

**DETECCIÓN DE PÓLIPOS EN SECUENCIAS LARGAS DE COLONOSCOPIA  
SIGUIENDO ESQUEMAS NO SUPERVISADOS.**

**DANIEL EDUARDO ORTIZ CELIS**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FÍSICOMECÁNICAS  
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA  
PROGRAMA DE PREGRADO EN INGENIERÍA DE SISTEMAS E INFORMÁTICA  
BUCARAMANGA  
2024**

**DETECCIÓN DE PÓLIPOS EN SECUENCIAS LARGAS DE COLONOSCOPIA  
SIGUIENDO ESQUEMAS NO SUPERVISADOS.**

**DANIEL EDUARDO ORTIZ CELIS**

**Trabajo de grado presentado en cumplimiento de los requisitos para el grado de:  
Ingeniero de Sistemas e Informática**

**Director:**

**Fabio Martínez Carrillo, Ph.D.**

**Doctor en Ingeniería de Sistemas y Computación**

**Codirector:**

**Edgar Yesid Rangel Pieschacón M.Sc.**

**Magister en Ingeniería de Sistemas y Computación**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FÍSICOMECÁNICAS  
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA  
PROGRAMA DE PREGRADO EN INGENIERÍA DE SISTEMAS E INFORMÁTICA  
BUCARAMANGA  
2024**

## **AGRADECIMIENTOS**

Agradezco a mis padres por todo el sacrificio, apoyo, educación, orientación, disciplina, dedicación, ética, valores y en especial por todo el amor y afecto que me han brindado a lo largo de mi vida que ha me ha permitido estar donde estoy. La confianza que depositaron en mí me dio la fuerza para perseverar y alcanzar este logro.

También quiero agradecer al profesor Fabio Martínez quien siempre estuvo dispuesto a escuchar mis preguntas, brindar orientación y alentar mi curiosidad intelectual. Su compromiso y su dedicación son ejemplos por seguir que nunca olvidaré.

A mi codirector Edgar Rangel que siempre estuvo ahí brindado su guía, conocimiento, orientación, compromiso y apoyo fueron esenciales para llevar con éxito este trabajo de grado.

Al grupo de Investigación BIVL2ab del cual tuve el privilegio de ser parte, quisiera agradecerles por toda la ayuda que me brindaron, las discusiones e ideas me permitieron ampliar la perspectiva y mejorar mi trabajo de manera significativa. Además, debo agradecer la disposición de todos por responder mis preguntas y compartir sus experiencia.

De igual forma, este trabajo de grado fue apoyado por el proyecto: Cuantificación de lesiones de próstata comparando secuencias multi-paramétricas y bi-paramétricas de MRI usando herramientas de inteligencia artificial, con código 3946 que tuvo el soporte de la Vicerrectoría de Investigación y Extensión (VIE) de la Universidad Industrial de Santander y el grupo de Investigación BIVL2ab.

## **DEDICATORIA**

*Quisiera dedicarle este trabajo a mis padres que fueron mi soporte durante toda esta aventura y durante toda mi vida, que con sus enseñanzas y dedicación me motivaron a seguir incluso en los momentos mas difíciles. A mi hermano que ya no está físicamente conmigo pero que me acompaña en el corazón.*

