



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Informática

Validación de técnicas de Deep Learning (DL) para la  
detección de lesiones displásicas en imágenes médicas de  
pacientes con enfermedad inflamatoria intestinal (EII)

Trabajo Fin de Grado

Grado en Ingeniería Informática

AUTOR/A: Ferrer Andreu, Pablo Lorenzo

Tutor/a: Andreu García, Gabriela

Director/a Experimental: Martínez Peiró, Joaquín

CURSO ACADÉMICO: 2023/2024

# Resumen

---

Este proyecto de Trabajo de Fin de Grado se enfoca en el desarrollo de un *software* basado en redes neuronales para la detección de lesiones displásicas en imágenes médicas de pacientes con enfermedad inflamatoria intestinal (EII). La colaboración con el Hospital Universitario y Politécnico La Fe de Valencia respalda este proyecto, proporcionando una perspectiva clínica que garantiza la aplicabilidad de los resultados en entornos médicos reales. El *software* busca mejorar la precisión y eficacia en el diagnóstico, priorizando la identificación de pólipos con riesgo de malignidad, lo que contribuye a una atención médica gastrointestinal más precisa y orientada a resultados.

# Abstract

---

This Final Degree Project (TFG) focuses on developing a neural network-based *software* for real-time detection of polyps during colonoscopies. Collaboration with the University and Polytechnic Hospital La Fe in Valencia supports this project, providing a clinical perspective that ensures the applicability of the results in real medical settings. The *software* aims to improve the accuracy and effectiveness of diagnosis, prioritizing the identification of polyps with malignant potential, thus contributing to more precise and results-oriented gastrointestinal medical care.

# Índice

---

1.	Introducción .....	8
1.1.	Motivación .....	9
1.2.	Objetivos del Proyecto.....	10
1.3.	Colaboradores del Proyecto .....	10
2.	Estado del arte y punto de partida.....	12
2.1.	Estado del arte .....	12
2.2.	Proyectos y sistemas comerciales actuales.....	12
2.3.	Investigaciones académicas.....	12
2.4.	Conclusiones .....	13
3.	Marco teórico.....	14
3.1.	Inteligencia artificial.....	14
3.1.1.	Definición de Inteligencia Artificial.....	14
3.1.2.	Aplicaciones de la Inteligencia Artificial en detección de pólipos .....	14
3.2.	Machine Learning .....	15
3.2.1.	Definición de Machine Learning .....	15
3.2.2.	Aplicaciones del Machine Learning en detección de pólipos .....	16
3.3.	Deep Learning .....	16
3.3.1.	Redes neuronales en el contexto de Deep Learning.....	17
3.4.	Redes Neuronales.....	17
3.4.1.	Redes Neuronales Artificiales (ANNs) .....	18
3.4.2.	Redes Neuronales Convolucionales (CNNs).....	19
3.4.3.	Transfer Learning.....	20
3.5.	YOLO .....	21
3.5.1.	Beneficios de YOLO .....	22
3.5.2.	Funcionamiento de YOLO.....	23
3.5.3.	Arquitectura de YOLO .....	23
3.6.	Métricas de rendimiento .....	24
4.	Marco práctico.....	26
4.1.	Entorno de desarrollo .....	26
4.2.	Preparación de las imágenes: Dataset.....	26
4.2.1.	Conjunto de entrenamiento, validación y test.....	29
4.3.	Desarrollo de la CNN y evaluación del modelo .....	31

4.3.1.	Entrenamiento de la red .....	32
4.4.	Hiperparámetros.....	34
4.4.1.	Ajuste de hiperparámetros de Ultralytics Yolo .....	35
4.5.	Validación cruzada ( <i>Cross validation</i> ).....	37
5.	Resultados .....	39
5.1.	Comparación de resultados .....	39
5.1.1.	Comparación de resultados de Cross Validation .....	39
5.1.2.	Comparación de resultados de ajuste de hiperparámetros .....	40
5.1.3.	Resultados en el conjunto de datos de prueba .....	40
5.2.	Análisis de resultados según la clase.....	43
5.3.	Últimas pruebas .....	45
6.	Conclusiones .....	47
6.1.	Trabajo futuro .....	47
	Bibliografía .....	49
	Anexo .....	51
	Anexo 1: Objetivos de Desarrollo Sostenible.....	51
	Anexo 2: Resultados Cross Validation SPLIT 1-7 .....	52



# Índice de figuras

---

Figura 1: Proceso de *Machine Learning*

Figura 2: Proceso de *Deep Learning*

Figura 3: Representación de Red Neuronal Artificial

Figura 4: Representación de Red Neuronal Convolutacional

Figura 5: Proceso de *Transfer Learning*

Figura 6: Algunos ejemplos de la imágenes de pólipos recopiladas de EndoTect[19], Kvasir[2], HyperKvasir[21] y CVC[22] para el entrenamiento de las CNN's y que configuran el DataTr-1 lo con un total de 2108 , se muestran tanto la original como la segmentada.

Figura 7: Algunos ejemplos de la imágenes de pólipos recopiladas de ENDOCV[22] y ENDOCV2[22] para el entrenamiento de las CNN's y estas imágenes junto con las del DataTr1, configuran el DatTr2 formado por un total de 2800 imágenes, se muestran tanto la original como la segmentada.

Figura 8: Ejemplo del procesamiento realizado en las imágenes de dataset propio, recopilas por los sanitarios especialista de la FE y que conforman el DatasetFe1, imagen original con pólipo, del Dataset\_Fe1, imagen con pólipo marcado por los sanitarios especializados y pólipo con su *Bounding Box*

Figura 9: Ejemplo de las imágenes de *dataset* propio, recopilas por los sanitarios especialista de la FE y que conforman el DatasetFe1.

Figura 10: Ejemplo de las imágenes de *dataset* propio, recopilas por los sanitarios especialista de la FE y que conforman el DatasetFe1, con el etiquetado proporcionado.

Figura 11: Ejemplos de detecciones de la red durante el entrenamiento de 250 épocas con DataTr1

Figura 12: Métricas de rendimiento del entrenamiento de 250 épocas con DataTr1

Figura 13: Ejemplos de detecciones de la red durante el entrenamiento de 500 épocas con DataTr1

Figura 14: Métricas de rendimiento del entrenamiento de 500 épocas con DataTr1

Figura 15: Hiperparámetros seleccionados por la función *Tune* para el mejor resultado del entrenamiento

Figura 16: Representación de la validación cruzada con  $k=5$

Figura 17: Resultados de predicción con red con sobreajuste de hiperparámetros

Figura 18: Ejemplo de predicciones de la red entrenada

Figura 19: Detecciones de pólipos de tipo adenoma

Figura 20: Detecciones de pólipos de tipo pseudopólipo

Figura 21: Detecciones de pólipos de tipo serrado

Figura 22: Ejemplo de pólipos detectados en la fase de predicción

Figura 23: Ejemplo de pólipos no detectados o detectados parcialmente en la fase de predicción

# Índice de tablas

---

- Tabla 1: Distribución de imágenes e instancias separadas por clases del conjunto *DatasetFe1*
- Tabla 2: Hiperparámetros por defecto que YOLO utiliza para entrenar
- Tabla 3: Resultados de entrenamiento con *Cross Validation del SPLIT 1 al 7*
- Tabla 4: Resultado de entrenamiento con 250 y 500 épocas con el *DataTr1* y *DataTr2*
- Tabla 5: Tabla de resultados de la fase de entrenamiento y validación con *DataTr1* y *DataTr2*
- Tabla 6: Tabla de resultados de la fase de predicción y test con el *DatasetFe1*
- Tabla 7: Tabla de resultados de la fase de predicción y test con el *DatasetFe1* por clases
- Tabla 8: Predicción por clases y métricas de rendimiento, obtenida con una red con  $mAP50=0,966$ ,  $mAP75 = 0,903$  y  $R = 0,926$
- Tabla 9: resultados de la fase de predicción sobre *DataSetFe1* con la reducción de 19 imágenes
- Tabla 10: resultados de la fase de predicción sobre *DataSetFe1* sin la reducción de 19 imágenes
- Tabla 11: Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS).



# 1. Introducción

---

La colonoscopia desempeña un papel fundamental en la detección precoz de enfermedades gastrointestinales, siendo esencial en la prevención y tratamiento del cáncer de colon. La identificación temprana de pólipos en este procedimiento, especialmente aquellos con características asociadas a displasia, ya que son los más peligrosos y pueden derivar en cáncer, es un factor crítico para intervenir antes de que sigan evolucionando. Esto reduce el riesgo de enfermedades graves y promueve una mejor salud y bienestar para los pacientes, contribuyendo así al Objetivo de Desarrollo Sostenible número 3 de salud y bienestar.

Este Trabajo de Fin de Grado aspira a la creación de un *software* basado en redes neuronales destinado a la detección de pólipos durante colonoscopias. La envergadura de este proyecto se ve respaldada por una colaboración con el Hospital Universitario y Politécnico La Fe de Valencia, una institución médica de renombre comprometida con la innovación en el ámbito de la salud.

La motivación central de esta iniciativa de investigación y desarrollo radica en la necesidad de un enfoque más eficiente en la detección de pólipos, especialmente aquellos con predisposición a la displasia. La capacidad de identificar de manera automática y en tiempo real estas anomalías durante las colonoscopias representan un avance significativo en la prevención y tratamiento de enfermedades gastrointestinales, disminuyendo sustancialmente el riesgo de desarrollo de cáncer de colon.

La colaboración con el Hospital La Fe no solo aporta una perspectiva práctica y clínica a este proyecto, sino que también garantiza la aplicación efectiva de los resultados en el entorno médico real. La integración de la tecnología de redes neuronales en la práctica diaria de la medicina promete mejorar la eficacia de los procedimientos endoscópicos y elevar la calidad general de la atención brindada a los pacientes.

Adicionalmente, este proyecto se plantea como un recurso de apoyo para la toma de decisiones médicas. Aunque el médico siempre tendrá la última palabra, este *software* tiene la finalidad de aumentar la efectividad en la identificación y localización de pólipos con displasia. La combinación de la experiencia humana con la precisión y rapidez que aporta la tecnología basada en redes neuronales pretende proporcionar al médico una herramienta valiosa en su diagnóstico.

A lo largo de este trabajo, se explorarán y abordarán aspectos cruciales relacionados con el diseño, desarrollo y evaluación del *software* propuesto. Además, se llevará a cabo una profunda discusión sobre sus posibles aplicaciones clínicas. Este proyecto, al aumentar la efectividad de la identificación de pólipos con displasia, propone un paso significativo hacia un diagnóstico médico más preciso y orientado a mejorar los resultados, respaldando así la toma de decisiones clínicas.

## 1.1. Motivación

---

La motivación de este proyecto tiene que ver con la necesidad de mejorar las prácticas de detección y tratamiento de enfermedades gastrointestinales, con un enfoque particular en el cáncer de colon. Según los estudios médicos, entre ellos, el de *'American College of Gastroenterology'* la identificación temprana de pólipos, especialmente aquellos con potencial displasia, sigue siendo un desafío crucial, pero es vital y puede ser crucial a la hora de ser tratado [1]. La relevancia de esta tarea se ve reflejada en la magnitud del impacto que puede tener en la vida de los pacientes y en la eficacia general de los programas de prevención.

La colonoscopia ha demostrado ser eficiente en la identificación de pólipos, pero la carga de trabajo asociada a la interpretación visual y manual por parte de los profesionales médicos es considerable. Este proyecto encuentra su motivación en el deseo de optimizar este proceso, proporcionando a los médicos una herramienta avanzada que no sólo acelere la identificación de pólipos, sino que también mejore la precisión, especialmente en aquellos casos con características indicativas de displasia.

La incidencia del cáncer de colon y su relación con la presencia de pólipos destacan la necesidad de una detección temprana y precisa. El desarrollo de un *software* basado en redes neuronales, capaz de ser usado durante las colonoscopias, representa una oportunidad significativa para abordar esta necesidad apremiante. La colaboración con el Hospital La Fe de Valencia proporciona un contexto clínico y una perspectiva práctica que refuerzan la relevancia y aplicabilidad directa de este proyecto en la atención médica cotidiana.

Además, la motivación se sustenta en la comprensión de que, si bien el médico siempre será el último responsable en la toma de decisiones, dotarlos de herramientas más precisas y ágiles puede aumentar sustancialmente la eficacia en la identificación de pólipos con displasia. Este enfoque colaborativo entre la informática y la medicina se muestra como un paso adelante en la convergencia de la tecnología y la atención médica, con el potencial de transformar positivamente la detección y prevención de enfermedades.

Por otro lado, mi motivación personal para participar en este proyecto tan importante junto con el *Hospital La Fe de Valencia* radica en el impacto positivo que puede tener en la detección y tratamiento de enfermedades gastrointestinales, particularmente en la identificación de pólipos con displasia. A pesar de la dificultad y la extensión del proceso, estoy emocionado por la oportunidad de contribuir a un proyecto que puede marcar la diferencia en la atención médica a la hora de prevenir ciertas enfermedades. Trabajar codo a codo con un centro médico de renombre como La Fe me da motivación y compromiso, sabiendo que mi trabajo puede ser parte de una solución que mejore la vida de los pacientes.

El desafío de desarrollar una red neuronal capaz de detectar pólipos es una meta ambiciosa, pero la perspectiva de poder ofrecer a los equipos médicos un apoyo decisivo en la identificación de pólipos, especialmente aquellos con displasia, impulsa a avanzar con el proyecto.

## 1.2. Objetivos del Proyecto

---

El objetivo central de este proyecto es el diseño, desarrollo e implementación de un *software* basado en redes neuronales destinado a la detección en tiempo real de pólipos durante colonoscopias. La meta principal es mejorar la eficiencia y precisión en la identificación de estas anomalías, con un enfoque particular en pólipos con características asociadas a displasia, marcadores potenciales de malignidad.

Para alcanzar este objetivo, se plantea:

- **Entrenar una red neuronal con imágenes disponibles:** se ha trabajado en el entrenamiento de una red neuronal utilizando las imágenes de bases de datos disponibles en internet de pólipos, con el fin de mejorar la detección de pólipos durante colonoscopias.
- **Buscar y evaluar diversas arquitecturas de redes neuronales:** se ha explorado y evaluado diversas arquitecturas de redes neuronales para encontrar aquella que presente buenos resultados en la detección de pólipos, y que pueda servir como base para futuros desarrollos.
- **Incrementar progresivamente la base de datos de imágenes:** se ha planificado aumentar progresivamente la base de datos de imágenes de pólipos disponible para el entrenamiento de la red neuronal. Esto se logrará mediante la adquisición y etiquetado de nuevas imágenes, con el objetivo de mejorar la robustez y generalización del modelo, así como su capacidad para detectar una variedad más amplia de pólipos y sus características asociadas.
- **Priorizar la detección de pólipos con displasia:** se establece como objetivo principal priorizar la detección de pólipos con displasia durante la colonoscopia. Esto implica optimizar los algoritmos de detección para identificar pólipos con características indicativas de displasia, mejorando así la capacidad de anticipar posibles casos de malignidad.

Este proyecto no solo busca impulsar el avance tecnológico en el ámbito específico de la endoscopia, sino también mejorar sustancialmente la eficacia de los procedimientos de detección y prevención de enfermedades gastrointestinales, con un impacto directo en la calidad de la atención médica y, en última instancia, en la salud y bienestar de los pacientes.

## 1.3. Colaboradores del Proyecto

---

Este proyecto se lleva a cabo gracias a una colaboración estrecha y fructífera entre la Universidad Politécnica de Valencia y el Hospital Universitario y Politécnico La Fe de Valencia. Ambas instituciones contribuyen con recursos humanos y conocimientos especializados para lograr los objetivos establecidos.

### **Desde la Universidad Politécnica de Valencia:**

El equipo de la universidad está compuesto por investigadores de Vision por Computador del Instituto Universitario de Automática e Informática Industrial(AI2)[2] de la Facultad de Ingeniería Informática. Este instituto agrupa a un conjunto de investigadores y profesionales multidisciplinares especializados en el ámbito de la informática y la inteligencia artificial. La diversidad de habilidades y experiencias dentro del equipo asegura una aproximación al diseño y

desarrollo del *software* basado en redes neuronales para la detección de pólipos durante colonoscopias.

**Desde el Hospital Universitario y Politécnico La Fe de Valencia:**

El grupo de colaboradores del hospital incluye a médicos especialistas en gastroenterología y endoscopia, así como a profesionales de la salud con experiencia en procedimientos endoscópicos. Este equipo aporta conocimientos clínicos valiosos, perspectivas prácticas y un entendimiento profundo de las necesidades específicas del entorno hospitalario. La participación de los profesionales de la salud garantiza que el *software* desarrollado sea práctico, funcional y se integre de manera efectiva en el flujo de trabajo clínico.

La interacción constante entre estos dos grupos de colaboradores es fundamental para el éxito del proyecto. La combinación de conocimientos técnicos y clínicos no solo enriquece la calidad de la investigación y el desarrollo, sino que también garantiza que el *software* final sea una herramienta efectiva y útil para los profesionales médicos.

Esta colaboración refleja un compromiso conjunto hacia la excelencia en la investigación, la innovación y la mejora continua de la atención médica. La convergencia de expertos de la Universidad Politécnica de Valencia y el Hospital La Fe establece las bases para el éxito de este proyecto, que busca contribuir de manera significativa al avance tecnológico y a la optimización de los procedimientos clínicos en la detección de pólipos durante colonoscopias.

## 2. Estado del arte y punto de partida

---

### 2.1. Estado del arte

---

En este capítulo, se examina el estado actual de la detección de pólipos durante colonoscopias, con un enfoque en el uso de tecnologías de inteligencia artificial, como redes neuronales, para mejorar la precisión y eficacia del proceso. Se revisarán proyectos existentes, sistemas comerciales y desarrollos académicos relevantes en este campo.

