificador convolucional. Finalmente, las implicaciones de este trabajo establecen un precedente para la asistencia en entornos clínicos con recursos limitados. Se ha estructurado y validado un pipeline computacional capaz de segmentar estructuras vasculares complejas de forma automatizada a partir de referencias heurísticas. Las segmentaciones producidas por este sistema híbrido abren la puerta al desarrollo de repositorios médicos, los cuales podrán utilizarse en el futuro para entrenar arquitecturas de aprendizaje semisupervisado. De este modo, esta investigación no solo proporciona una alternativa de solución algorítmica a un problema de visión por computadora, sino que sienta bases para optimizar la implementación de inteligencia artificial en las salas de hemodinamia y la práctica cardiológica en tiempo real.

9 BIBLIOGRAFIA

- [1] J. El-Taraboulsi, C. P. Cabrera, C. Roney, and N. Aung, "Deep neural network architectures for cardiac image segmentation," *Artificial Intelligence in the Life Sciences*, vol. 4, p. 100083, Dec. 2023, doi: 10.1016/j.aills.2023.100083.
- [2] C. Bian *et al.*, "Uncertainty-aware domain alignment for anatomical structure segmentation," *Med. Image Anal.*, vol. 64, p. 101732, Aug. 2020, doi: 10.1016/j.media.2020.101732.
- [3] X. Xia and B. Kulis, "W-Net: A Deep Model for Fully Unsupervised Image Segmentation," Nov. 2017.
- [4] A. Medellin, D. Grabowsky, D. Mikulski, and R. Langari, "Sam-Sam - A Novel Approach to Hyperspectral Image Semantic Segmentation," in *2023 13th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, IEEE, Oct. 2023, pp. 1–5. doi: 10.1109/WHISPERS61460.2023.10431253.
- [5] W. Chen *et al.*, "Prostate Segmentation using 2D Bridged U-net," in *2019 International Joint Conference on Neural Networks (IJCNN)*, IEEE, Jul. 2019, pp. 1–7. doi: 10.1109/IJCNN.2019.8851908.
- [6] Y. Zhang, Z. Shen, and R. Jiao, "Segment anything model for medical image segmentation: Current applications and future directions," *Comput. Biol. Med.*, vol. 171, p. 108238, Mar. 2024, doi: 10.1016/j.combiomed.2024.108238.
- [7] T.-Y. Hsia, E. A. Peck, and J. V. Conte, "Coronary Artery Disease," in *The Johns Hopkins Manual of Cardiac Surgical Care*, Elsevier, 2008, pp. 61–93. doi: 10.1016/B978-0-323-01810-4.10003-3.
- [8] B. A. Stark *et al.*, "Global, Regional, and National Burden of Cardiovascular Diseases and Risk Factors in 204 Countries and Territories, 1990-2023," *JACC*, Sep. 2025, doi: 10.1016/j.jacc.2025.08.015.
- [9] S. Lee, M. Lee, and M. Kang, "Poisson–Gaussian Noise Analysis and Estimation for Low-Dose X-ray Images in the NSCT Domain," *Sensors*, vol. 18, no. 4, p. 1019, Mar. 2018, doi: 10.3390/s18041019.
- [10] LUQMAN and FAKHRUD DIN, "A Hyper-Heuristic Based Strategy for Image Segmentation Using Multilevel Thresholding," *Journal of Quality Measurement and Analysis*, vol. 21, no. 2, pp. 123–152, Jun. 2025, doi: 10.17576/jqma.2102.2025.10.
- [11] Y. Sato *et al.*, "Tissue classification based on 3D local intensity structures for volume rendering," *IEEE Trans. Vis. Comput. Graph.*, vol. 6, no. 2, pp. 160–180, 2000, doi: 10.1109/2945.856997.
- [12] T. Jerman, F. Pernus, B. Likar, and Z. Spiclin, "Enhancement of Vascular Structures in 3D and 2D Angiographic Images," *IEEE Trans. Med. Imaging*, vol. 35, no. 9, pp. 2107–2118, Sep. 2016, doi: 10.1109/TMI.2016.2550102.
- [13] D. Indra, T. Hasanuddin, R. Satra, and N. R. Wibowo, "Eggs Detection Using Otsu Thresholding Method," in *2018 2nd East Indonesia Conference on Computer and*