***Daniel Eduardo Ortiz Celis.***

## CONTENIDO

	<b>page</b>
<b>INTRODUCCIÓN</b> . . . . .	<b>11</b>
<b>1. Fundamentos y trabajos previos</b> . . . . .	<b>14</b>
1.1. El cáncer colorrectal y la detección de pólipos . . . . .	14
1.2. Estrategias de aprendizaje y ajuste de representaciones profundas . . . . .	15
1.2.1 Aprendizaje supervisado. . . . .	16
1.2.2 Aprendizaje débilmente supervisado. . . . .	18
1.2.3 Aprendizaje semisupervisado. . . . .	18
1.2.4 Aprendizaje auto-supervisado. . . . .	19
1.2.5 Detección de anomalías. . . . .	22
1.3. Aproximaciones computacionales para la localización y caracterización de pólipos	24
<b>2. Problema de Investigación</b> . . . . .	<b>28</b>
<b>3. Objetivos</b> . . . . .	<b>29</b>
<b>4. Metodología Propuesta</b> . . . . .	<b>30</b>
4.1. Arquitectura convolucional con múltiples mecanismos de atención. . . . .	31
4.2. Entrenamiento semisupervisado para la detección de cáncer colorrectal . . . . .	33
4.3. Entrenamiento autosupervisado y detección de pólipos como anomalías . . . . .	34
<b>5. Diseño Experimental</b> . . . . .	<b>37</b>
5.1. Conjunto de datos de colonoscopias . . . . .	37
5.1.1 Conjunto de datos publico . . . . .	37
5.1.2 Conjunto de secuencias largas de colonoscopia: COLON . . . . .	38

5.2. Hiperparámetros de la arquitectura propuesta . . . . .	40
5.3. Validación . . . . .	41
5.3.1 Enfoque semi-supervisado . . . . .	41
5.3.2 Enfoque de detección de anomalías . . . . .	41
<b>6. Evaluación y Resultados . . . . .</b>	<b>44</b>
6.1. Detección de pólipos siguiendo un esquema semi-supervisado . . . . .	44
6.2. Detección de pólipos siguiendo un esquema de detección de anomalías . . . . .	47
<b>7. Conclusiones y trabajo futuro . . . . .</b>	<b>50</b>
<b>BIBLIOGRAFÍA . . . . .</b>	<b>53</b>

## LISTA DE FIGURAS

	<b>page</b>
Figura 1. Procedimiento de Colonoscopia . . . . .	15
Figura 2. Taxonomía de los esquemas de aprendizaje profundo . . . . .	16
Figura 3. Aprendizaje supervisado . . . . .	17
Figura 4. Aprendizaje semisupervisado . . . . .	20
Figura 5. Aprendizaje auto-supervisado. . . . .	22
Figura 6. Aprendizaje de anomalías. . . . .	23
Figura 7. Metodologías propuestas para la detección de pólipos . . . . .	30
Figura 8. Arquitectura multiatención . . . . .	32
Figura 9. Metodología semisupervisada propuesta. . . . .	34
Figura 10 Metodología auto-supervisada para la detección de pólipos como anomalías .	36
Figura 11.Conjunto de secuencias largas de colonoscopia COLON . . . . .	39
Figura 12.Comparación entre los conjuntos de datos públicos y el conjunto de datos COLON . . . . .	40
Figura 13Entrenamiento y validación con los conjunto de datos para el enfoque semi- supervisado . . . . .	42
Figura 14Entrenamiento y validación con los conjunto de datos para el enfoque anómalo	43
Figura 15Métricas obtenidas con los conjuntos de datos públicos. . . . .	45
Figura 16Métricas obtenidas con el conjunto de datos COLON . . . . .	46
Figura 17Predicciones obtenidas en particiones de datos etiquetados. . . . .	48

## LISTA DE TABLAS

	<b>page</b>
Tabla 1. Hiperparámetros de entrenamiento de arquitectura . . . . .	41
Tabla 2. Comparación entre diferentes arquitecturas del estado del arte y las metodologías propuestas. . . . .	49