### 2.2. Proyectos y sistemas comerciales actuales

---

El sistema comercial *GI Genius* [3] desarrollado por *Medtronic* [4], utiliza técnicas avanzadas de visión por computador, incluidas redes neuronales convolucionales, para la detección automática de pólipos durante las colonoscopias. El sistema segmenta y analiza las imágenes de video obtenidas durante el procedimiento, identificando características morfológicas específicas asociadas con pólipos y proporcionando alertas visuales al endoscopista.

Este dispositivo ha sido objeto de un estudio reciente que destaca su impacto en la detección de pólipos durante colonoscopias estándar. Según *Savino et al. (2024)* [5], el *GI Genius* se ha encontrado seguro y efectivo para mejorar la detección de adenomas, especialmente en colonoscopias de cribado y vigilancia para el cáncer colorrectal. Sin embargo, su impacto parece ser menos significativo entre endoscopistas experimentados y en escenarios clínicos del mundo real en comparación con las condiciones controladas de los ensayos clínicos. Además, el estudio encontró que el *GI Genius* principalmente mejora la detección de pólipos no avanzados y pequeños, pero no tiene un impacto significativo en la identificación de adenomas avanzados y difíciles de detectar.

### 2.3. Investigaciones académicas

---

En [6] se presenta un estudio del sistema *CADx*, que se enfoca en el desarrollo de sistemas de ayuda al diagnóstico asistido por computador para la detección automática de pólipos en imágenes endoscópicas. Utiliza una variedad de técnicas de inteligencia artificial, incluidas redes neuronales profundas, para analizar características morfológicas y texturales de las imágenes y distinguir entre tejido sano y pólipos con alta precisión.

Un estudio realizado por *Van Der Zander* [6] validó un sistema de diagnóstico asistido por computadora (*CADx*) que utiliza imágenes de alta definición en blanco y negro (*HDWL*) y de imágenes de luz azul (*BLI*) para caracterizar pólipos colorrectales. Los resultados mostraron que *CADx* tuvo una precisión diagnóstica significativamente mayor que la de expertos y novatos para

el diagnóstico óptico de pólipos colorrectales. La precisión diagnóstica global, combinando *HDWL* y *BLI* (imagen multimodal), fue del 95.0%, lo que fue significativamente superior a la de los expertos (81.7%) y novatos (66.7%). Además, *CADx* mostró una sensibilidad más alta que la de los expertos y novatos, mientras que la especificidad fue más alta para los expertos en comparación con *CADx* y novatos. A pesar de la formación en el uso del sistema de clasificación *BASIC*, no se observó un aumento en la precisión diagnóstica de los endoscopistas al utilizarlo.

**Investigaciones en hospitales y centros de investigación:** Diversos hospitales y centros de investigación están llevando a cabo estudios para mejorar la detección de pólipos durante colonoscopias utilizando tecnologías de inteligencia artificial. Estos proyectos incluyen el desarrollo de algoritmos personalizados y la integración de sistemas automatizados en la práctica clínica, con un enfoque en la optimización del rendimiento en tiempo real y la interoperabilidad con equipos médicos existentes.

## 2.4. Conclusiones

---

El uso de tecnologías de inteligencia artificial, como redes neuronales, en la detección de pólipos durante colonoscopias presenta un gran potencial para mejorar la atención médica gastroenterológica. Los proyectos y sistemas existentes destacan los avances significativos en la detección temprana y precisa de pólipos, lo que podría tener un impacto positivo en el diagnóstico y tratamiento de enfermedades gastrointestinales, incluido el cáncer de colon.

## 3. Marco teórico

---

### 3.1. Inteligencia artificial

---

La Inteligencia Artificial (IA) se erige como una disciplina que abarca el diseño y desarrollo de sistemas capaces de ejecutar tareas que, tradicionalmente, requerirían de la intervención humana. En el contexto de este proyecto, la IA despliega su potencial al incorporar algoritmos avanzados para la detección automatizada de pólipos en imágenes durante procedimientos de colonoscopia. A continuación, se profundiza en aspectos esenciales de la Inteligencia Artificial, subrayando su aplicación específica en el ámbito médico.

#### 3.1.1. Definición de Inteligencia Artificial

---

La IA no tiene una definición única, pero en términos generales se refiere a la capacidad de las máquinas para realizar tareas que implican procesos cognitivos humanos, como el aprendizaje, la resolución de problemas y la toma de decisiones. En este proyecto, la aplicación de la IA se orienta hacia el reconocimiento y clasificación de pólipos, lo que implica dotar al sistema de capacidades para identificar patrones y realizar inferencias a partir de datos visuales en tiempo real.

#### 3.1.2. Aplicaciones de la Inteligencia Artificial en detección de pólipos

---

La Inteligencia Artificial ha demostrado ser una herramienta ideal en la interpretación de imágenes médicas. En el ámbito de la gastroenterología, la aplicación de la IA se focaliza en la mejora de la eficiencia y precisión en la detección temprana de pólipos durante procedimientos endoscópicos. Algoritmos de aprendizaje profundo, como las redes neuronales convolucionales (CNN), son esenciales para analizar imágenes endoscópicas e identificar anomalías con una precisión cada vez mayor.

Estas aplicaciones específicas de la IA no solo buscan agilizar el proceso de detección de pólipos, sino también mejorar la capacidad de identificar características indicativas de displasia, como un marcador temprano de posible malignidad. La IA, al operar en tiempo real durante las colonoscopias, se convierte en una herramienta valiosa para el personal médico, proporcionando información adicional y apoyo en la toma de decisiones clínicas.

La comprensión a fondo de los fundamentos de la Inteligencia Artificial es esencial para la implementación efectiva de algoritmos en el desarrollo del *software* propuesto. En los apartados siguientes, se explorarán aspectos más específicos de la Inteligencia Artificial que respaldan la creación de un sistema autónomo capaz de identificar pólipos durante procedimientos

endoscópicos, contribuyendo así a una mejora sustancial en la detección y prevención de este tipo de enfermedades.

Para una revisión más detallada sobre el uso de la Inteligencia Artificial en la detección de pólipos, se recomienda consultar el artículo de *Byrne et al. (2018)* [7], que proporciona una amplia revisión sobre las técnicas de aprendizaje profundo aplicadas a la detección automática y clasificación de pólipos en endoscopia.

## 3.2. Machine Learning

---

El *Machine Learning* (ML), o aprendizaje automático, representa una rama crucial de la Inteligencia Artificial que se centra en el desarrollo de algoritmos y modelos capaces de aprender patrones y realizar tareas específicas sin intervención humana directa. En el contexto de este proyecto, el *Machine Learning* desempeña un papel esencial en la capacitación del *software* para la detección precisa de pólipos durante colonoscopias. A continuación, se explora el rol fundamental del *Machine Learning*, destacando sus aplicaciones específicas en este contexto médico.

### 3.2.1. Definición de Machine Learning

---

En el contexto de la detección de pólipos, el *Machine Learning* (ML) se define como un enfoque computacional que permite a los sistemas aprender automáticamente y mejorar su rendimiento en la tarea de detección de pólipos a medida que se exponen a más datos. Este aprendizaje se realiza a través de la identificación de patrones complejos y variados presentes en las imágenes endoscópicas, lo que permite al *software* desarrollar modelos capaces de reconocer y distinguir entre tejido sano y tejido con pólipos con una precisión cada vez mayor.

Los algoritmos de ML utilizados en la detección de pólipos pueden ser supervisados o no supervisados. En el caso de algoritmos supervisados, se utilizan conjuntos de datos etiquetados que contienen imágenes endoscópicas previamente clasificadas como positivas (con pólipos) o negativas (sin pólipos). Estos datos se utilizan para entrenar el modelo, que ajusta sus parámetros para optimizar su capacidad de hacer predicciones precisas sobre nuevas imágenes. Por otro lado, en los algoritmos no supervisados, el modelo aprende a identificar patrones y estructuras en los datos sin la necesidad de etiquetas previas, lo que puede ser útil para descubrir anomalías o patrones inesperados en las imágenes.

El ML en la detección de pólipos abarca una variedad de técnicas y enfoques, incluyendo redes neuronales convolucionales (CNN), árboles de decisión, máquinas de vectores de soporte (SVM), y algoritmos de *clustering*, entre otros. Estas técnicas permiten al *software* analizar características morfológicas y texturales de las imágenes endoscópicas, como la forma, el color y la textura de los tejidos, para identificar con precisión la presencia de pólipos y diferenciarlos del tejido circundante. La **Figura 1** ilustra el proceso básico de un sistema de ML, que puede ser supervisado o no supervisado, aplicado a la clasificación de imágenes, en este caso, la identificación de un coche. Este proceso es análogo a cómo se implementa ML en la detección de pólipos.

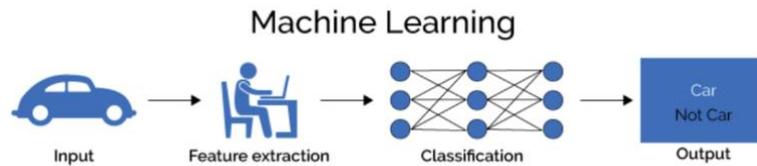


Figura 1: Proceso de *Machine Learning*

### 3.2.2. Aplicaciones del Machine Learning en detección de pólipos

---

En el campo de la gastroenterología, el *Machine Learning* (ML) se utiliza para analizar grandes conjuntos de datos de imágenes endoscópicas y aprender patrones asociados con la presencia de pólipos [8]. Algoritmos de aprendizaje supervisado, donde el modelo se entrena utilizando imágenes previamente etiquetadas, son esenciales para enseñar al *software* a distinguir entre pólipos y tejido normal. Además, el aprendizaje no supervisado puede revelar patrones más sutiles y complejos, mejorando la capacidad del sistema para identificar pólipos con características específicas, como la displasia.

La capacidad del *ML* para adaptarse y mejorar su rendimiento a medida que se alimenta con más datos permite que el *software* se ajuste a las variaciones y complejidades presentes en las imágenes endoscópicas reales. Esto resulta crucial para la aplicación exitosa de esta tecnología en un entorno clínico, donde la diversidad de casos y condiciones puede ser considerable.

La Figura-1, resume el proceso de *ML*. La comprensión de los principios del *Machine Learning* es esencial para desarrollar un sistema capaz de aprender y evolucionar con el objetivo de mejorar la detección de pólipos durante procedimientos endoscópicos. En las siguientes secciones, se profundizará en aspectos más específicos del *Machine Learning* que respaldan el desarrollo de la herramienta propuesta.

## 3.3. Deep Learning

---

*Deep Learning* (DL), o aprendizaje profundo, es una rama especializada del *ML* que se enfoca en el desarrollo de redes neuronales artificiales con capacidades de aprendizaje más avanzadas y complejas. En contraste con los enfoques tradicionales de *Machine Learning*, el *Deep Learning* se destaca por su capacidad para aprender representaciones de datos de manera automática a partir de conjuntos de datos no etiquetados o parcialmente etiquetados. Esta capacidad se logra mediante el uso de arquitecturas de redes neuronales profundas que están compuestas por múltiples capas.

En el contexto de este proyecto, el *Deep Learning* emerge como un componente esencial para el reconocimiento y clasificación precisa de pólipos en imágenes endoscópicas. La naturaleza jerárquica de las redes neuronales profundas permite la extracción de características complejas y abstractas de los datos, lo que resulta fundamental para identificar patrones intrincados presentes en las imágenes endoscópicas.

La introducción de múltiples capas en las arquitecturas de redes neuronales profundas permite la creación de modelos más sofisticados que pueden aprender representaciones de datos en diferentes niveles de abstracción. Esto significa que las características simples, como bordes y texturas, se pueden combinar para formar representaciones más complejas, como formas y

estructuras, que son relevantes para la detección y clasificación de pólipos. La Figura-2, resume el proceso de DL.

En el contexto de este proyecto, el *Deep Learning* se presenta como un componente esencial para el reconocimiento y clasificación precisa de pólipos. A continuación, se explica en detalle el papel del *Deep Learning* y sus aplicaciones específicas en el ámbito médico. La **Figura 2** ilustra el proceso de Deep Learning, que se caracteriza por su capacidad de aprender representaciones complejas de los datos a través de múltiples capas de procesamiento, conocidas como capas ocultas en redes neuronales profundas.

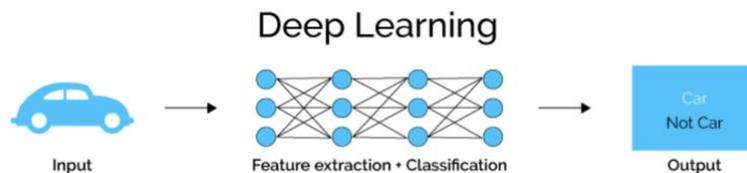


Figura 2: Proceso de Deep Learning

### 3.3.1. Redes neuronales en el contexto de Deep Learning

---

En el ámbito médico, las redes neuronales profundas se aplican con éxito para tareas de procesamiento de imágenes y reconocimiento de patrones. En particular, en la detección de pólipos durante colonoscopias, estas redes tienen la capacidad de aprender representaciones complejas de los tejidos y reconocer características específicas asociadas a la presencia de pólipos.

- **Redes Neuronales Convolucionales (CNN):** Se han aplicado con éxito en tareas de procesamiento de imágenes y reconocimiento de patrones. Especialmente en la detección de pólipos durante colonoscopias, las CNN's han demostrado ser especialmente efectivas. Su capacidad para capturar patrones espaciales y jerárquicos las convierte en herramientas valiosas para identificar pólipos en imágenes endoscópicas. Un estudio reciente realizado, titulado "*A deep convolutional neural network for the detection of polyps in colonoscopy images*" proporciona una perspectiva detallada sobre el uso de CNN en este contexto [9].
- **Transfer Learning:** La técnica de *Transfer Learning*, que implica utilizar conocimientos previos aprendidos en una tarea para mejorar el rendimiento en otra, se muestra relevante en este contexto. Pre-entrenar una red con conjuntos de datos masivos antes de afinarla para la detección específica de pólipos puede mejorar significativamente la eficiencia del modelo.

La aplicación del *Deep Learning* en este proyecto se traduce en la capacidad de la herramienta propuesta para aprender y adaptarse de manera más sofisticada a las complejidades de las imágenes endoscópicas, mejorando así la precisión y la eficiencia en la detección de pólipos durante procedimientos de colonoscopia.

## 3.4. Redes Neuronales

---

Las redes neuronales, fundamentales en el ámbito del aprendizaje automático, son estructuras computacionales inspiradas en el funcionamiento del cerebro humano. En el contexto de este proyecto, el empleo de redes neuronales es esencial para construir un sistema inteligente capaz de

identificar pólipos durante colonoscopias. A continuación, se profundiza en los aspectos clave de las redes neuronales y su aplicación específica en el análisis de imágenes endoscópicas.

### 3.4.1. Redes Neuronales Artificiales (ANNs)

Las Redes Neuronales Artificiales (*ANNs*) son la columna vertebral de muchas aplicaciones de inteligencia artificial debido a su capacidad para aprender y modelar relaciones complejas entre conjuntos de datos. Inspiradas en el funcionamiento del cerebro humano, las *ANNs* están compuestas por nodos interconectados, también conocidos como neuronas, organizados en capas. Cada neurona en una capa está conectada a las neuronas de la capa anterior y posterior mediante conexiones ponderadas, lo que permite la transferencia y transformación de información a lo largo de la red. La Figura-3, resume el proceso de las ANN's.

En el contexto de la detección de pólipos durante procedimientos endoscópicos, las *ANNs* desempeñan un papel fundamental. Estas redes son capaces de procesar grandes volúmenes de datos de imágenes endoscópicas y aprender patrones representativos que distinguen entre tejido sano y tejido con pólipos. A medida que se alimentan con más datos, las *ANNs* ajustan los pesos de las conexiones entre neuronas para mejorar su capacidad de reconocimiento y clasificación.

Las *ANNs* se caracterizan por su capacidad para aprender de manera no lineal, lo que significa que pueden identificar y modelar relaciones complejas y no lineales entre las características de entrada y salida de los datos. Esta capacidad es crucial en la detección de pólipos, ya que los patrones presentes en las imágenes endoscópicas pueden ser altamente variables y no seguir una estructura lineal predefinida. Además, las *ANNs* pueden adaptarse a diferentes tipos de datos y problemas de manera flexible. Con la disponibilidad de algoritmos de optimización eficientes y potentes unidades de procesamiento gráfico (GPU) para acelerar los cálculos, las *ANNs* pueden entrenarse en conjuntos de datos grandes y complejos de manera más rápida y eficiente que nunca.

En resumen, las Redes Neuronales Artificiales son herramientas poderosas en el campo de la detección de pólipos durante procedimientos endoscópicos, ya que pueden aprender y modelar patrones complejos presentes en las imágenes endoscópicas, contribuyendo así a una identificación automática más precisa y eficiente de los pólipos. La **Figura 3** proporciona una representación simplificada de una red neuronal artificial, ilustrando cómo las entradas se transforman en salidas a través de una serie de capas conectadas.

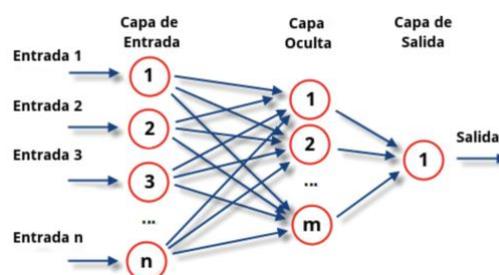


Figura 3: Representación de Red Neuronal Artificial

## 3.4.2. Redes Neuronales Convolucionales (CNNs)

---

Las Redes Neuronales Convolucionales (CNNs), representan una variante especializada y altamente efectiva de las Redes Neuronales Artificiales (ANNs). Están especialmente diseñadas para el procesamiento eficiente de datos bidimensionales, como imágenes, y han demostrado ser excepcionalmente poderosas en una variedad de tareas de visión por computadora, incluida la detección de objetos, segmentación semántica y, específicamente, la identificación de pólipos en imágenes endoscópicas.