- Information Technology (EIConCIT)*, IEEE, Nov. 2018, pp. 10–13. doi: 10.1109/EIConCIT.2018.8878517.
- [14] Chen Yu, Chen Dian-ren, Li Yang, and Chen Lei, “Otsu’s thresholding method based on gray level-gradient two-dimensional histogram,” in *2010 2nd International Asia Conference on Informatics in Control, Automation and Robotics (CAR 2010)*, IEEE, Mar. 2010, pp. 282–285. doi: 10.1109/CAR.2010.5456687.
 - [15] D. Shi, R. Liu, L. Tao, and C. Yuan, “Heuristic Dropout: An Efficient Regularization Method for Medical Image Segmentation Models,” in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, May 2022, pp. 1101–1105. doi: 10.1109/ICASSP43922.2022.9747409.
 - [16] X. Wang, T. Duan, G. Ouyang, W. Hao, L. Mu, and X. Zhang, “Comprehensive review of segment anything model across multiple domains,” *Digit. Signal Process.*, vol. 167, p. 105459, Dec. 2025, doi: 10.1016/j.dsp.2025.105459.
 - [17] Z. Wang, Z. Hou, S. Hou, and D. Cao, “Generalizable Digital Rock Image Segmentation under Limited Data with the Segment Anything Model,” *Intelligent Geoengineering*, Dec. 2025, doi: 10.1016/j.ige.2025.12.001.
 - [18] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, “Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2019, pp. 658–666. doi: 10.1109/CVPR.2019.00075.
 - [19] Y. Song *et al.*, “Transformer framework for depth-assisted UDA semantic segmentation,” *Eng. Appl. Artif. Intell.*, vol. 137, p. 109206, Nov. 2024, doi: 10.1016/j.engappai.2024.109206.
 - [20] W. Dong, Z. Liang, L. Wang, G. Tian, and Q. Long, “Unsupervised domain adaptation segmentation algorithm with cross-domain data augmentation and category contrast,” *Neurocomputing*, vol. 623, p. 129393, Mar. 2025, doi: 10.1016/j.neucom.2025.129393.
 - [21] S. Li, X. Zhao, and J. Cao, “An UDA bearing fault diagnosis method based on synergistic optimization of unbalanced data,” *Mech. Syst. Signal Process.*, vol. 243, p. 113700, Jan. 2026, doi: 10.1016/j.ymssp.2025.113700.
 - [22] S. Soobia.et.al., “Analysis of Software Development Methodologies,” *International Journal of Computing and Digital Systems*, vol. 8, no. 5, pp. 445–460, Jan. 2019, doi: 10.12785/ijcds/080502.
 - [23] C. Schröer, F. Kruse, and J. M. Gómez, “A Systematic Literature Review on Applying CRISP-DM Process Model,” *Procedia Comput. Sci.*, vol. 181, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.
 - [24] Q. Yang, H. Yi, L. Yi, M. Liu, and X. Chen, “Accurate segmentation and labeling of coronary artery segments in X-ray angiography with an improved UNet-based cGAN architecture,” *Biomed. Signal Process. Control*, vol. 112, p. 108812, Feb. 2026, doi: 10.1016/j.bspc.2025.108812.