## RESUMEN

**TÍTULO:** Detección de pólipos en secuencias largas de colonoscopia siguiendo esquemas no supervisados \*

**AUTOR:** Daniel Eduardo Ortiz Celis \*\*

**PALABRAS CLAVE:** Pólipos, mecanismos de atención, aprendizaje no supervisado, detección de anomalías.

**DESCRIPCIÓN:** En Colombia, durante el año 2020, se diagnosticaron 11 mil nuevos casos de cáncer de colon, siendo el tercer tipo de cáncer con mayor incidencia a nivel nacional. Además, mundialmente se reportaron cerca de 2 millones de casos nuevos, siendo el tercer tipo de cáncer con mayor incidencia. Los pólipos observados durante procedimientos de colonoscopia son los principales precursores de la enfermedad y su detección temprana está asociada con mejores pronósticos. Sin embargo, la tarea de detección de estas masas es desafiante, incluso para expertos gastroenterólogos, reportando hasta un 26% de pólipos no clasificados. Las herramientas computacionales para la detección y caracterización de pólipos han permitido soportar estos procedimientos, pero siguen reportando limitaciones en ambientes reales debido a movimientos de cámara, similitud textural con pliegues intestinales, entre otros artefactos de la escena. Además, la escasez es un desafío adicional, apareciendo en menos del 10% de una secuencia de colonoscopia. En este trabajo se exploraron dos estrategias de entrenamiento débilmente supervisadas para la detección de cáncer colorrectal. Primero, se utiliza un modelo del estado del arte basado en una arquitectura multiatención, que se entrenó bajo un esquema de aprendizaje semisupervisado, realizando una tarea de segmentación binaria de pólipos. Durante este proceso, solo se incluyó un subconjunto de entrenamiento para simular escenarios con, complementando la información con pseudoetiquetas que complementan observaciones para ajustar el modelo. También en este trabajo se adaptó la arquitectura de aprendizaje profunda, para ajustarla en un esquema autosupervisado, siguiendo una tarea de reconstrucción. De esta arquitectura se implementó una estrategia para la detección (patrones no vistos durante el entrenamiento) utilizando únicamente datos del tracto intestinal. Estos dos enfoques fueron validados en dos conjuntos diferentes de datos, públicos y secuencias largas de colonoscopia, para evaluar el desempeño y generalización del modelo.

---

\* Trabajo de investigación

\*\* Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Fabio Martínez Carrillo PhD. Codirector: Edgar Yesid Rangel Pieschacón MSc.

## ABSTRACT

**TITLE:** Polyp detection in long colonoscopy sequences following unsupervised schemes \*

**AUTHOR:** Daniel Eduardo Ortiz Celis \*\*

**KEYWORDS:** Polyps, attention mechanisms, unsupervised learning, anomaly detection.

**DESCRIPTION:** In Colombia, during 2020, 11 thousand new cases of colon cancer were diagnosed, making it the third most common type of cancer nationwide. Additionally, globally, nearly 2 million new cases were reported, making it the third most common type of cancer in terms of incidence. Polyps observed during colonoscopy procedures are the main precursors of the disease, and their early detection is associated with better outcomes. However, the task of detecting these masses is challenging, even for expert gastroenterologists, with up to 26% of polyps remaining unclassified. Computational tools for polyp detection and characterization have supported these procedures, but they still report limitations in real-world environments due to camera movements, textural similarity with intestinal folds, among other scene artifacts. Moreover, sparsity is an additional challenge, as it appears in less than 10% of a colonoscopy sequence. In this work, two weakly supervised training strategies for colorectal cancer detection were explored. First, a state-of-the-art model based on a multi-attention architecture was used, trained under a semi-supervised learning scheme, performing a binary polyp segmentation task. During this process, only a subset of training data was included to simulate scenarios with limited annotations, complementing the information with pseudolabels to enhance model fitting. Additionally, in this work, a deep learning architecture was adapted and fine-tuned in a self-supervised scheme, following a reconstruction task. From this architecture, a strategy for detection (unseen patterns during training) was implemented using solely intestinal tract data. These two approaches were validated on two different datasets, public and long colonoscopy sequences, to assess model performance and generalization.

---

\* Research work

\*\* Faculty of Physics-Mechanics Engineering. School of Systems Engineering and Informatics. Advisor: Fabio Martínez Carrillo, PhD. Coadvisor: Edgar Yesid Rangel Pieschacón, MSc

## INTRODUCCIÓN

El cáncer de colon es el tercer cáncer con más incidencia a nivel mundial presentándose alrededor de 2 millones de casos durante el año 2020. Según la Organización Mundial de Salud (OMS), este tipo de cáncer es el segundo más mortal, reportando aproximadamente 935 mil decesos<sup>1</sup>. En Colombia se diagnosticaron alrededor de 11 mil nuevos casos, una cifra dramática que se espera siga en aumento cada año<sup>1</sup>. La tasa de supervivencia depende del estadio diagnosticado, en estadios tempranos hay 93% de probabilidad de recuperación, mientras que esta probabilidad disminuye drásticamente hasta un 8% para los estadios avanzados<sup>2</sup>.