Lo que distingue a las CNNs es su capacidad inherente para aprender automáticamente características espaciales y jerárquicas a partir de los datos de entrada. Esto se logra mediante el uso de capas convolucionales que aplican filtros a regiones locales de la imagen, lo que permite capturar patrones locales y construir representaciones cada vez más abstractas a medida que la información se propaga a través de las capas de la red. La Figura-4, resume el proceso de las CNN's.

En el contexto de la detección de pólipos durante procedimientos endoscópicos, las CNNs desempeñan un papel crucial y destacado. Su arquitectura está especialmente diseñada para capturar características relevantes en las imágenes, como bordes, texturas y formas, que son indicativas de la presencia de pólipos. Al aprender a reconocer estos patrones característicos de manera automática y adaptativa durante el entrenamiento, las CNNs pueden lograr una detección precisa y rápida de anomalías, lo que resulta fundamental para mejorar la eficiencia y la efectividad de los procedimientos endoscópicos. La capacidad de las CNNs para generalizar a partir de ejemplos de entrenamiento limitados y su capacidad para adaptarse a diferentes condiciones de iluminación, escalas y orientaciones hacen que sean altamente robustas y versátiles en la detección de pólipos en una variedad de escenarios clínicos.

Las CNNs representan una herramienta poderosa y altamente especializada en el campo de la detección de pólipos durante procedimientos endoscópicos. Su capacidad para aprender automáticamente características espaciales y jerárquicas de las imágenes las hace particularmente adecuadas para identificar patrones indicativos de la presencia de pólipos, lo que contribuye significativamente a mejorar la precisión y la eficiencia de los procesos de diagnóstico en gastroenterología.

En la **Figura 4**, se muestra un flujo típico de procesamiento en una CNN, destacando las diversas fases a través de las cuales se transforman los datos de entrada en salidas de clasificación precisas.

### *Fase de Entrada (Input)*

- **Entrada de Imagen:** La imagen de entrada se suministra al modelo, en este caso, una imagen que podría ser de un coche o de un pólipo. En el contexto de la detección de pólipos, la imagen podría ser una captura endoscópica que el modelo necesita analizar.

### *Fase de Convolución (Conv)*

- **Operación de Convolución:** Durante esta fase, la CNN aplica filtros convolucionales a la imagen de entrada. Estos filtros actúan como detectores de características, capaces de identificar bordes, texturas, y otros patrones relevantes. En el caso de la detección de pólipos, los filtros pueden resaltar características específicas de los pólipos, como su forma o textura.

### *Fase de Submuestreo (Pool)*

- **Pooling (Submuestreo):** Esta fase reduce la dimensionalidad de la salida de la convolución, lo que ayuda a resumir la información más relevante y a reducir la carga computacional. Las operaciones de *pooling*, como el *max-pooling*, toman el valor máximo en pequeñas regiones de la imagen, preservando las características más destacadas. Esto es crucial para mantener la información clave mientras se reduce el tamaño de los datos.

### *Fase de Convolución y Submuestreo Repetida (Conv-Pool)*

- **Repetición de Convolución y Pooling:** Estas operaciones suelen repetirse varias veces para extraer características de alto nivel de la imagen. Cada capa convolucional posterior puede capturar características más abstractas y complejas, lo que es esencial para identificar estructuras más detalladas en las imágenes de pólipos.

### *Fase de Conexión Completa (Fully Connected - FC)*

- **Capa Densa:** Las capas completamente conectadas reciben las características extraídas y las combinan para formar la representación final. En esta fase, todos los nodos de la capa anterior están conectados a cada nodo de la capa siguiente, permitiendo que la red tome decisiones basadas en la combinación de todas las características aprendidas.

### *Fase de Clasificación (Softmax)*

- **Capa Softmax:** La capa de salida *softmax* toma las salidas de las capas completamente conectadas y las convierte en probabilidades, asignando una probabilidad a cada clase de salida. En la detección de pólipos, esta capa podría determinar la probabilidad de que la imagen contenga un pólipo o no.

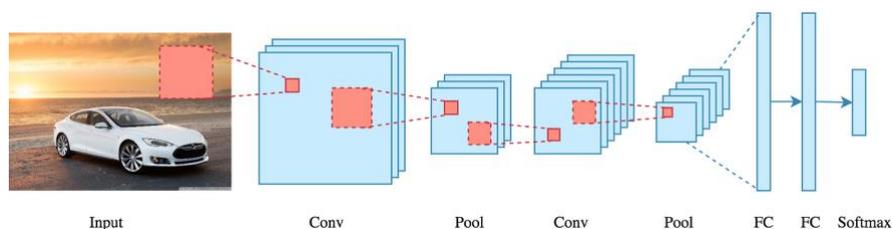


Figura 4: Representación de Red Neuronal Convolutivas

## 3.4.3. Transfer Learning

El *Transfer Learning* es una técnica fundamental en el ámbito del aprendizaje automático que permite aprovechar el conocimiento previamente adquirido por una red neuronal en una tarea específica para mejorar el rendimiento en otra tarea relacionada. Esta metodología es especialmente relevante en el contexto de la detección de pólipos durante procedimientos endoscópicos, donde puede acelerar significativamente el entrenamiento del modelo, sobre todo cuando se dispone de conjuntos de datos limitados.

En esencia, el *Transfer Learning* implica tomar una red neuronal pre-entrenada en un conjunto de datos grande y general (por ejemplo, un conjunto de datos de imágenes de Internet) y reutilizar sus capas y pesos para una tarea específica relacionada (en este caso, la detección de pólipos en imágenes endoscópicas), como vemos en la Figura 5 que reutiliza los pesos de la 'Task 1' para entrenar 'Task 2'. Al hacer esto, se pueden evitar las limitaciones de tener que entrenar un modelo desde cero con un conjunto de datos pequeño y específico. La Figura-5, resume el proceso de *Transfer Learning*.

Esta técnica se ha explorado en numerosos estudios recientes dentro del campo de la detección de pólipos. Por ejemplo, el estudio realizado por *Dulf et al.*, titulado "*Transfer Learning with Deep Convolutional Neural Network for Polyp Classification*," publicado en la revista *Sensors* [10], destaca cómo el *Transfer Learning* puede mejorar significativamente la capacidad de las redes neuronales para clasificar pólipos en imágenes endoscópicas.

La aplicación de *Transfer Learning* en este proyecto permite construir un sistema robusto y adaptable, capaz de aprender y reconocer patrones complejos en imágenes endoscópicas. Al reutilizar el conocimiento previamente adquirido por redes neuronales entrenadas en conjuntos de datos generales, el modelo puede aprovechar características y representaciones aprendidas que son relevantes para la tarea específica de detección de pólipos. Esto no solo acelera el proceso de entrenamiento, sino que también mejora la capacidad del modelo para generalizar y adaptarse a nuevas condiciones y variaciones en los datos.

En conjunto, el *Transfer Learning* representa una herramienta poderosa y efectiva para mejorar el rendimiento de los modelos de detección de pólipos durante procedimientos endoscópicos. Su capacidad para capitalizar el conocimiento previo y aplicarlo a nuevas tareas permite avanzar hacia sistemas más precisos y eficientes en la detección temprana de anomalías gastrointestinales, lo que tiene un impacto significativo en la práctica clínica y el bienestar de los pacientes.

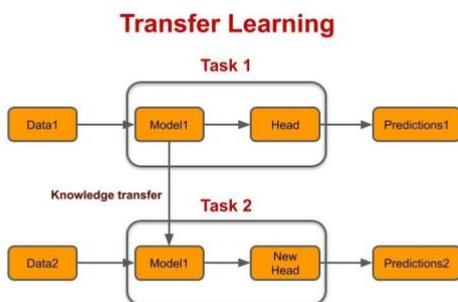


Figura 5: Proceso de Transfer Learning

### 3.5. YOLO

*You Only Look Once (YOLO)* es un enfoque innovador en el ámbito de la detección de objetos en imágenes que ha ganado una considerable popularidad debido a su eficiencia y precisión [11]. A diferencia de los métodos tradicionales que implican múltiples etapas, como la propuesta de regiones o la extracción de características, *YOLO* adopta un enfoque más directo y eficiente al realizar la detección y clasificación de objetos en una sola pasada de la red neuronal.

El algoritmo de *YOLO* opera dividiendo la imagen de entrada en una cuadrícula de celdas y, en cada celda, predice simultáneamente los cuadros delimitadores (*bounding boxes*) y las probabilidades de clase para los objetos presentes. Este enfoque de "detección en una sola pasada" permite una detección rápida y precisa de múltiples objetos en tiempo real, lo que lo convierte en



una opción atractiva para aplicaciones donde la velocidad es crítica, como la detección de objetos en videos o imágenes en tiempo real.

La arquitectura de *YOLO* se ha mejorado y refinado a lo largo del tiempo, dando lugar a versiones posteriores como *YOLOv2*, *YOLOv3*, y *YOLOv4*, cada una con mejoras significativas en términos de precisión y velocidad de detección [12]. En este proyecto, se utiliza específicamente la versión *YOLOv8.1.4* de *Ultralytics YOLO*, una implementación avanzada que incorpora mejoras adicionales en términos de rendimiento y funcionalidad [13]. Estas versiones mejoradas han integrado características como la utilización de capas convolucionales profundas, técnicas de aumento de datos, y estrategias avanzadas de regularización para mejorar la capacidad de generalización del modelo y reducir el riesgo de sobreajuste.

En el contexto específico de la detección de pólipos en imágenes médicas, *YOLO* ha demostrado ser una herramienta valiosa [14]. Su capacidad para detectar y localizar objetos de manera rápida y precisa lo hace especialmente adecuado para aplicaciones en las que se necesita una detección temprana y confiable de anomalías. Además, su capacidad de adaptarse a diferentes dominios de aplicación y su escalabilidad hacen que sea una opción versátil y poderosa para una variedad de problemas de detección de objetos en la práctica médica y más allá.

### 3.5.1. Beneficios de YOLO

---

Los principales beneficios de *YOLO* son rapidez y eficiencia, facilidad de implementación y eficiencia en recursos computacionales.

**Rapidez y Eficiencia:** Una de las principales ventajas de *YOLO* radica en su rapidez y eficiencia en la detección de objetos en imágenes. Esto se debe a su enfoque innovador que reduce la detección a un problema de regresión lineal. Al hacerlo, elimina la necesidad de un pipeline complejo y permite que la red neuronal realice predicciones rápidas directamente en la imagen de entrada. En lugar de dividir la imagen en regiones y procesarlas por separado, *YOLO* consume la imagen completa, lo que limita los errores en la detección de objetos y garantiza una detección rápida y precisa en tiempo real. Esta capacidad de procesamiento rápido es crucial en aplicaciones donde se requiere una respuesta instantánea, como la detección de pólipos durante procedimientos endoscópicos.

**Facilidad de implementación:** Otro beneficio significativo de *YOLO* es su facilidad de implementación y entrenamiento. Al entrenar y detectar con la imagen completa, *YOLO* simplifica el proceso de entrenamiento de la red neuronal, lo que resulta en una curva de aprendizaje más suave y una mejora en la precisión del modelo. Además, la capacidad de *YOLO* para aprender representaciones generalizables de objetos reduce las probabilidades de fallo cuando se introducen datos de entrada nuevos. Esto significa que el modelo es más robusto y confiable en la detección de objetos en una variedad de condiciones y contextos, en comparación con técnicas anteriores que pueden ser más sensibles a pequeñas variaciones en los datos de entrada.

**Eficiencia en recursos computacionales:** Una ventaja adicional de *YOLO* es su eficiencia en el uso de recursos computacionales. A pesar de su capacidad para realizar detecciones precisas en tiempo real, *YOLO* requiere menos recursos de hardware en comparación con otras arquitecturas de detección de objetos, lo que lo hace más accesible y económico de implementar en sistemas con limitaciones de recursos.

## 3.5.2. Funcionamiento de YOLO

---

El funcionamiento de *YOLO* se basa en un enfoque único y altamente eficiente para la detección de objetos utilizando una sola pasada a través de una red neuronal convolucional (CNN). Esta metodología revolucionaria ha establecido nuevos estándares en la precisión y velocidad de detección, lo que la convierte en una opción atractiva para una amplia gama de aplicaciones, incluida la detección de pólipos en imágenes endoscópicas.

En el corazón del funcionamiento de *YOLO* se encuentra la división de la imagen de entrada en una rejilla de tamaño  $S \times S$ . Cada celda de esta rejilla asume la responsabilidad de detectar objetos si están presentes en su área correspondiente. Esta estrategia permite una detección exhaustiva y eficiente de objetos en toda la imagen, maximizando así la cobertura del modelo y garantizando que ningún objeto pase desapercibido. Dentro de cada celda de la rejilla, *YOLO* predice  $B$  cuadros delimitadores, cada uno con su correspondiente puntuación de confianza. Estos cuadros delimitadores definen las regiones de la imagen donde se estima que se encuentran los objetos detectados. Además, *YOLO* también calcula las probabilidades de clase condicionales para cada cuadro delimitador, lo que permite la clasificación precisa de los objetos en diferentes categorías.

El proceso de detección de objetos en *YOLO* se lleva a cabo de manera conjunta para toda la imagen, lo que significa que la red neuronal considera simultáneamente todas las regiones de la imagen en lugar de dividirla en regiones más pequeñas. Esta capacidad de procesamiento holístico proporciona a *YOLO* una ventaja significativa en términos de velocidad y eficiencia, permitiendo la detección en tiempo real de objetos incluso en imágenes de alta resolución.

En resumen, el funcionamiento de *YOLO* se caracteriza por su enfoque integral y eficiente para la detección de objetos, aprovechando al máximo la información de la imagen completa para generar predicciones precisas y rápidas. Esta metodología innovadora ha demostrado ser altamente efectiva en una variedad de aplicaciones, incluida la detección de pólipos en imágenes médicas, y sigue siendo objeto de investigación y desarrollo continuos para mejorar aún más su rendimiento y versatilidad.

## 3.5.3. Arquitectura de YOLO

---

La arquitectura de red de *YOLO* es un hito en el campo de la detección de objetos gracias a su eficiencia y precisión. Esta arquitectura se implementa como una red neuronal convolucional (CNN), caracterizada por su simplicidad y capacidad para realizar detecciones en tiempo real. La esencia de *YOLO* radica en su capacidad para procesar la imagen completa de una sola vez, en lugar de dividirla en regiones como lo hacen otros enfoques.

El diseño de la red *YOLO* consta de varias capas convolucionales, que se encargan de extraer características relevantes de la imagen de entrada. Para reducir el número de capas y mantener la eficiencia computacional, se emplean convoluciones  $1 \times 1$  seguidas de convoluciones  $3 \times 3$ . Esta combinación permite capturar información contextual a diferentes escalas espaciales, lo que mejora la capacidad del modelo para detectar objetos de diversos tamaños.

La capa final de convolución genera un tensor que contiene la información necesaria para las predicciones de detección de objetos. Este tensor se procesa a través de capas completamente conectadas, que refinan y ajustan las predicciones antes de emitir las salidas finales. Este enfoque de diseño permite que *YOLO* genere detecciones precisas y rápidas, lo que lo convierte en una



opción popular para aplicaciones que requieren detección en tiempo real, como la detección de pólipos en imágenes endoscópicas.

En resumen, la arquitectura de *YOLO* se destaca por su simplicidad, eficiencia y precisión en la detección de objetos, incluidos los pólipos en imágenes médicas. Su capacidad para procesar la imagen completa de una sola vez la convierte en una herramienta invaluable para aplicaciones que requieren detección rápida y precisa de objetos en tiempo real.

## 3.6. Métricas de rendimiento

En el contexto de la detección de pólipos y otras aplicaciones de inteligencia artificial, evaluar el rendimiento de los modelos es crucial para comprender su eficacia y su capacidad para generalizar datos no vistos. Para ello, se emplean diversas métricas que proporcionan una medida cuantitativa del desempeño del modelo. Algunas de las métricas más comúnmente utilizadas incluyen:

- **Precisión (Accuracy):** Es la proporción de predicciones correctas sobre el total de predicciones realizadas por el modelo. Esta métrica es útil para evaluar el rendimiento general del modelo, pero puede ser engañosa en conjuntos de datos desbalanceados, donde una clase es mucho más común que otras.

$$\text{Precisión} = \frac{TP+TN}{TP+TN+FP+FN}$$

Donde:

TP: es el número de verdaderos positivos (casos positivos correctamente clasificados).

TN: es el número de verdaderos negativos (casos negativos correctamente clasificados).

FP: es el número de falsos positivos (casos negativos incorrectamente clasificados como positivos).

FN: es el número de falsos negativos (casos positivos incorrectamente clasificados como negativos).

- **Precisión por clase:** Esta métrica proporciona la precisión para cada clase individualmente, lo que permite identificar si el modelo está teniendo dificultades para distinguir ciertas clases.

$$\text{Precisión por clase}(C) = \frac{TP_C}{TP_C+FP_C}$$

Donde:

TP<sub>C</sub>: es el número de verdaderos positivos para la clase *C* (casos de la clase *C* correctamente clasificados).

FP<sub>C</sub>: es el número de falsos positivos para la clase *C* (casos de otras clases incorrectamente clasificados como *C*).