- [25] H. He, A. Banerjee, R. P. Choudhury, and V. Grau, "Deep learning based coronary vessels segmentation in X-ray angiography using temporal information," *Med. Image Anal.*, vol. 102, p. 103496, May 2025, doi: 10.1016/j.media.2025.103496.
- [26] D. Lesage, E. D. Angelini, I. Bloch, and G. Funka-Lea, "A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes," *Med. Image Anal.*, vol. 13, no. 6, pp. 819–845, Dec. 2009, doi: 10.1016/j.media.2009.07.011.
- [27] H. Xiang, Q. Zou, M. A. Nawaz, X. Huang, F. Zhang, and H. Yu, "Deep learning for image inpainting: A survey," *Pattern Recognit.*, vol. 134, p. 109046, Feb. 2023, doi: 10.1016/j.patcog.2022.109046.
- [28] J. Yang and N. I. R. Ruhaiyem, "Review of Deep Learning-Based Image Inpainting Techniques," *IEEE Access*, vol. 12, pp. 138441–138482, 2024, doi: 10.1109/ACCESS.2024.3461782.
- [29] P. M. Szczypiński, A. Klepaczko, and R. Olbrycht, "VesselKnife – software for the analysis of tubular structures in biomedical images," *Comput. Methods Programs Biomed.*, vol. 268, p. 108817, Aug. 2025, doi: 10.1016/j.cmpb.2025.108817.
- [30] A. Arino, N. Yamamoto, K. Sasaki, R. Kimura, and T. Aizawa, "Quantitative evaluation of fatty change in rotator cuff muscles using the Otsu thresholding method," *J. Shoulder Elbow Surg.*, vol. 35, no. 1, pp. 200–208, Jan. 2026, doi: 10.1016/j.jse.2025.04.026.
- [31] J. Huang *et al.*, "Deep learning for automated mandibular canal segmentation in CBCT scans," *BMC Oral Health*, vol. 25, no. 1, p. 1699, Oct. 2025, doi: 10.1186/s12903-025-07098-5.
- [32] H. RostomMohamed and A. A. Mohammed Baqer Al-Shammaa, "Contour with Connected Components and Mesh Surface for Skin Image Segmentation," *Int. J. Appl. Inf. Syst.*, vol. 6, no. 2, pp. 48–52, Sep. 2013, doi: 10.5120/ijais13-451023.
- [33] Y. Wu, Q. Xu, Z. Zou, and H. Zhang, "Conditional generative model with skip-connection structure for low-light image enhancement," *Digit. Signal Process.*, vol. 151, p. 104556, Aug. 2024, doi: 10.1016/j.dsp.2024.104556.
- [34] B. Walter, "Analysis of convolutional neural network image classifiers in a hierarchical max-pooling model with additional local pooling," *J. Stat. Plan. Inference*, vol. 224, pp. 109–126, May 2023, doi: 10.1016/j.jspi.2022.11.001.
- [35] W. Zhang *et al.*, "Focal difficult-to-predict pixels dice loss for mitigating data imbalance in medical image segmentation," *Expert Syst. Appl.*, vol. 290, p. 128518, Sep. 2025, doi: 10.1016/j.eswa.2025.128518.
- [36] Q. Li *et al.*, "Low-dose computed tomography image reconstruction via a multistage convolutional neural network with autoencoder perceptual loss network," *Quant. Imaging Med. Surg.*, vol. 12, no. 3, pp. 1929–1957, Mar. 2022, doi: 10.21037/qims-21-465.
- [37] S. Shihui, W. Yilin, S. Xiaofeng, C. Huilin, and W. Biao, "Intelligent prediction of favorable shale oil areas based on deep learning and batch normalization,"