El diagnóstico del cáncer de colon se realiza comúnmente por medio del examen de colonoscopia<sup>3</sup>. El objetivo de este examen es localizar y caracterizar pólipos, los cuales son masas anormales presentes en las paredes intestinales a lo largo del colon<sup>4</sup> y que son los precursores del cáncer colorrectal<sup>5</sup>. Estudios reportan que hasta un 26% de masas sospechosas

- 
- <sup>1</sup> ORGANIZACIÓN MUNDIAL DE LA SALUD (OMS). *Cancer Today*. <https://gco.iarc.fr/today/home>. International Agency for Research on Cancer, 2020.
  - <sup>2</sup> Luis Charúa GÜINDIC. “Cáncer de colon”. In: *Gastroenterología*. Ed. by Eduardo Pérez TORRES et al. New York, NY: McGraw-Hill Education, 2015.
  - <sup>3</sup> Bernard LEVIN et al. “Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology”. In: *Gastroenterology* 134.5 (2008), pp. 1570–1595.
  - <sup>4</sup> Oscar TAPIA et al. “Cáncer de colon y recto: Descripción morfológica y clínica de 322 casos”. In: *International Journal of Morphology* 28.2 (2010), pp. 393–398.
  - <sup>5</sup> Amy E NOFFSINGER. “Serrated polyps and colorectal cancer: new pathway to malignancy”. In: *Annual Review of Pathology: Mechanisms of Disease* 4 (2009), pp. 343–364.

no son detectadas durante el procedimiento <sup>6</sup>, debido a factores como: movimientos de cámara, problemas de iluminación, similitud de los pólipos con paredes del tracto intestinal<sup>6,7</sup>, entre muchos otros problemas. Por otra parte, la colonoscopia es un procedimiento prolongado<sup>7</sup>, que puede provocar fatiga en el especialista lo cual dificulta la detección de las masas sospechosas<sup>8</sup>, conduciendo a un diagnóstico tardío del cáncer colorrectal. Este hecho tiene un alto impacto en el tratamiento de la enfermedad, ya que estudios han evidenciado que la detección tardía de pólipos en estadios avanzados disminuye la tasa de supervivencia de cada individuo<sup>9</sup>.

Las herramientas computacionales han surgido como estrategias alternativas para soportar los diagnósticos de detección, localización y caracterización de pólipos. Estas herramientas afrontan desafíos relacionados al movimiento de la cámara, cambios abruptos de iluminación, y similitudes texturales entre los pólipos con las paredes intestinales<sup>6</sup>. Recientemente, estos trabajos han incluido mecanismos de atención, convoluciones dilatadas o métodos de aprendizaje débilmente supervisados para mejorar el modelamiento de los pólipos<sup>10</sup>. Estos métodos han demostrado importantes avances en la caracterización y discriminación del pólipo. Sin embargo, estas aproximaciones operan sobre observaciones independientes, las cuales han sido típicamente seleccionadas por un experto. Además, es-

---

<sup>6</sup> Quentin ANGERMANN; Aymeric HISTACE, and Olivier ROMAIN. “Active learning for real time detection of polyps in videocolonoscopy”. In: *Procedia Computer Science* 90 (2016), pp. 182–187.

<sup>7</sup> Peter B COTTON et al. “Colonoscopy: practice variation among 69 hospital-based endoscopists”. In: *Gastrointestinal endoscopy* 57.3 (2003), pp. 352–357.

<sup>8</sup> Chang Kyun LEE; Jae Myung CHA, and Wan Jung KIM. “Endoscopist fatigue may contribute to a decline in the effectiveness of screening colonoscopy”. In: *Journal of clinical gastroenterology* 49.6 (2015), e51–e56.