- **Recall**
- (Recuperación o Sensibilidad): Es la proporción de instancias de una clase que fueron correctamente detectadas por el modelo sobre el total de instancias reales de esa clase en el conjunto de datos. Es especialmente importante en aplicaciones médicas, donde es crucial minimizar los falsos negativos, es decir, la omisión de casos positivos.

$$\text{Recall} = \frac{\text{Verdaderos positivos (TP)}}{\text{Verdaderos positivos (TP)} + \text{Falsos negativos (FN)}}$$

- **F1-score:** Es la media armónica de la precisión y el recall. Proporciona un equilibrio entre la precisión y el recall y es útil cuando hay un desequilibrio entre las clases en el conjunto de datos.

$$F1-score = 2 \times \frac{\text{Precisión} \times \text{Recall}}{\text{Precisión} + \text{Recall}}$$

- **Curva ROC** (*Receiver Operating Characteristic*) y **área bajo la curva** (AUC): Estas métricas se utilizan comúnmente para evaluar el rendimiento de modelos de clasificación binaria. La curva ROC representa la tasa de verdaderos positivos frente a la tasa de falsos positivos en varios umbrales de decisión, y el AUC proporciona una medida agregada del rendimiento del modelo independientemente del umbral de decisión.

Seleccionar la métrica más adecuada depende del contexto específico del problema y de las necesidades del usuario final. En el caso de la detección de pólipos, es fundamental considerar tanto la sensibilidad para detectar todos los casos positivos como la especificidad para evitar falsas alarmas, garantizando así la fiabilidad y utilidad clínica del modelo.



## 4. Marco práctico

---

### 4.1. Entorno de desarrollo

---

El entorno de desarrollo utilizado para implementar y entrenar el modelo de detección de pólipos se configuró en un sistema operativo *Ubuntu 20.04*. Se emplearon diversas herramientas y bibliotecas de *software*, incluyendo *CUDA* [15] para el procesamiento paralelo en GPU, *PyTorch* [16] como el *framework* principal de aprendizaje profundo, y *Jupyter Notebook* [17] como el entorno interactivo para la prototipación y desarrollo del código.

Para garantizar un rendimiento óptimo y una eficiente utilización de los recursos de hardware, se aprovechó la potencia de cálculo de las GPU compatibles con *CUDA*, como la *NVIDIA RTX 3090* con 24268 MiB de memoria. Esto permitió acelerar significativamente el proceso de entrenamiento del modelo, aprovechando al máximo la capacidad de cómputo de las GPU disponibles. La integración de *CUDA* con *PyTorch* proporcionó una plataforma robusta para el desarrollo de modelos de aprendizaje profundo, permitiendo realizar cálculos de manera eficiente y aprovechando la aceleración proporcionada por la GPU.

En resumen, el entorno de desarrollo se diseñó cuidadosamente para proporcionar un marco de trabajo eficiente y productivo para la implementación y entrenamiento del modelo de detección de pólipos basado en *YOLOv8.1.4* de *Ultralytics* en el sistema operativo *Ubuntu 20.04*. La combinación de herramientas como *CUDA*, *PyTorch*, y *Jupyter Notebook* ofreció una plataforma poderosa y flexible para abordar los desafíos de este proyecto de investigación.

### 4.2. Preparación de las imágenes: Dataset

---

Para entrenar una CNN es necesario disponer de un *dataset* de imágenes lo suficientemente amplio y diverso del problema a resolver. En nuestro caso disponemos de un conjunto de imágenes de pólipos propio todavía muy reducido (250 imágenes), por lo que para la fase de entrenamiento de la red fue necesario recopilar imágenes similares de bases especializadas. A continuación, se detalla tanto la confección de dichos *dataset*, como el procesamiento realizado sobre las imágenes que lo configuran.

El entrenamiento de los modelos de CNNs se llevó a cabo utilizando conjuntos de datos exhaustivos y diversos. En una fase inicial el dataset de entrenamiento al que nos referiremos como *DataTr1* lo formaron un total de 2108 imágenes provenientes de bases públicas. Estas imágenes fueron recopiladas de varias fuentes, incluyendo bases de datos especializadas como *EndoTect* [18], *Kvasir* [19], *HyperKvasir* [20], y *CVC* [21], entre otras. Este conjunto de datos abarcaba una amplia variedad de escenarios clínicos y condiciones de iluminación, lo que garantiza la robustez y la generalización del modelo de detección, la Figura 6.

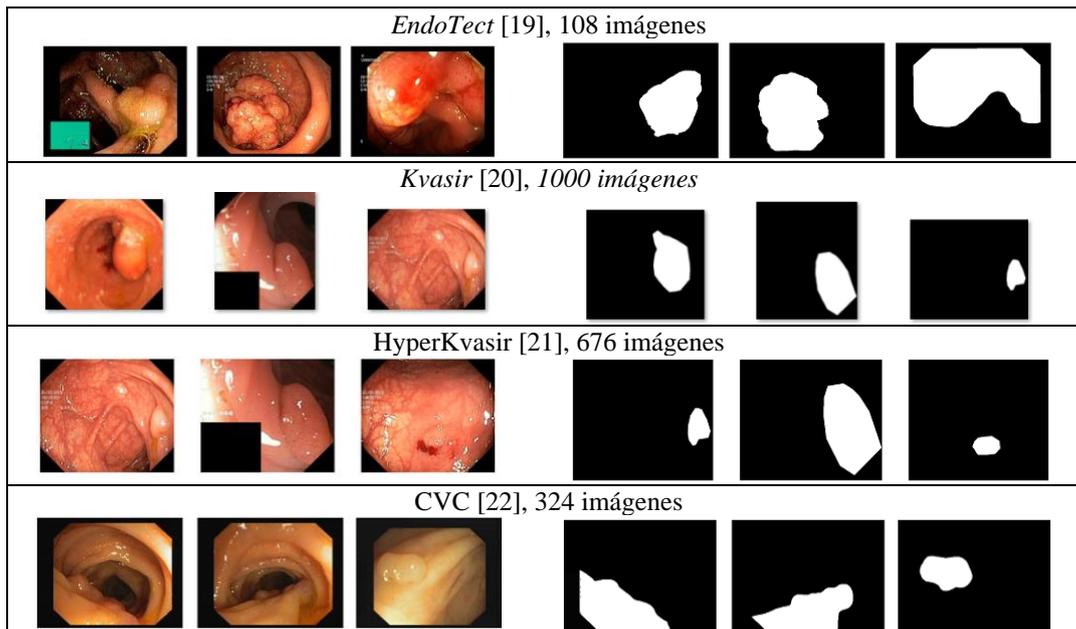


Figura 6: Algunos ejemplos de las imágenes de pólipos recopiladas de EndoTect[18], Kvasir[19], HyperKvasir[20] y CVC[21] para el entrenamiento de las CNN's y que configuran el *DataTr1* lo con un total de 2108 , se muestran tanto la original como la segmentada.

Posteriormente se consiguió aumentar el tamaño y la diversidad del conjunto de datos utilizado en el entrenamiento hasta disponer de un total de 2850 imágenes al que nos referiremos como *DataTr2*. Para ampliar la base de datos, se recurrió a diversas fuentes disponibles públicamente que proporcionaban imágenes médicas relevantes para la detección de pólipos. Estas fuentes incluyeron repositorios de imágenes médicas, bases de datos de investigación y conjuntos de datos abiertos disponibles en línea. Se seleccionaron cuidadosamente las imágenes que cumplieran con los criterios de calidad y relevancia establecidos para garantizar la integridad y la representatividad del conjunto de datos ampliado, en la Figura 7 se muestra ejemplos de las imágenes seleccionadas. Se obtuvieron dos bases de datos de *ENDOCV*[21] sitios como *Kaggle* [22] o *Google Dataset Search* [23].

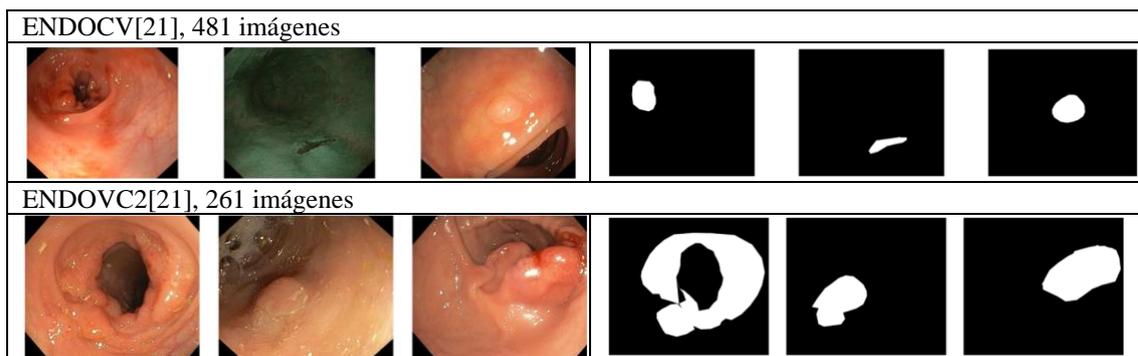


Figura 7: Algunos ejemplos de las imágenes de pólipos recopiladas de *ENDOCV*[21] y *ENDOCV2*[21] para el entrenamiento de las CNN's y estas imágenes junto con las del *DataTr1*, configuran el *DataTr2* formado por un total de 2800 imágenes, se muestran tanto la original como la segmentada.

Para poder utilizar esta diversidad de imágenes, fue necesario aplicar una fase de procesamiento para obtener imágenes normalizadas con píxeles en el rango adecuado, un redimensionamiento para ajustarlas al mismo tamaño antes de ser alimentadas a la red, y además se aplicaron técnicas de *data augmentation* (como rotaciones, traslaciones, escalado, *flip* horizontal/vertical, etc.) para aumentar la diversidad del *dataset* y reducir el riesgo de sobreajuste.

Antes de poder utilizar estas imágenes para el entrenamiento del modelo, fue necesario realizar una serie de pasos de preprocesamiento para resaltar la zona de interés, es decir, los pólipos presentes en las imágenes. Había bases de datos donde las imágenes ya contenían un archivo de texto asociado con el etiquetado de las zonas de interés que eran señaladas con un *bounding box* (*BBox*) por cada pólipo que aparecía en la imagen. Por otro lado, había bases de datos, que venían sin etiquetar en formato texto, sino que simplemente venían señaladas por *bounding boxes* de los cuales no teníamos las coordenadas.

### EJEMPLO DEL PREPROCESAMIENTO DE IMÁGENES:

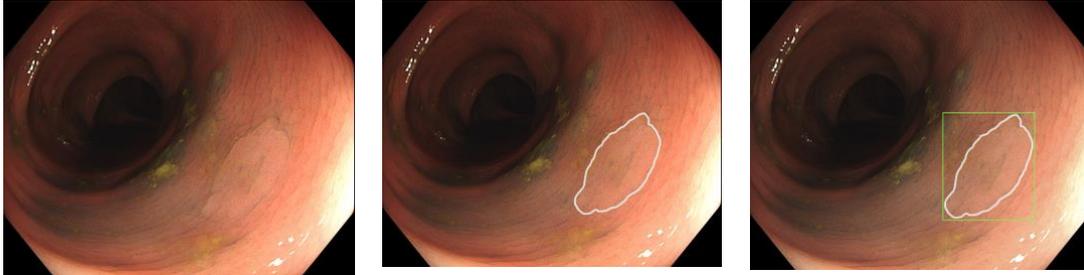


Figura 8: Ejemplo del procesamiento realizado en las imágenes de *dataset* propio, recopiladas por los sanitarios especialistas de la FE y que conforman el *DatasetFeI*, imagen original con pólipo, imagen con pólipo marcado por los sanitarios especializados y pólipo con su *Bounding Box*

Este trabajo de preprocesamiento implicó el desarrollo de un script personalizado que utilizaba técnicas de procesamiento de imágenes y visión por computador para identificar las regiones relevantes en cada imagen. Dependiendo de la base de datos y de cómo estaban marcadas las zonas de interés, se aplicaron diferentes enfoques, como la detección de bordes, la segmentación por color o la identificación de patrones específicos. Una vez identificadas las zonas de interés, se procedió a generar *BBox*'s alrededor de estas regiones y a extraer las coordenadas correspondientes. Estas coordenadas fueron luego convertidas al formato requerido por el algoritmo *YOLO*, que consiste en coordenadas normalizadas del centro de la *BBox* y su ancho y altura relativos a las dimensiones de la imagen.

Cada imagen fue asociada con un archivo de texto con el mismo nombre, que contenía las coordenadas de las zonas de interés en formato *YOLO*. Estos archivos de texto proporcionaron la información necesaria para entrenar el modelo de detección de pólipos utilizando el algoritmo *YOLOv5*.

El proceso de ampliación de la base de datos permitió enriquecer el conjunto de datos de entrenamiento con una mayor variedad de casos y escenarios clínicos, lo que mejoró la capacidad del modelo para generalizar a situaciones diversas y capturar la variabilidad inherente en las imágenes médicas de pólipos. Además, al utilizar imágenes adicionales de diferentes fuentes, se redujo el sesgo potencial y se aumentó la robustez del modelo frente a nuevas muestras.

En resumen, la ampliación de la base de datos se realizó con el objetivo de aumentar la diversidad y el tamaño del conjunto de datos de entrenamiento utilizado en el desarrollo del modelo de detección de pólipos. Este proceso fue fundamental para mejorar la capacidad de generalización del modelo y garantizar su eficacia en la detección precisa de pólipos en imágenes médicas.

## 4.2.1. Conjunto de entrenamiento, validación y test

---

Para el entrenamiento y validación de los modelos se divide el *dataset* en tres conjuntos: conjunto de entrenamiento, validación y test. La mayor parte del *dataset* (80% de las imágenes) debe ser utilizado para el entrenamiento de la red. Una parte del *dataset* (20%), conjunto de validación se reserva para validar el modelo durante su entrenamiento y ajustar hiperparámetros. Al separar el conjunto de datos en un 80% para el entrenamiento y un 20% para la validación, se garantiza una evaluación adecuada del modelo en datos no vistos durante el entrenamiento, lo que ayudó a evitar el sobreajuste y a mejorar la generalización del modelo.

Este enfoque meticuloso y automatizado garantizó la preparación precisa y eficiente del conjunto de datos para el entrenamiento del modelo. Además, al utilizar técnicas de delimitación y etiquetado automatizado, se redujo significativamente la carga manual y se minimizó el riesgo de errores humanos en el proceso de preparación de los datos.

El conjunto de *test* debe estar formado por datos totalmente diferentes al de entrenamiento, debe ser un conjunto disjunto de datos respecto al de entrenamiento y validación y se utiliza para evaluar el rendimiento final del modelo de manera objetiva. En este proyecto se utiliza un conjunto de test íntegramente formado por imágenes etiquetadas por los expertos en endoscopia del hospital de la FE de Valencia. El *DatasetFe1* lo configuran un total de 250 imágenes, que han sido anotadas en formato *YOLO*, y que es utilizado como test en nuestros modelos. La Figura-9 y la Figura-10 muestra imágenes perteneciente dicho dataset propio, con el etiquetado del experto y la detección de la red.



Figura 9: Ejemplo de las imágenes de dataset propio, recopiladas por los sanitarios especialistas de la FE y que conforman el *DatasetFe1*.

Aunque en este proyecto sólo se trabaja con el objetivo evaluar modelos para la detección de pólipos en imágenes médicas, el objetivo más ambicioso de futuro es la clasificación de los diferentes pólipos. En la Tabla-1, se muestra una distribución por tipo de pólipos de las imágenes que configuran el *DatasetFe1*.

Tipo de pólipo	Imágenes	Instancias
Adenoma	21	22
Serrada	81	95
Pseudopólipo	112	216
Normal	33	33
No Claro	1	2
Inflamación	1	0
Hiperplásico	2	2
	<b>251</b>	<b>370</b>

Tabla 1: Distribución de imágenes e instancias separadas por clases del conjunto *DatasetFe1*.

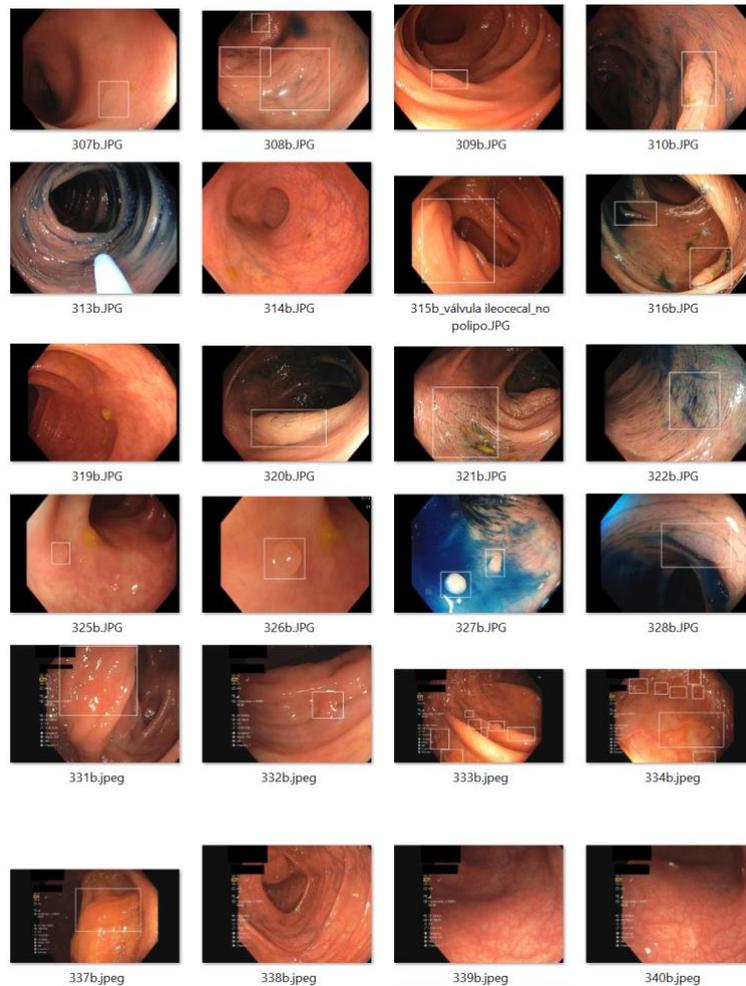


Figura 10: Ejemplo de las imágenes de dataset propio, recopiladas por los sanitarios especialistas de la FE y que conforman el *DatasetFe1*, con el etiquetado proporcionado.