- Geoenergy Science and Engineering*, vol. 252, p. 213955, Sep. 2025, doi: 10.1016/j.geoen.2025.213955.
- [38] Y. Xu and H. Zhang, "Convergence of deep ReLU networks," *Neurocomputing*, vol. 571, p. 127174, Feb. 2024, doi: 10.1016/j.neucom.2023.127174.
 - [39] J. Li, Y. Cheng, Y. Lu, Z. Xia, Y. Mo, and G. Huang, "From ReLU to GeMU: Activation functions in the lens of cone projection," *Neural Networks*, vol. 190, p. 107654, Oct. 2025, doi: 10.1016/j.neunet.2025.107654.
 - [40] H. Salem, A. E. Kabeel, E. M. S. El-Said, and O. M. Elzeki, "Predictive modelling for solar power-driven hybrid desalination system using artificial neural network regression with Adam optimization," *Desalination*, vol. 522, p. 115411, Jan. 2022, doi: 10.1016/j.desal.2021.115411.
 - [41] Z. Khan, H. Liu, Y. Shen, and X. Zeng, "Deep learning improved YOLOv8 algorithm: Real-time precise instance segmentation of crown region orchard canopies in natural environment," *Comput. Electron. Agric.*, vol. 224, p. 109168, Sep. 2024, doi: 10.1016/j.compag.2024.109168.
 - [42] F. Binsar and Sasmoko, "Integrating chain of thought and explainable AI in BERT-based deep learning for interpretable medical diagnosis," *Procedia Comput. Sci.*, vol. 269, pp. 389–399, 2025, doi: 10.1016/j.procs.2025.08.291.
 - [43] Y. Zhuo, Z. Xu, D. Zhou, P. Xu, and Y. Tian, "RankSAM: Lightweight adapters and prompt generation in zero-shot semantic segmentation," *Neurocomputing*, vol. 670, p. 132594, Mar. 2026, doi: 10.1016/j.neucom.2025.132594.
 - [44] M. Marjani, M. Mahdianpari, D. J. Varon, and F. Mohammadimanesh, "The integration of vision transformers and SAM for automated methane super-emitter detection using TROPOMI data," *J. Environ. Manage.*, vol. 393, p. 127034, Oct. 2025, doi: 10.1016/j.jenvman.2025.127034.
 - [45] A. Carraro, M. Sozzi, and F. Marinello, "The Segment Anything Model (SAM) for accelerating the smart farming revolution," *Smart Agricultural Technology*, vol. 6, p. 100367, Dec. 2023, doi: 10.1016/j.atech.2023.100367.
 - [46] Y. Huang *et al.*, "Segment anything model for medical images?," *Med. Image Anal.*, vol. 92, p. 103061, Feb. 2024, doi: 10.1016/j.media.2023.103061.
 - [47] J. Wu *et al.*, "Medical SAM adapter: Adapting segment anything model for medical image segmentation," *Med. Image Anal.*, vol. 102, p. 103547, May 2025, doi: 10.1016/j.media.2025.103547.
 - [48] *2014 International Conference on Computer Vision Theory and Applications*. IEEE, 2014.
 - [49] K. S. Ni and T. Q. Nguyen, "Adaptable K-nearest neighbor for image interpolation," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, Mar. 2008, pp. 1297–1300. doi: 10.1109/ICASSP.2008.4517855.
 - [50] G. Tziafas and H. Kasaei, "Early or Late Fusion Matters: Efficient RGB-D Fusion in Vision Transformers for 3D Object Recognition," in *2023 IEEE/RSJ International*

- Conference on Intelligent Robots and Systems (IROS)*, IEEE, Oct. 2023, pp. 9558–9565. doi: 10.1109/IROS55552.2023.10341422.
- [51] K. Gadzicki, R. Khamsehashari, and C. Zetsche, “Early vs Late Fusion in Multimodal Convolutional Neural Networks,” in *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*, IEEE, Jul. 2020, pp. 1–6. doi: 10.23919/FUSION45008.2020.9190246.
 - [52] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.
 - [53] S. He, Y. Feng, P. E. Grant, and Y. Ou, “Segmentation ability map: Interpret deep features for medical image segmentation,” *Med. Image Anal.*, vol. 84, p. 102726, Feb. 2023, doi: 10.1016/j.media.2022.102726.
 - [54] M. Alidoost *et al.*, “Model utility of a deep learning-based segmentation is not Dice coefficient dependent: A case study in volumetric brain blood vessel segmentation,” *Intell. Based. Med.*, vol. 7, p. 100092, 2023, doi: 10.1016/j.ibmed.2023.100092.
 - [55] K. V. Houde, P. M. Kamble, and R. S. Hegadi, “A Comparative Analysis of IoU-Based Losses with YOLOv8: Fruit Tree Detection from UAV Images,” *Procedia Comput. Sci.*, vol. 260, pp. 898–905, 2025, doi: 10.1016/j.procs.2025.03.272.
 - [56] T. Rädtsch *et al.*, “Quality Assured: Rethinking Annotation Strategies in Imaging AI,” Jul. 2024.
 - [57] A. A. Taha and A. Hanbury, “Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool,” *BMC Med. Imaging*, vol. 15, no. 1, p. 29, Dec. 2015, doi: 10.1186/s12880-015-0068-x.
 - [58] S. Ghosal and P. Shah, “A deep-learning toolkit for visualization and interpretation of segmented medical images,” *Cell Reports Methods*, vol. 1, no. 7, p. 100107, Nov. 2021, doi: 10.1016/j.crmeth.2021.100107.
 - [59] Z. Li, D. Huo, M. Meurer, M. Panesso, W.-G. Drossel, and T. Bergs, “Enhancing Tool Wear Segmentation with LoRA-SAM and Point Prompts,” *Procedia CIRP*, vol. 134, pp. 705–710, 2025, doi: 10.1016/j.procir.2025.02.184.