En resumen, el proceso de clasificación de imágenes se llevó a cabo mediante la recopilación y preparación de un conjunto de datos diverso y representativo, que fue procesado con un script personalizado para la detección y delimitación de las zonas de interés. Esta preparación de datos fue fundamental para el entrenamiento efectivo del modelo de detección de pólipos y aseguró la calidad y relevancia de los datos utilizados en el proceso de aprendizaje automático. Para las etapas de *test* de los modelos *CNN*'s se confeccionó un *dataset* propio con las imágenes etiquetadas por los expertos en endoscopia del hospital de la FE de Valencia.

### 4.3. Desarrollo de la CNN y evaluación del modelo

En esta etapa, se procedió al desarrollo de la *Convolutional Neural Network (CNN)* para la detección de pólipos utilizando el conjunto de datos preparado en la fase anterior. Se empleó el *framework PyTorch* [16] para la implementación de la CNN, aprovechando su flexibilidad y eficiencia en el desarrollo de modelos de aprendizaje profundo.

La arquitectura de la CNN con la que se ha trabajado es *YOLOv8.1.4* de *Ultralytics* seleccionada entre las versiones de YOLO teniendo en cuenta las características específicas de las imágenes médicas y la naturaleza de los pólipos. Una vez implementada la arquitectura de la CNN, se

procedió a entrenar el modelo utilizando el conjunto de datos etiquetado previamente, *DataTr1* y *DataTr2*. Durante el entrenamiento, se ajustaron los parámetros del modelo con el objetivo de minimizar la función de pérdida y mejorar la capacidad predictiva del modelo.

Tras completar el entrenamiento, se evaluó el desempeño del modelo utilizando métricas de rendimiento como la precisión, el *Recall* y el *F1-score*, entre otras. Se emplearon conjuntos de datos de prueba previamente reservados para evaluar la capacidad de generalización del modelo a datos no vistos. Además de las métricas de rendimiento cuantitativas, se realizaron análisis cualitativos de las predicciones del modelo, examinando visualmente las detecciones realizadas en imágenes de prueba para identificar posibles áreas de mejora y afinar la arquitectura del modelo si fuera necesario.

En resumen, en esta etapa se llevó a cabo el desarrollo de la CNN para la detección de pólipos, incluyendo el diseño de la arquitectura del modelo, el entrenamiento con el conjunto de datos preparado y la evaluación del rendimiento del modelo. Este proceso permitió iterar y mejorar continuamente el modelo para alcanzar un desempeño óptimo en la detección de pólipos en imágenes médicas.

### 4.3.1. Entrenamiento de la red

---

Durante el proceso de entrenamiento de la red neuronal convolucional (CNN) para la detección de pólipos, se llevaron a cabo experimentos utilizando diferentes números de épocas (*epochs*). Las épocas son un parámetro importante que determina la cantidad de iteraciones que la red neuronal realiza sobre el conjunto de datos durante el entrenamiento. Se exploraron varias configuraciones de épocas, incluyendo 100, 250 y 500, con el fin de determinar el número óptimo de iteraciones para maximizar el rendimiento del modelo.

Inicialmente, se usaron pesos pre-entrenados de la red *COCO* para el entrenamiento de la red de detección de pólipos *YOLOv8*. La red se entrenó durante 100 épocas con el *DataTr1* para evaluar su capacidad de aprendizaje y su rendimiento inicial. Durante este proceso, se monitorizó el progreso del entrenamiento utilizando métricas como la pérdida (*loss*) y la precisión en el conjunto de validación. Posteriormente, se amplió el número de épocas a 250 y luego a 500, con el objetivo de permitir que la red neuronal convergiera hacia una solución óptima y mejorara su capacidad de generalización.

Las métricas de rendimiento del entrenamiento de 250 y 500 épocas con *DataTr1* se muestran en las figuras 12 y 14. Durante el entrenamiento con un mayor número de épocas, se observó una mejora continua en las métricas de rendimiento del modelo, como la precisión y el *Recall*. Sin embargo, se identificó un punto de saturación donde el modelo dejaba de mejorar significativamente, incluso antes de completar las 500 épocas. Este comportamiento sugiere que la red neuronal alcanzó su capacidad máxima de aprendizaje y ya no podía mejorar más, lo que resultó en una convergencia prematura del entrenamiento. Como resultado de estos experimentos, se determinó que el número óptimo de épocas para el entrenamiento del modelo era de 250, donde se logró un equilibrio entre la capacidad de aprendizaje del modelo y el riesgo de sobreajuste. Entrenar la red durante más de 250 épocas no proporcionó beneficios significativos en términos de mejora del rendimiento, mientras que entrenarla durante menos de 250 épocas resultó en un rendimiento sub-óptimo del modelo. Ejemplos de detecciones de la red durante el entrenamiento de 250 y 500 épocas con *DataTr1*, se muestran en las Figuras-11 y 13.

Después del entrenamiento, ya teníamos el modelo listo para poder ser usado con imágenes de test para su validación y predicción, lo que nos permitía determinar la eficacia del modelo en situaciones reales y desconocidas utilizando el conocimiento adquirido de las imágenes etiquetadas en las bases de datos.

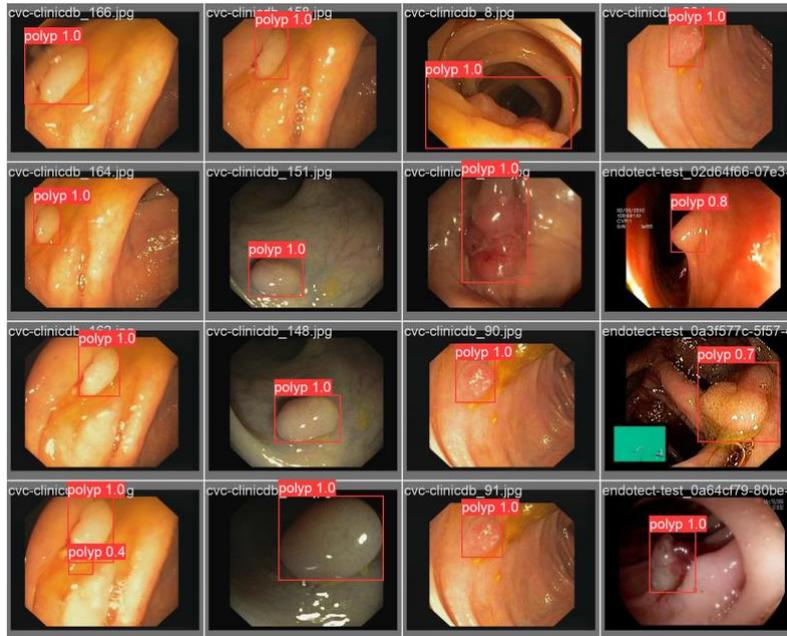


Figura 11: Ejemplos de detecciones de la red durante el entrenamiento de 250 épocas con *DataTrl*.

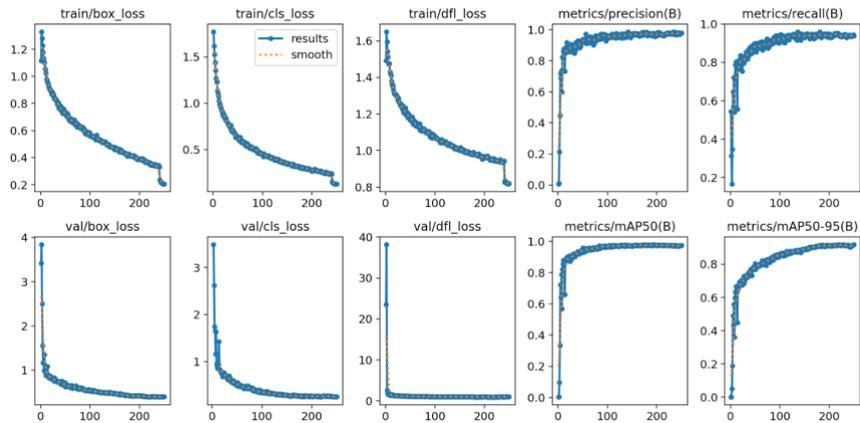


Figura 12: Métricas de rendimiento del entrenamiento de 250 épocas con *DataTrl*

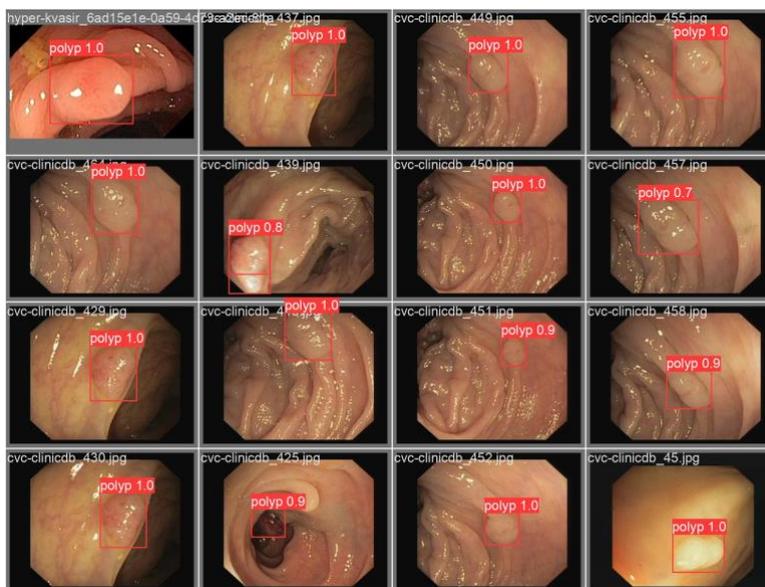


Figura 13: Ejemplos de detecciones de la red durante el entrenamiento de 500 épocas con DataTr7

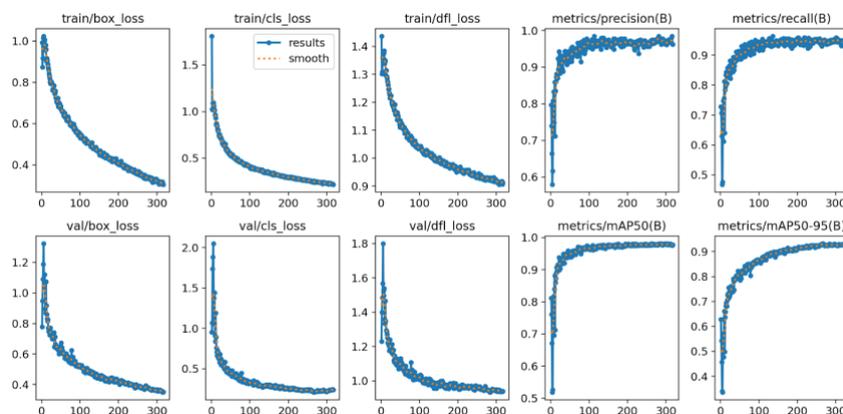


Figura 14: Métricas de rendimiento del entrenamiento de 500 épocas con DataTr7

En resumen, el proceso de entrenamiento de la red neuronal para la detección de pólipos implicó la experimentación con diferentes números de épocas para determinar el número óptimo de iteraciones. Este enfoque permitió ajustar el proceso de entrenamiento para maximizar el rendimiento del modelo y garantizar su capacidad de generalización a nuevos datos.

## 4.4. Hiperparámetros

En una red neuronal convolucional (CNN), los hiperparámetros son valores que se configuran antes del proceso de entrenamiento y que tienen un impacto significativo en el rendimiento y comportamiento del modelo. Los hiperparámetros son variables que controlan el comportamiento y la configuración del algoritmo de aprendizaje automático, entre otros tenemos: la tasa de aprendizaje (*learning rate*), el tamaño del lote (*batch size*), número de épocas (*epochs*), tasa de abandono (*Dropout Rate*), la arquitectura de la red, etc. Durante el proceso de entrenamiento de la red neuronal para la detección de pólipos, se exploraron y ajustaron diversos hiperparámetros con el objetivo de optimizar el rendimiento del modelo.

Uno de los hiperparámetros clave que se ajustaron fue la tasa de aprendizaje (*learning rate*), que determina la magnitud de los cambios realizados en los pesos de la red durante el entrenamiento.

Se experimentó con diferentes valores de tasa de aprendizaje para encontrar el valor óptimo que permitiera una convergencia rápida y estable del modelo.

Además de la tasa de aprendizaje, se ajustaron otros hiperparámetros relacionados con la arquitectura de la red, como el número de capas convolucionales, el tamaño de los filtros, y el número de neuronas en las capas ocultas. Estos ajustes se realizaron con el objetivo de optimizar la capacidad de extracción de características de la red y mejorar su capacidad de generalización a nuevos datos.

Otro hiperparámetro importante que se ajustó fue el tamaño del lote (*batch size*), que determina la cantidad de ejemplos de entrenamiento que se utilizan en cada iteración del proceso de entrenamiento. Se exploraron diferentes tamaños de lote para encontrar el equilibrio entre la estabilidad del entrenamiento y la eficiencia computacional del proceso.

También se realizaron ajustes en otros aspectos del proceso de entrenamiento, como la función de pérdida utilizada, que mide la discrepancia entre la predicción de un algoritmo de aprendizaje automático y el resultado que se requiere, las técnicas de regularización aplicadas y los métodos de inicialización de pesos de la red. El ajuste de hiperparámetros se realizó de manera sistemática, utilizando técnicas como la búsqueda en cuadrícula (*grid search*) y la búsqueda aleatoria (*random search*), para explorar el espacio de hiper parámetros y encontrar la configuración óptima que maximizara el rendimiento del modelo en términos de precisión, Recall y *F1-score*.

En resumen, el ajuste de hiperparámetros fue un componente crucial del proceso de entrenamiento de la red neuronal para la detección de pólipos. Al optimizar los hiperparámetros, se logró mejorar significativamente el rendimiento del modelo y su capacidad de generalización a nuevos datos, lo que resultó en un modelo más preciso y robusto en la detección de pólipos en imágenes médicas.

#### 4.4.1. Ajuste de hiperparámetros de Ultralytics Yolo

El ajuste de hiperparámetros es una etapa crítica en la optimización de modelos de aprendizaje profundo. En este trabajo, se ha empleado la función `tune` de *Ultralytics YOLO* para llevar a cabo este proceso, con el objetivo de mejorar el rendimiento del modelo en términos de precisión y eficiencia. Los hiperparámetros por defecto utilizados por *YOLO* en *Ultralytics* son los mostrados en la Tabla-2.

Hiperparámetro	valor	Hiperparámetro	valor
Learning Rate (lr0):	0.01	FL Gamma:	0.0
Learning Rate Final (lrf):	0.2	Hsv-H:	0.015
Momentum:	0.937	Hsv-S:	0.7
Weight Decay:	0.0005	Hsv-V:	0.4
Warmup Epochs:	3.0	Degrees:	0.0
Warmup Momentum:	0.8	Translate:	0.1
Warmup Bias LR:	0.1	Scale:	0.5
Box:	0.05	Shear:	0.0
Cls:	0.5	Perspective:	0.0
Cls Pw:	1.0	Flipud:	0.0
Obj:	1.0	Fliplr:	0.5
Obj Pw:	1.0	Mosaic:	1.0
IoU T:	0.2	Mixup:	0.0
Anchor T:	4.0		

Tabla 2: Hiperparámetros por defecto que YOLO utiliza para entrenar.



**Configuración de la Función `tune`:** Estas configuraciones permiten una exploración exhaustiva del espacio de hiperparámetros, asegurando que se evalúen múltiples combinaciones para encontrar la óptima. Los hiperparámetros que YOLO ha elegido como óptimos para el entrenamiento son los mostrados en la Figura 15. Para el ajuste de los hiperparámetros, se han configurado las siguientes especificaciones:

- Iteraciones: 300
- Épocas: 250

```
lr0: 0.00141
lrf: 0.01084
momentum: 0.71268
weight_decay: 0.00058
warmup_epochs: 2.04313
warmup_momentum: 0.24702
box: 5.6203
cls: 0.67431
dfl: 1.27657
hsv_h: 0.01362
hsv_s: 0.31591
hsv_v: 0.67931
degrees: 0.0
translate: 0.12148
scale: 0.23915
shear: 0.0
perspective: 0.0
flipud: 0.0
fliplr: 0.28361
mosaic: 0.43696
mixup: 0.0
copy_paste: 0.0
```

Figura 15: hiperparámetros seleccionados por la función `Tune` para el mejor resultado del entrenamiento

El proceso de ajuste de hiperparámetros se llevó a cabo de la siguiente manera:

- 1. Inicialización:** Se definieron los rangos iniciales para cada hiperparámetro basado en valores comunes y recomendaciones de la literatura.
- 2. Iteraciones de ajuste:** Durante 300 iteraciones, se probaron diferentes combinaciones de hiperparámetros, utilizando una técnica de búsqueda aleatoria o basada en un algoritmo específico.
- 3. Evaluación del rendimiento:** En cada iteración, el rendimiento del modelo se evaluó utilizando métricas clave como la precisión, la precisión promedio (AP) y la pérdida del modelo.
- 4. Selección de los mejores hiperparámetros:** Al finalizar las 300 iteraciones, se seleccionaron los hiperparámetros que ofrecieron el mejor rendimiento general.

Al comparar los valores de los hiperparámetros utilizados por defecto (Tabla-2) por YOLO con los ajustados por la función `tune` (Figura 15), se observan varios cambios significativos:

- 1. Learning Rate (`lr0`) y Learning Rate Final (`lrf`):** Se han reducido, lo cual puede llevar a un entrenamiento más estable y una mejor convergencia.
- 2. Momentum y Warmup Momentum:** Disminuyeron, posiblemente para una actualización más gradual de los pesos, reduciendo el riesgo de sobreajuste.
- 3. Warmup Epochs:** Ligeramente reducido, permitiendo que el modelo comience a aprender a un ritmo normal más rápidamente.
- 4. Box, Cls, y DFL:** Aumentaron considerablemente, lo que puede mejorar la precisión en la predicción de las cajas delimitadoras o BBox y la clasificación.
- 5. Hsv-H, Hsv-S, y Hsv-V:** Ajustados para una mejor augmentación de datos en términos de color.
- 6. Translate y Scale:** Se ajustaron para modificar la intensidad de la augmentación de datos, potencialmente mejorando la robustez del modelo.

**7. Flipud y Fliplr:** Reducción en la probabilidad de volteo, lo cual puede reducir el ruido en la augmentación de datos.

**8. Mosaic:** Reducción de la probabilidad de usar mosaico, lo cual puede llevar a un entrenamiento más estable.

Estos cambios optimizados han sido posibles gracias a la exploración exhaustiva del espacio de hiperparámetros, asegurando que se evalúen múltiples combinaciones para encontrar la óptima, mejorando así la precisión y eficiencia del modelo.

## 4.5. Validación cruzada (*Cross validation*)

La validación cruzada es una técnica fundamental en el desarrollo y evaluación de modelos de aprendizaje automático [24]. Consiste en dividir el conjunto de datos en  $k$  subconjuntos mutuamente excluyentes, llamados pliegues o *folds*, y realizar  $k$  iteraciones del proceso de entrenamiento y evaluación del modelo. En cada iteración, se utiliza un pliegue diferente como conjunto de prueba (conjunto de datos de validación), mientras que los restantes se utilizan como conjunto de entrenamiento.

El parámetro " $k$ " representa el número de pliegues en los que se divide el conjunto de datos. Es importante destacar que " $k$ " es un hiperparámetro de la validación cruzada y debe elegirse cuidadosamente dependiendo del tamaño del conjunto de datos y la naturaleza del problema. Un valor comúnmente utilizado para " $k$ " es 5 ó 10, pero en nuestro caso hemos optado por utilizar  $k = 7$ . En la figura 16, se representa la validación cruzada con 5-*folds*, los *folds* 1-4 se convierten en el conjunto de entrenamiento y por ejemplo, el *fold* 5 (en amarillo) se denomina *fold* de validación y se utiliza para ajustar los hiperparámetros.

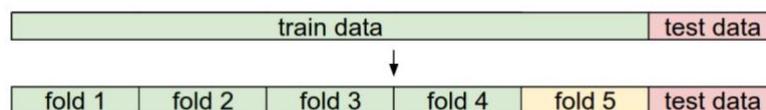


Figura 16: Representación de la validación cruzada con  $k=5$

La validación cruzada con  $k = 7$  nos proporciona varias ventajas. En primer lugar, nos permite obtener una estimación más precisa y confiable del rendimiento del modelo, al evaluarlo en múltiples conjuntos de datos diferentes. Esto es especialmente útil cuando el conjunto de datos es relativamente pequeño o cuando queremos garantizar que el rendimiento del modelo sea consistente en diferentes subconjuntos de datos. Además, la validación cruzada nos permite aprovechar al máximo nuestros datos de entrenamiento y validación, ya que cada instancia del conjunto de datos se utiliza tanto para entrenamiento como para evaluación en diferentes iteraciones. Esto ayuda a reducir el sesgo en la estimación del rendimiento del modelo y a mitigar el riesgo de sobreajuste o sub-ajuste.

En nuestro proyecto, utilizaremos la validación cruzada con  $k = 7$  para evaluar el rendimiento de nuestro modelo de detección de pólipos en imágenes médicas. Esto nos permitirá obtener una evaluación más completa y precisa de la capacidad de nuestro modelo para generalizar a nuevos datos y detectar pólipos de manera efectiva.

Cuando se realiza la validación cruzada, es importante tener en cuenta que cada iteración del proceso de entrenamiento y evaluación se lleva a cabo en un subconjunto diferente de los datos. Esto significa que debemos decidir cómo configurar el proceso de entrenamiento en cada iteración, incluyendo el número de épocas que se realizarán para entrenar el modelo en cada

subconjunto. En la práctica, se suele definir un número fijo de épocas para cada iteración del proceso de validación cruzada. Este número puede ser determinado mediante experimentación previa o utilizando técnicas de búsqueda de hiperparámetros. La elección del número de épocas dependerá de varios factores, como la complejidad del modelo, el tamaño del conjunto de datos y la disponibilidad de recursos computacionales.

Una vez que se ha decidido el número de épocas para cada iteración, se procede a realizar la validación cruzada. Durante cada iteración, se divide el conjunto de datos en  $k$  subconjuntos y se entrena el modelo utilizando  $k-1$  subconjuntos como conjunto de entrenamiento y el subconjunto restante como conjunto de validación. El proceso de entrenamiento se lleva a cabo durante el número de épocas especificado.

Además, durante el entrenamiento, es posible que se pasen pesos predeterminados al modelo en cada iteración para inicializarlo con una configuración específica. Esto puede ayudar a estabilizar el entrenamiento y acelerar la convergencia del modelo. Después de completar todas las iteraciones de la validación cruzada con  $k = 7$  y obtener los 7 conjuntos de pesos del modelo, así como sus archivos “.yaml” correspondientes, será necesario realizar la predicción con imágenes de prueba utilizando cada conjunto de pesos por separado. Este proceso implica cargar el modelo entrenado con los pesos específicos de cada iteración y evaluar su rendimiento en las imágenes de prueba.

Al realizar la predicción con cada conjunto de pesos del modelo, se obtendrán resultados diferentes para cada iteración de la validación cruzada. Estos resultados pueden variar en términos de métricas de rendimiento como precisión, Recall, *F1-score*, entre otras, dependiendo de la configuración específica de cada modelo y su capacidad para generalizar a nuevos datos.

Una vez que se han evaluado todas las iteraciones por separado, se puede comparar el rendimiento de cada modelo y determinar cuál de ellos es el óptimo en términos de su capacidad para detectar pólipos en imágenes médicas. Esta comparación nos permitirá identificar el modelo que presenta el mejor equilibrio entre rendimiento y generalización, lo que nos ayudará a tomar decisiones informadas sobre la configuración y ajuste del modelo.

En resumen, para evaluar y seleccionar el modelo óptimo, será necesario realizar la predicción con imágenes de prueba utilizando los 7 conjuntos de pesos del modelo generados durante la validación cruzada. Este proceso nos permitirá determinar cuál de todas las configuraciones de la red neuronal convolucional es la más adecuada para nuestra tarea específica de detección de pólipos en imágenes médicas.

# 5. Resultados

Para generar los resultados presentados, hemos utilizado imágenes proporcionadas por el Hospital La Fe, las cuales son imágenes originales de colonoscopias realizadas en sus instalaciones. Estas imágenes nos permiten probar las redes entrenadas y evaluar su precisión en un entorno clínico real. En la actualidad, contamos con un total de 250 imágenes propias, *DatasetFe1*, como se muestra en la Tabla-1 disponibles para llevar a cabo estas pruebas.

Para comprobar la evolución y mejora del proyecto tras la ampliación de la base de datos de *DataTr1* con 2108 imágenes a *DataTr2* con 2850 imágenes, se han obtenido resultados tanto de la base de datos original *DataTr1* como de la base de datos ampliada *DataTr2*. Las redes entrenadas con ambas bases de datos han sido sometidas a pruebas utilizando como imágenes de *test* el *DatasetFe1* para realizar predicciones. De esta manera, es posible evaluar cómo ha evolucionado la red con la ampliación de la base de datos.

El uso de imágenes reales de colonoscopias es fundamental para validar la efectividad de nuestras redes neuronales en situaciones clínicas reales. Estas imágenes nos proporcionan un conjunto diverso y representativo de casos, lo que nos permite evaluar la capacidad de las redes para detectar y clasificar pólipos con precisión.

## 5.1. Comparación de resultados

Para comparar los resultados obtenidos en la validación cruzada y en el conjunto de datos de prueba, se utilizarán diversas métricas, incluyendo precisión, recall y F1-score. Estas métricas nos proporcionarán una visión completa del rendimiento del modelo en términos de su capacidad para identificar correctamente los pólipos y minimizar los falsos positivos. Además, se analizarán las métricas mAP50 y mAP75, que son medidas específicas de precisión en la detección de objetos, con un énfasis en la detección de pólipos en imágenes endoscópicas.

En cada iteración de la validación cruzada, se evaluarán estas métricas para determinar la variabilidad del rendimiento del modelo en diferentes subconjuntos de datos. Posteriormente, se compararán los resultados obtenidos en la validación cruzada con los del conjunto de datos de prueba para identificar la consistencia del rendimiento del modelo en diferentes escenarios.

El objetivo final es seleccionar el modelo con el mejor rendimiento en términos de las métricas mencionadas anteriormente. Esto nos permitirá identificar la configuración óptima del modelo y determinar su capacidad para la detección precisa de pólipos en imágenes endoscópicas.

### 5.1.1. Comparación de resultados de Cross Validation

Para evaluar el rendimiento del modelo de manera exhaustiva, se llevó a cabo una validación cruzada con  $k = 7$ . A continuación, se presentan los resultados obtenidos en cada iteración de esta validación, incluyendo las métricas de precisión, recall, *F1-score* y cualquier otra métrica relevante:

	<b>Class</b>	<b>Images</b>	<b>Instances</b>	<b>Box(P)</b>	<b>R</b>	<b>mAP50</b>	<b>m</b>
<b>SPLIT 1</b>	all	589	633	0,958	0,920	0,963	0,87
<b>SPLIT 2</b>	all	589	633	0,955	0,921	0,958	0,887
<b>SPLIT 3</b>	all	589	633	0,974	0,927	0,967	0,88
<b>SPLIT 4</b>	all	589	633	0,967	0,921	0,963	0,87
<b>SPLIT 5</b>	all	589	633	0,977	0,928	0,972	0,893



<b>SPLIT 6</b>	all	589	633	0,968	0,927	0,965	0,896
<b>SPLIT 7</b>	all	589	633	0,958	0,923	0,965	0,886

Tabla 3: Resultados de entrenamiento con *Cross Validation del SPLIT 1 al 7*

## 5.1.2. Comparación de resultados de ajuste de hiperparámetros

Para elaborar el apartado considerando los resultados de la red entrenada con el ajuste de hiperparámetros, se puede realizar un análisis detallado de los resultados obtenidos tanto en el conjunto de datos de validación como en el conjunto de datos de prueba. Aquí se describe el proceso y los resultados del entrenamiento en la Figura 23:

```

val: Scanning /home/acustuna/Escritorio/pablo-escritorio/yolov5/datasets/endotec
Class  Images  Instances  Box(P)  R  mAP50  m
all    589      633      0.973  0.921  0.96  0.899

```

Figura 23: resultados de predicción con red con sobreajuste de hiperparámetros

Los valores de los hiperparámetros finales con los cuáles el entrenamiento de la red ha sacado los mejores resultados son:

**Precisión (Box(P)):** 0.973

**Recall (R):** 0.921

**mAP50:** 0.96

**mAP:** 0.899

Estos resultados indican que el modelo tiene un alto rendimiento en el conjunto de validación, con una precisión y un recall elevados, así como un mAP que refleja una capacidad consistente para detectar y localizar correctamente las instancias en las imágenes.

A continuación, se realiza un análisis de los resultados en función de valor obtenido para cada parámetro:

**Precisión:** Un valor de 0.973 en precisión sugiere que el modelo es muy bueno para identificar correctamente las instancias detectadas. Esto implica pocas falsas alarmas.

**Recall:** Un valor de 0.921 en recall indica que el modelo es eficaz en la detección de la mayoría de las instancias relevantes.

**mAP50:** Un valor de 0.96 en mAP50 muestra una excelente capacidad del modelo para localizar objetos con una precisión alta a un umbral de *IoU* del 50%.

**mAP:** Un valor de 0.899 en mAP refleja un rendimiento robusto del modelo en la localización precisa de objetos a través de un rango de umbrales de *IoU*.

Se puede concluir que el ajuste de hiperparámetros y el entrenamiento del modelo con el conjunto de datos proporcionado han resultado en una red con un desempeño mejorado, tanto en términos de precisión como de recall. Un mAP de 0,899 confirma que el modelo es capaz de realizar detecciones precisas y consistentes.

## 5.1.3. Resultados en el conjunto de datos de prueba

Además de la validación cruzada, se evaluó el rendimiento del modelo en un conjunto de datos de prueba independiente, en este caso tenemos resultados del entrenamiento con la base de datos *DataTr1* y *DataTr2*. A continuación, en la Tabla 4, se muestran las métricas de rendimiento obtenidas en este conjunto de datos:

Dataset	Epochs	Class	Images Validacion	Instances	Box(P)	R	mAP50	m
DataTr1	250	all	424	452	0,977	0,965	0,98	0,935
DataTr1	500	all	424	452	0,964	0,965	0,983	0,952
DataTr2	250	all	589	633	0,952	0,923	0,962	0,88
DataTr2	500	all	589	633	0,976	0,921	0,961	0,88

Tabla 4: Resultado de entrenamiento con 250 y 500 épocas con el DataTr1 y DataTr2.

## Análisis de resultados

Tabla de resultados de la fase de entrenamiento y validación:

Dataset	IMÁGENES	EPOCH	ITERATION	IM_TRAIN	IM_VAL	ID-RED	FUNCIÓN USADA
DataTr1	2108	250	X	1684	424	T1	X
DataTr1	2108	500	X	1684	424	T2	X
DataTr2	2850	250	X	2261	589	T3	X
DataTr2	2850	500	X	2261	589	T4	X
DataTr2	2850	250	7	2261	589	T5	CROSS VALIDATION
DataTr2	2850	250	300	2261	589	T6	TUNE

Tabla 5: Tabla de resultados de la fase de entrenamiento y validación con DataTr1 y DataTr2.

Tabla de resultados de la fase de predicción, siendo:

- \*T1: Pesos del entrenamiento con 250 épocas con DataTr1
- \*T2: Pesos del entrenamiento con 500 épocas con DataTr1
- \*T3: Pesos del entrenamiento con 250 épocas con DataTr2
- \*T4: Pesos del entrenamiento con 500 épocas con DataTr2
- \*T6: Pesos del entrenamiento con *Cross Validation*
- \*T5: Pesos del entrenamiento con ajuste de hiperparámetros

ID-RED	IMÁGENES	TOTAL PÓLIPOS	DETECTADOS	FALSOS POSITIVOS	RECALL	PRECISIÓN
T1	180	401	95	6	0,237	0,941
T2	180	401	64	4	0,16	0,941
T3	180	401	137	5	0,342	0,965
T4	180	401	120	7	0,299	0,945
T5	180	401	116	6	0,289	0,951
T6	180	401	124	4	0,309	0,969

Tabla 6: Tabla de resultados de la fase de predicción y test con el DatasetFe1.

TIPO	IMÁGENES	TOTAL PÓLIPOS	DETECTADOS	FALSOS POSITIVOS	RECALL	PRECISIÓN
SERRADOS	44	95	58	3	0,713	0,968
PSEUDOPÓLIPOS	119	216	99	3	0,4507	0,9697
ADENOMAS	15	21	9	1	0,4	0,889
HIPERPLÁSICOS	2	2	2	0	1	1

Tabla 7: Tabla de resultados de la fase de predicción y test con el DatasetFe1 por clases con la ID-¿???

Para determinar cuál de las configuraciones tiene el mejor rendimiento, es crucial analizar detenidamente las métricas proporcionadas y considerar cómo afectan al objetivo general del



modelo de detección de pólipos. Aquí hay un análisis más detallado de cada configuración y su rendimiento:

### Configuración de 250 épocas

- **Ventajas:** Altos valores de mAP50 y mAP75 indican una buena precisión en la detección de pólipos en la imagen. Bajo número de falsos positivos, lo que sugiere una alta capacidad para evitar detectar objetos incorrectos.
- **Desventajas:** Bajo recall, lo que indica que el modelo podría estar perdiendo algunas detecciones importantes.
- **Desarrollo:** Esta configuración muestra una buena precisión en la detección de pólipos, con valores altos de mAP50 y mAP75. Sin embargo, el bajo recall podría ser preocupante, ya que implica que el modelo puede estar perdiendo algunas instancias de pólipos en la imagen.

### Configuración de 500 épocas

- **Ventajas:** Mejora marginal en mAP75 en comparación con la configuración de 250 épocas, lo que sugiere una convergencia más sólida del modelo. Menor cantidad de falsos positivos en comparación con la configuración de 250 épocas.
- **Desventajas:** **Reducción** significativa en el recall en comparación con la configuración de 250 épocas.
- **Desarrollo:** Aunque hay mejoras en mAP75 y en la cantidad de falsos positivos en comparación con la configuración de 250 épocas, la disminución en el recall podría indicar que el modelo está perdiendo algunas detecciones importantes.

### Configuración de TUNE

- **Ventajas:** Valores de mAP50 y mAP75 competitivos, similares a las configuraciones de 250 y 500 épocas. Baja cantidad de falsos positivos. Buen recall y precisión.
- **Desventajas:** Los valores de mAP50 y mAP75 son ligeramente inferiores a los de la configuración de 500 épocas.
- **Desarrollo:** Esta configuración muestra un equilibrio notable entre mAP50, mAP75, falsos positivos, recall y precisión. Aunque los valores de mAP50 y mAP75 son ligeramente inferiores a los de la configuración de 500 épocas, el equilibrio general de métricas podría indicar una mejor capacidad del modelo para detectar pólipos de manera precisa y sin exceso de falsos positivos.