10 ANEXOS

10.1 Anexo 1.

Carta de autorización para el uso de imágenes de angiografía coronaria emitida por Corazón y Aorta S.A.S.


Corazón y Aorta
Más vida

Guadalajara de Buga, 21 de Octubre de 2025

Corazón y Aorta S.A.S
Dirección científica

Asunto: Autorización para uso de imágenes de arteriografías coronarias con fines académicos.

Por medio de la presente, la **Dirección Científica de Corazón y Aorta S.A.S.** autoriza al **Ingeniero Biomédico Nicolás Ramírez Millán** con documento de identificación **CC. 1107526653**, identificado como miembro de esta institución y **candidato a Magister en Ingeniería Biomédica**, para utilizar un número limitado de **imágenes de arteriografías coronarias** obtenidas en nuestra institución, con el único fin de desarrollar su proyecto de tesis de maestría titulado:

"Comparación de algoritmos de segmentación heurística de imágenes de angiografía coronaria utilizando técnicas no supervisadas."

La presente autorización se concede exclusivamente con fines académicos y científicos, sin implicaciones comerciales, y bajo las siguientes condiciones:

1. Las imágenes deberán ser **anonimizadas** previamente, garantizando la protección total de la información clínica y la confidencialidad de los pacientes.
2. El uso de las imágenes será **únicamente para el desarrollo del proyecto académico**, sin posibilidad de cesión o reproducción sin previa autorización institucional.
3. Las imágenes deberán almacenarse en entornos **seguros y encriptados**, conforme a las políticas de protección de datos de la institución.
4. Cualquier divulgación de resultados se deberá realizar **sin incluir información identificable** y mencionando que los datos fueron obtenidos con la autorización de Corazón y Aorta S.A.S.

Servicios: Cirugía Cardiovascular, Cirugía Endovascular.

• Cardiología no invasiva: Consulta especializada, Ecocardiografía, Prueba de esfuerzo, MAPA, HOLTER, EKG, Rehabilitación Cardíaca.	• Angiografía: Hemodinámica, Electrofisiología, Neuroradiología, Vascular Periférico, Radiología Intervencionista.
--	---

Esta autorización se expide a solicitud del interesado, para los fines académicos descritos.

Atentamente,



DR. LUIS FELIPE MEDINA GALLO
DIRECTOR CIENTIFICO
CORAZON Y AORTA S.A.S
NIT 900.051.107-2



DRA. MARIA CONSUELO NAVARRO
DIRECTORA MÉDICA
CORAZON Y AORTA S.A.S
NIT 900.051.107-2

Servicios: Cirugía Cardiovascular, Cirugía Endovascular.

• **Cardiología no invasiva:** Consulta especializada, Ecocardiografía, Prueba de esfuerzo, MAPA, HOLTER, EKG, Rehabilitación Cardíaca.

• **Angiografía:** Hemodinámica, Electrofisiología, Neuroradiología, Vascular Periférico, Radiología Intervencionista.