Se puede concluir que, aunque todas las configuraciones tienen sus propias fortalezas y debilidades, la configuración de *TUNE* parece ofrecer el mejor equilibrio entre precisión, recall y cantidad de falsos positivos. Su capacidad para mantener una precisión competitiva mientras mantiene un buen balance entre recall y falsos positivos la hace una opción sólida para la detección de pólipos en imágenes médicas.

### Ejemplos de predicciones de la red entrenada:

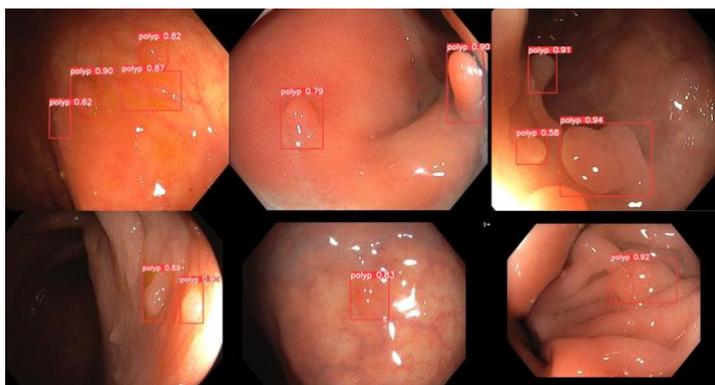


Figura 28: ejemplos predicciones de la red entrenada

## 5.2. Análisis de resultados según la clase

En este apartado se evaluará el desempeño del modelo de detección de pólipos al considerar cuatro clases distintas: *pólipos serrados*, *pseudopólipos*, *adenomas* e *hiperplásicos*. El objetivo es determinar cuál de estas clases presenta el mayor nivel de acierto en términos de detecciones realizadas por el modelo y analizar su rendimiento en base a métricas clave como precisión y recall. Cabe destacar que, por el momento, el análisis de los resultados según la clase de pólipo no se realiza con el fin de crear una red que clasifique los pólipos, sino que nos facilita comprender con qué clases tiene más capacidad de detección y con cuáles menos.

### 1. Análisis de los Datos:

- **Distribución de las Clases:** Observamos que la clase "Pseudopólipos" tiene la mayor cantidad de instancias (216), seguida por "Pólipos Serrados" (95), "Adenoma" (21) y "Hiperplásico" (2). Esta distribución desigual podría influir en la evaluación del rendimiento del modelo.

### 2. Métricas de Rendimiento por Clase:

Clase	Instancias	Detecciones	Falsos Positivos	Recall	Precisión de Clase
Pseudopólipo	216	99	3	0,4507	0,97
Hiperplásico	2	2	0	1	1
Serrado	95	58	3	0,713	0,95
Adenoma	21	9	1	0,4	0,9

Tabla 8: Predicción por clases y métricas de rendimiento, obtenida con una red con  $mAP50=0,966$ ,  $mAP75 = 0,903$  y  $R = 0,926$ .

### 3. Interpretación de Resultados:

- Aunque "Pseudopólipos" tiene el mayor número de detecciones, su recall es relativamente bajo en comparación con "Hiperplásico" y "Pólipos Serrados". Sin embargo, su precisión es la más alta entre todas las clases.
- "Hiperplásico" y "Pólipos Serrados" tienen un recall más alto que "Pseudopólipos" y "Adenoma", lo que indica que el modelo identificó una mayor proporción de las instancias reales de estas clases.

- "Hiperplásico" tiene el mejor rendimiento en términos de precisión y recall, pero su bajo número de detecciones podría sugerir que esta clase tiene muy poca representación en la base de datos.

Dada la distribución desigual de las clases y las métricas de rendimiento obtenidas, la clase "Pseudopólipos" parece ser la mejor opción en términos de detecciones totales y precisión. Sin embargo, si se valora más el recall sobre la precisión, "Hiperplásico" y "Pólipos Serrados" podrían ser considerados mejores opciones, dependiendo de las prioridades específicas del proyecto y del contexto clínico. Es importante tener en cuenta la distribución de clases y las características únicas de cada una al evaluar el rendimiento del modelo en la detección de pólipos. A continuación, en las Figuras 30, 31 y 32 se muestra algunos ejemplos de detecciones de la red por clases.

### EJEMPLOS DE DETECCIONES POR CLASES



Figura 30: Detecciones de pólipos de tipo adenoma

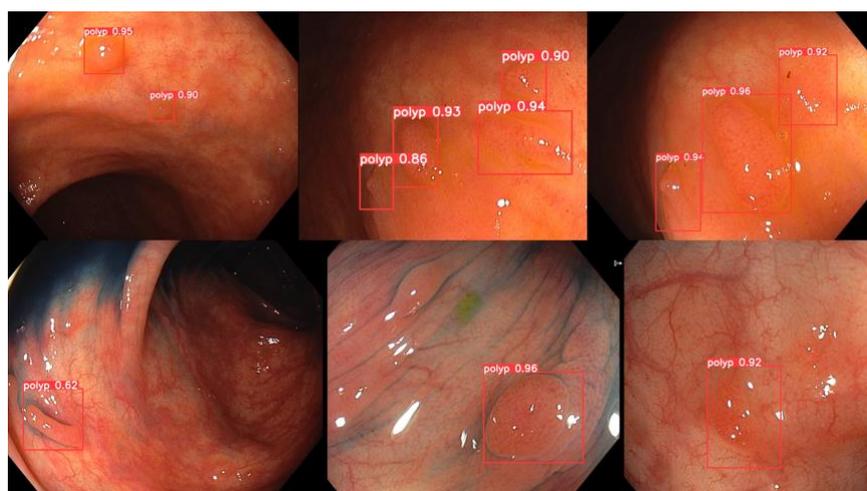


Figura 31: Detecciones de pólipos de tipo pseudopólipo

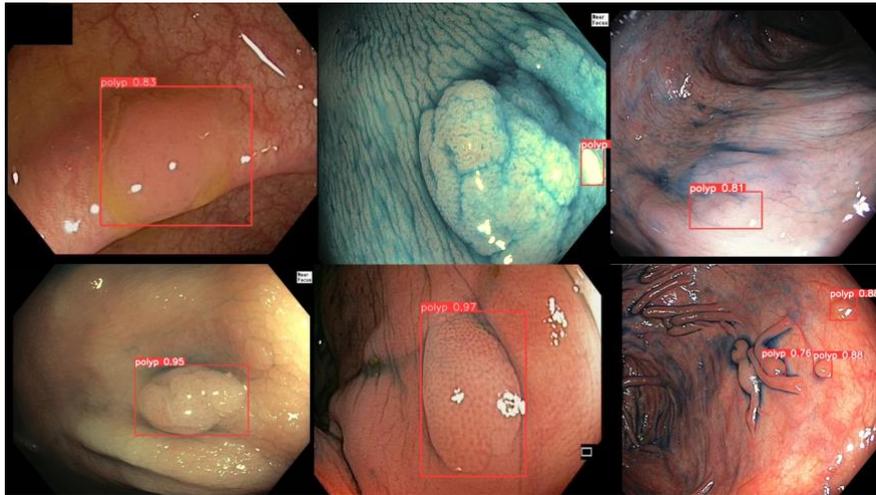


Figura 32: Detecciones de pólipos de tipo serrado

## 5.3. Últimas pruebas

Existen algunas imágenes en las que es muy complicado detectar las instancias de los diferentes pólipos, ya que estos están muy juntos entre sí o son muy difíciles de apreciar. Esto, sumado a que no disponemos de una base de datos lo suficientemente grande como para mejorar a este nivel la precisión, se ha decidido hacer una prueba con la red entrenada, eliminando las imágenes donde exista una excesiva acumulación de pólipos en una zona, en concreto, 19 imágenes eliminadas.

Estas 19 imágenes eliminadas hacen que el total de instancias disminuya de las 401 iniciales a 264. A continuación, damos un nombre a los pesos que se han usado de los diferentes entrenamientos para la fase de la predicción, siendo  $T_x$  el peso usado para la fase de predicción de esta última parte.

- \*T1: Pesos del entrenamiento con 250 épocas
- \*T2: Pesos del entrenamiento con 500 épocas
- \*T3: Pesos del entrenamiento con ajuste de hiperparámetros
- \*T4: Pesos del entrenamiento con *Cross Validation*

Tabla de resultados de la fase de predicción sobre nuestra base de datos propia DataSetFe1 completa, es decir, habiendo eliminado las 19 imágenes.

ID-RED	IMÁGENES	TOTAL PÓLIPOS	DETECTADOS	FALSOS POSITIVOS	RECALL	PRECISIÓN
T1	161	264	111	4	0,4205	0,9652
T2	161	264	105	9	0,3977	0,9211
T3	161	264	97	5	0,3674	0,9509
T4	161	264	70	5	0,2652	0,9333

Tabla 9: resultados de la fase de predicción sobre DataSetFe1 con la reducción de 19 imágenes

Tabla de resultados de la fase de predicción sobre nuestra base de datos propia DataSetFe1 completa, es decir, sin haber eliminado las 19 imágenes.

ID-RED	IMÁGENES	TOTAL PÓLIPOS	DETECTADOS	FALSOS POSITIVOS	RECALL	PRECISIÓN
T1	180	401	137	5	0,342	0,965
T2	180	401	120	7	0,299	0,945
T3	180	401	124	4	0,309	0,969

<b>T4</b>	180	401	116	6	0,289	0,951
-----------	-----	-----	-----	---	-------	-------

Tabla 10: resultados de la fase de predicción sobre DataSetFe1 sin la reducción de 19 imágenes

Como vemos, la primera tabla con las imágenes eliminadas tiene un mayor recall en casi todos los casos. Esto es debido a que, al haber eliminado imágenes con instancias complicadas de detectar, la red es capaz de detectar mayor porcentaje de instancias sobre el total, aumentando así el recall. Sin embargo, la precisión, o no mejora o empeora; esto se debe a que los falsos positivos que antes detectaba, los sigue detectando, puesto que muchas veces estos falsos positivos son confusiones con reflejos de luz u objetos o textos que se ven en algunas imágenes.

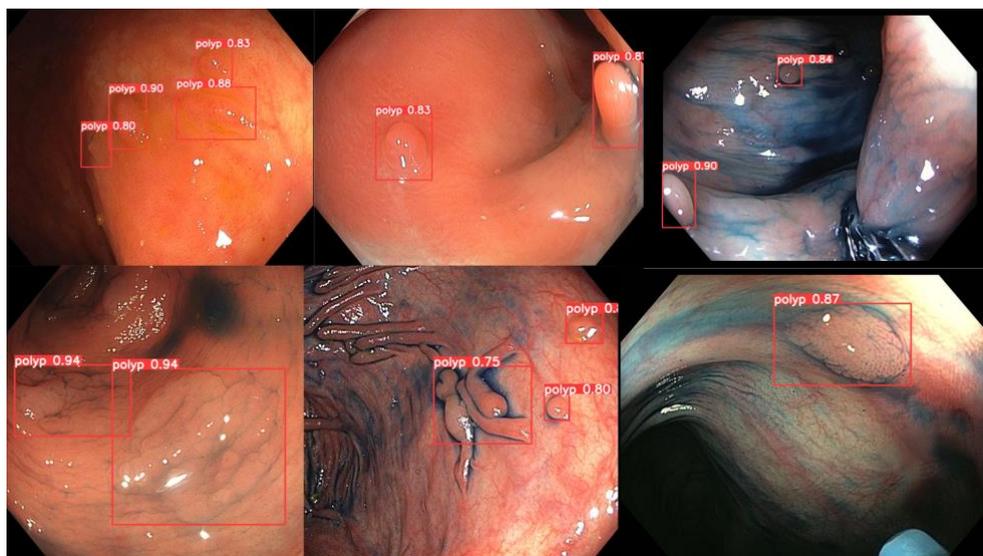


Figura 33: Ejemplo de pólipos detectados en la fase de predicción

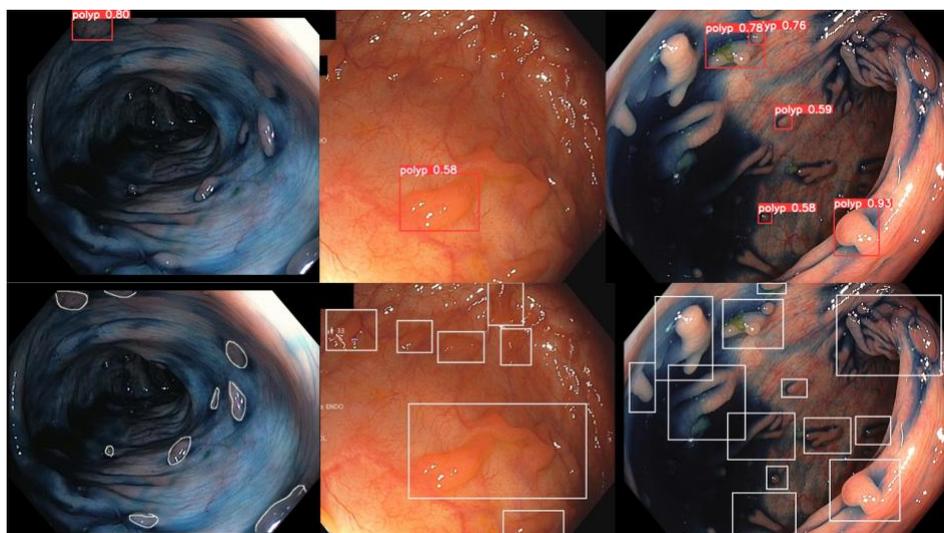


Figura 34: Ejemplo de pólipos no detectados o detectados parcialmente en la fase de predicción

## 6. Conclusiones

---

En este Trabajo de Fin de Grado se ha abordado el desafío de mejorar las prácticas de detección y tratamiento de enfermedades gastrointestinales, con un enfoque particular en el cáncer de colon, a través del desarrollo de un *software* basado en redes neuronales para la detección en tiempo real de pólipos durante colonoscopias. Este proyecto se ha llevado a cabo en colaboración con el Hospital Universitario y Politécnico La Fe de Valencia, con el objetivo de proporcionar una herramienta avanzada que no solo acelere la identificación de pólipos, sino que también mejore la precisión, especialmente en aquellos casos con características indicativas de displasia.

La motivación detrás de este proyecto se basa en la necesidad de un enfoque más eficiente en la detección de pólipos, dada su estrecha relación con el desarrollo de cáncer de colon. La colaboración con el Hospital La Fe no solo proporcionó una perspectiva clínica práctica, sino que también garantizó la aplicabilidad efectiva de los resultados en el entorno médico real.

El diseño, desarrollo e implementación del *software* se centró en varios objetivos principales, incluyendo la creación de un *software* robusto, su integración en el flujo de trabajo clínico existente, la priorización de la detección de pólipos con displasia, la colaboración estrecha con profesionales de la salud, y la contribución a la toma de decisiones clínicas.

Los resultados obtenidos mediante la evaluación del modelo desarrollado destacan la eficacia del mismo en la detección de pólipos durante colonoscopias. El ajuste de hiperparámetros demostró el mejor rendimiento general, con una alta precisión y valores significativos de mAP50 y mAP75. Además, las evaluaciones independientes después de 250 y 500 épocas mostraron un rendimiento consistente, indicando que el modelo alcanzó un punto de convergencia donde el aumento adicional en el número de épocas de entrenamiento no resultó en mejoras sustanciales en la precisión de la detección.

En conclusión, este proyecto representa un paso significativo hacia una atención médica más precisa y orientada a resultados, respaldando la toma de decisiones clínicas informadas y mejorando sustancialmente la eficacia de los procedimientos de detección y prevención de enfermedades gastrointestinales. El *software* desarrollado tiene el potencial de transformar positivamente la detección y prevención de enfermedades, contribuyendo así al Objetivo de Desarrollo Sostenible número tres de salud y bienestar.

### 6.1. Trabajo futuro

---

El desarrollo del software de detección de pólipos basado en redes neuronales ha marcado un avance significativo en las prácticas de diagnóstico durante las colonoscopias. Sin embargo, la evolución de esta herramienta no termina con su implementación inicial. A medida que el proyecto avanza, se presentan múltiples oportunidades para expandir y mejorar su aplicación y eficacia. En esta sección, se explorarán varias direcciones prometedoras para la investigación y el desarrollo futuros, que ya están en marcha.

#### **Ampliación de la Base de Datos**

Una de las áreas clave para futuras investigaciones sería la ampliación significativa de la base de datos utilizada para el entrenamiento del modelo. Esto implicaría colaboraciones más estrechas con hospitales y otras organizaciones médicas para recopilar una mayor variedad y cantidad de imágenes endoscópicas de pólipos. La diversidad en la base de datos es crucial para mejorar la capacidad de generalización del modelo y su efectividad en diferentes escenarios clínicos.

Además, la inclusión de imágenes de pacientes con una variedad de condiciones médicas y características demográficas podría mejorar aún más la robustez del modelo y su capacidad para adaptarse a casos más diversos.

### **Implementación del *Software* en Entornos Clínicos**

Una vez que el modelo de detección de pólipos esté entrenado y validado adecuadamente, el siguiente paso sería la implementación del *software* en entornos clínicos reales. Esto requeriría el desarrollo de una aplicación que pueda integrarse fácilmente en los sistemas existentes de colonoscopia utilizados en hospitales y clínicas. Además, se necesitaría una cuidadosa validación y evaluación del rendimiento del *software* en entornos clínicos para garantizar su eficacia y seguridad en situaciones reales.

### **Integración de Funcionalidades Adicionales**

Además de la detección de pólipos, el *software* podría beneficiarse de la integración de funcionalidades adicionales que mejoren su utilidad y versatilidad en entornos clínicos. Esto podría incluir características como la segmentación de imágenes para resaltar áreas de interés, la generación de informes automáticos para facilitar la documentación clínica y la integración con sistemas de información médica para un manejo eficiente de los datos del paciente. La incorporación de estas funcionalidades adicionales podría mejorar aún más la eficacia y la usabilidad del *software* en entornos clínicos reales.

# Bibliografía

- [1] *American College of Gastroenterology*, “Pólipos de Colon” 2010. [Online]. Available: <https://gi.org/patients/recursos-en-espanol/polipos-del-colon/>
- [2] Instituto Universitario de Automática e Informática Industrial (AI2) de la Universitat Politècnica de València. <https://www.ai2.upv.es/>
- [3] *American Cancer Society*, “Your Colon or Rectal Pathology Report: Polyps (Including Serrated Adenomas)” 2023. [Online]. Available: <https://www.cancer.org/cancer/diagnosis-staging/tests/biopsy-and-cytology-tests/understanding-your-pathology-report/colon-pathology/colon-polyps-sessile-or-traditional-serrated-adenomas.html>
- [3] Módulo de endoscopia inteligente *GI Genius*. [Online]. Available: <https://www.medtronic.com/covidien/es-es/products/gastrointestinal-artificial-intelligence/gi-genius-intelligent-endoscopy.html#>
- [4] Medtronic: desarrollador de *GI Genius*. [Online]. Available: <https://www.medtronic.com/es-es/our-company.html>
- [5] Savino, A., Rondonotti, E., Rocchetto, S., Piagnani, A., Bina, N., & Di Domenico, P. (2024). *GI Genius endoscopy module: a clinical profile*. Retrieved from <https://doi.org/10.1080/17434440.2024.2342508>
- [6] Van der Zander QEW, Schreuder RM, Fonollà R, Scheeve T, van der Sommen F, Winkens B, Aepli P, Hayee B, Pischel AB, Stefanovic M, Subramaniam S, Bhandari P, de With PHN, Masclee AAM, Schoon EJ “ *Optical diagnosis of colorectal polyp images using a newly developed computer-aided diagnosis system (CADx) compared with intuitive optical diagnosis*”. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/33368056/>
- [7] Byrne, M. F., et al. (2018). *Deep Learning for Automatic Polyp Detection and Classification in Endoscopy: A Review*. *IEEE Reviews in Biomedical Engineering*, 11, 1-18.
- [8] ElKarazle, K., Raman, V., Then, P., & Ch, C. 2023. " *Detection of Colorectal Polyps from Colonoscopy Using Machine Learning: A Survey on Modern Techniques*".
- [9] Rahim, T., Hassan, S. A., & Shin, S. Y., 2021. " *A deep convolutional neural network for the detection of polyps in colonoscopy images* ". *Biomedical Signal Processing and Control*, Volume 68, 2021, <https://doi.org/10.1016/j.bspc.2021.102654>. (<https://www.sciencedirect.com/science/article/pii/S1746809421002512>)
- [10] Dulf, I., Bleda, M., Mocan, T., & Mocan, L. (2021). *Transfer Learning with Deep Convolutional Neural Network for Polyp Classification*. *Sensors*, 21(17), 5704. <https://doi.org/10.3390/s21175704>
- [11] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You Only Look Once: Unified, Real-Time Object Detection*. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).
- [12] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). *YOLOv4: Optimal Speed and Accuracy of Object Detection*. arXiv preprint arXiv:2004.10934.
- [13] Wang, X., & Bochkovskiy, A. (2021). *YOLOv5: A native PyTorch implementation of the YOLO algorithm*. *GitHub Repository*. Retrieved from <https://github.com/ultralytics/yolov5>
- [14] Pacal, I., Karaman, A., Karaboga, D., Akay, B., Basturk, A., Nalbantoglu, U., & Coskun, S. (2022). *An efficient real-time colonic polyp detection with YOLO algorithms trained by using negative samples and large datasets*. *Computers in Biology and Medicine*, 141, 105031. <https://doi.org/10.1016/j.compbimed.2021.105031>
- [15] *CUDA*: Documentación oficial de *CUDA* de NVIDIA. Disponible en: <https://docs.nvidia.com/cuda/>(<https://docs.nvidia.com/cuda/>)
- [16] *PyTorch*: Documentación oficial de *PyTorch*. Disponible en: <https://pytorch.org/docs/stable/index.html>(<https://pytorch.org/docs/stable/index.html>)



- [17] *Jupyter Notebook*: Documentación oficial de *Jupyter Notebook*. Disponible en: [\[https://jupyter.org/documentation\]](https://jupyter.org/documentation)(<https://jupyter.org/documentation>)
- [18] *EndoTect*: Base de datos especializada en imágenes médicas endoscópicas. Para obtener más información, consulta: [\[https://endo-tect.grand-challenge.org/\]](https://endo-tect.grand-challenge.org/)(<https://endo-tect.grand-challenge.org/>)
- [19] *Kvasir*: Base de datos de imágenes endoscópicas para el desarrollo de algoritmos de detección y diagnóstico. Puedes acceder a ella en: [\[https://datasets.simula.no/kvasir/\]](https://datasets.simula.no/kvasir/)(<https://datasets.simula.no/kvasir/>)
- [20] *HyperKvasir*: Base de datos de imágenes endoscópicas gastrointestinalmente etiquetadas. Para más detalles, visita: [\[https://datasets.simula.no/hyper-kvasir/\]](https://datasets.simula.no/hyper-kvasir/)(<https://datasets.simula.no/hyper-kvasir/>)
- [21] *CVC*: Centro de Visión por Computador, una institución que proporciona conjuntos de datos para investigación en visión por computador. Más información en: [\[https://www.cvc.uab.es/\]](https://www.cvc.uab.es/)(<https://www.cvc.uab.es/>)
- [22] "Kaggle." *Kaggle*, [www.kaggle.com/](http://www.kaggle.com/).
- [23] "Google Dataset Search." *Google Dataset Search*, [datasetsearch.research.google.com/](https://datasetsearch.research.google.com/).
- [24] *Ultralytics*. "K-Fold Cross Validation." *Ultralytics Documentation*, <https://docs.ultralytics.com/guides/kfold-cross-validation/>.

## Anexo 1: Objetivos de Desarrollo Sostenible

Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS).

<b>Objetivos de Desarrollo Sostenibles</b>	<b>Alto</b>	<b>Medio</b>	<b>Bajo</b>	<b>No Procede</b>
ODS 1. <b>Fin de la pobreza.</b>				<b>X</b>
ODS 2. <b>Hambre cero.</b>				<b>X</b>
ODS 3. <b>Salud y bienestar.</b>	<b>X</b>			
ODS 4. <b>Educación de calidad.</b>				<b>X</b>
ODS 5. <b>Igualdad de género.</b>				<b>X</b>
ODS 6. <b>Agua limpia y saneamiento.</b>				<b>X</b>
ODS 7. <b>Energía asequible y no contaminante.</b>				<b>X</b>
ODS 8. <b>Trabajo decente y crecimiento económico.</b>			<b>X</b>	
ODS 9. <b>Industria, innovación e infraestructuras.</b>		<b>X</b>		
ODS 10. <b>Reducción de las desigualdades.</b>				<b>X</b>
ODS 11. <b>Ciudades y comunidades sostenibles.</b>				<b>X</b>
ODS 12. <b>Producción y consumo responsables.</b>				<b>X</b>
ODS 13. <b>Acción por el clima.</b>				<b>X</b>
ODS 14. <b>Vida submarina.</b>				<b>X</b>
ODS 15. <b>Vida de ecosistemas terrestres.</b>				<b>X</b>
ODS 16. <b>Paz, justicia e instituciones sólidas.</b>				<b>X</b>
ODS 17. <b>Alianzas para lograr objetivos.</b>	<b>X</b>			

Tabla 11: Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS).

### Reflexión sobre la relación del TFG/TFM con los ODS y con el/los ODS más relacionados

Reflexión sobre la Relación de los Objetivos de Desarrollo Sostenible (ODS) con la Investigación sobre Pólipos

Al desarrollar este trabajo de fin de grado sobre pólipos, se puede apreciar cómo esta investigación se entrelaza con diversos Objetivos de Desarrollo Sostenible (ODS). A medida

que se profundiza en el estudio de los pólipos y su impacto en la salud, emerge una clara conexión con varios ODS, reflejando cómo la ciencia médica puede contribuir significativamente al desarrollo sostenible en múltiples niveles.

En primer lugar, la conexión con el *ODS 3: Salud y bienestar* es inmediata. Los pólipos, particularmente aquellos en el tracto gastrointestinal, representan un riesgo significativo para la salud, dado su potencial para evolucionar hacia condiciones más graves, como el cáncer colorrectal. La investigación enfocada en mejorar la detección temprana y el tratamiento eficaz de los pólipos es crucial para reducir la morbilidad y mortalidad asociadas a estas patologías. Al avanzar en la comprensión de los factores que conducen a la formación de pólipos y al desarrollar métodos menos invasivos para su manejo, se contribuye directamente al objetivo de promover una vida sana y el bienestar para todos.

Además, existe una notable conexión con el *ODS 8: Trabajo decente y crecimiento económico*. Las enfermedades relacionadas con los pólipos pueden afectar significativamente la capacidad laboral de las personas, llevando a ausencias prolongadas y reduciendo la productividad. Mejorar la prevención y el tratamiento de estas condiciones no solo beneficia a los individuos, sino que también fortalece la economía al mantener a la fuerza laboral en buen estado de salud. La industria de la salud, al integrar innovaciones derivadas de esta investigación, puede generar empleo y estimular el crecimiento económico, demostrando cómo una población más saludable puede impulsar el desarrollo económico.

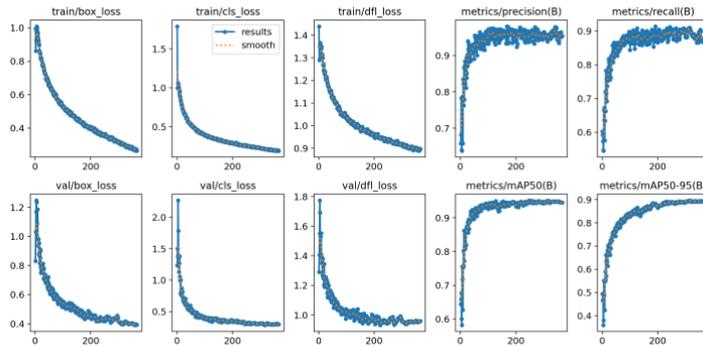
El *ODS 9: Industria, innovación e infraestructuras* también se ve reflejado en el contexto de esta investigación. El estudio de los pólipos promueve la innovación tecnológica en el campo de la medicina. La exploración de nuevas técnicas de diagnóstico, como las imágenes médicas avanzadas y los métodos de tratamiento menos invasivos, impulsa el avance de la tecnología médica. Estas innovaciones no solo mejoran la eficacia de los tratamientos, sino que también pueden llevar a la modernización de las infraestructuras de salud, asegurando que más personas tengan acceso a cuidados de alta calidad. La intersección entre la investigación médica y la tecnología subraya cómo los avances en un área pueden catalizar mejoras significativas en otras.

Finalmente, la relación con el *ODS 17: Alianzas para lograr los objetivos* destaca la importancia de la cooperación y el intercambio de conocimientos. La investigación en salud prospera gracias a las colaboraciones entre hospitales, instituciones académicas, y profesionales de la salud. Este trabajo se beneficia enormemente de las alianzas que facilitan el acceso a recursos y conocimientos compartidos. La colaboración permite la implementación más rápida y eficaz de nuevas técnicas y tratamientos, asegurando que los descubrimientos beneficien a una población más amplia. Las alianzas son esenciales para enfrentar los desafíos de salud de manera integral y equitativa.

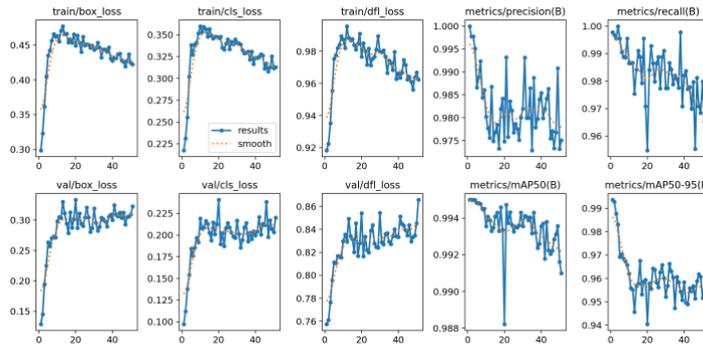
En conjunto, la investigación sobre pólipos no solo impacta directamente en la salud y el bienestar de las personas, sino que también tiene implicaciones más amplias en la economía, la innovación tecnológica y la cooperación internacional. Esta reflexión destaca cómo un enfoque especializado puede tener un alcance amplio, contribuyendo al logro de varios Objetivos de Desarrollo Sostenible y promoviendo un futuro más saludable y sostenible para todos. La integración de estos ODS en el desarrollo de la investigación demuestra el poder de la ciencia médica para catalizar cambios positivos en múltiples dimensiones del desarrollo humano.

## Anexo 2: Resultados Cross Validation SPLIT 1-7

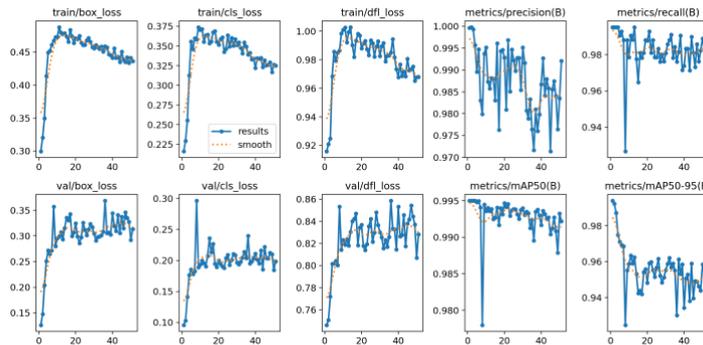
---



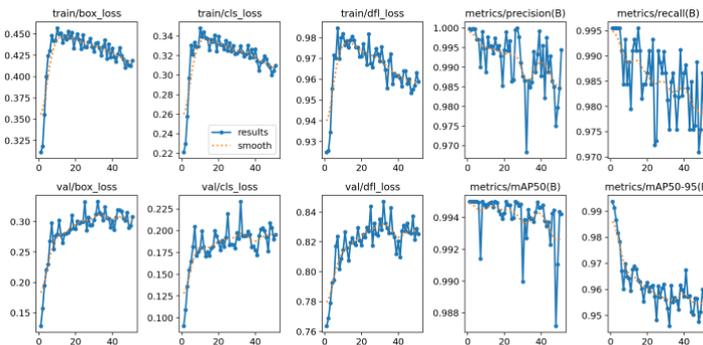
SPLIT 1



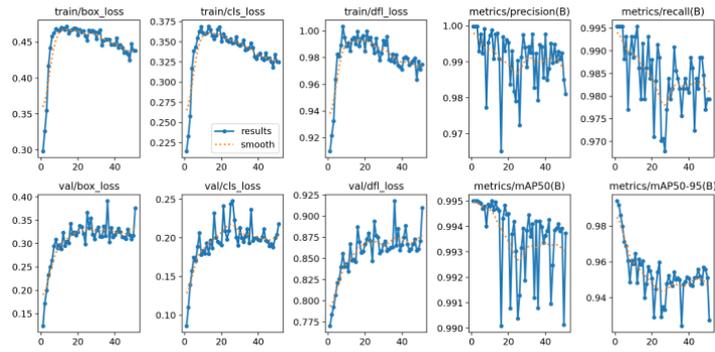
SPLIT 2



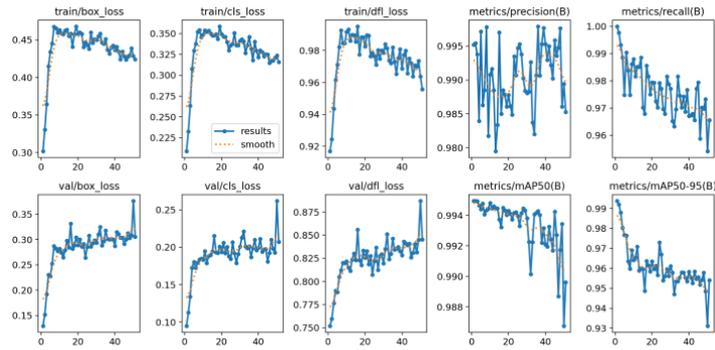
SPLIT 3



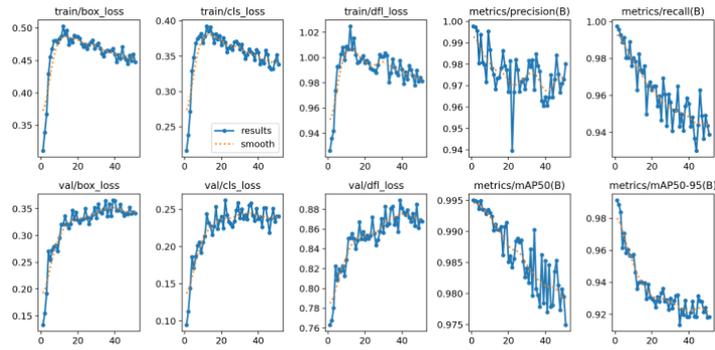
SPLIT 4



SPLIT 5



SPLIT 6



SPLIT 7