<sup>9</sup> Peter B COTTON et al. “Colonoscopy: practice variation among 69 hospital-based endoscopists”. In: *Gastrointestinal endoscopy* 57.3 (2003), pp. 352–357.

<sup>10</sup> Lina RUIZ and Fabio MARTINEZ. “Weakly supervised polyp segmentation from an attention receptive field mechanism”. In: *2022 44th Annual International Conference of the IEEE Engineering in Medicine And Biology Society (EMBC)*. IEEE. 2022, pp. 3745–3758.

tas herramientas excluyen la información completa de la colonoscopia, lo cual limita su uso en entornos clínicos reales. Por lo tanto, el análisis de colonoscopias debe tener en cuenta el carácter desbalanceado de los datos, donde la mayoría de los fotogramas corresponden al tracto intestinal sin pólipos.

En este trabajo se abordó el problema de detección de pólipos en secuencias largas de colonoscopia donde los pólipos son escasos y están representados por muy pocas muestras. Para ello, primero se abordó el problema desde un esquema semisupervisado, que ajustaba una arquitectura que segmenta pólipos. Durante fases de entrenamiento, la red se ajusta con un conjunto limitado de datos supervisados, los cuales son complementados con datos que contienen pseudoetiquetas. El subconjunto de datos con pseudoetiquetas fueron utilizados para refinar el entrenamiento, mejorando las características morfológicas aprendidas de cada clase. También como segunda contribución en este trabajo se adaptó una arquitectura de aprendizaje profundo para reconstruir solo fotogramas de colonoscopia sin pólipos, siguiendo un esquema de entrenamiento autosupervisado. Una vez entrenada la arquitectura, se utilizó una aproximación de análisis de anomalías para detectar los pólipos, como objetos que no son observados por la red. Estos enfoques fueron seleccionados para simular escenarios reales con pocos datos y el gran desbalanceo existen entre los fotogramas con y sin pólipo. Por último, los esquemas de aprendizaje fueron validados en conjuntos de datos públicos y en conjuntos de datos de secuencias largas de colonoscopia<sup>11</sup>, teniendo en cuenta su generalización y desempeño en la tarea de detección de pólipos.

---

<sup>11</sup> Lina RUIZ et al. "COLON: The largest COLonoscopia LONG sequence public database". In: *arXiv preprint arXiv:2403.00663*. 2024.

## 1. Fundamentos y trabajos previos

### 1.1. El cáncer colorrectal y la detección de pólipos

El cáncer colorrectal es el tercer cáncer con más ocurrencia a nivel global y el segundo con más mortalidad según estadísticas dadas por la Organización Mundial de Salud (OMS) <sup>1</sup>. Esta enfermedad presenta masas que son pólipos localizados en las paredes intestinales a lo largo del colon<sup>4</sup>, siendo su detección temprana, caracterización y remoción la principal alternativa de tratamiento para afrontar esta enfermedad<sup>12</sup>.

Los pólipos son los principales precursores del cáncer de colon<sup>5</sup>, los cuales representan acumulaciones anormales de células en las paredes intestinales, y dependiendo de sus características morfológicas es posible determinar el grado de malignidad que presenta este tipo de cáncer. Los pólipos tienen altas variaciones según el tamaño, lo cual está relacionado directamente con su grado de malignidad. Por ejemplo, pólipos pequeños (6 y 9 mm) son los más difíciles de localizar y tienen una probabilidad de malignidad del 0.4%, mientras que pólipos grandes (20mm) tienen una probabilidad de malignidad del 19%<sup>13</sup>. De esta manera, las características morfológicas de los pólipos (como su tamaño) están correlacionadas con su estadio. En la literatura se reporta una tasa de supervivencia del 93% en estadios tempranos y a medida que su tamaño va aumentando se va reduciendo la tasa de supervivencia hasta llegar a un 8%<sup>2</sup>. Por lo tanto, es importante detectar de forma temprana los pólipos en las colonoscopias<sup>9</sup>.

La colonoscopia es el examen estándar para la detección de pólipos<sup>3</sup>, el cual consiste en utilizar una cámara de vídeo que recorre todo el tracto intestinal en busca de masas anor-

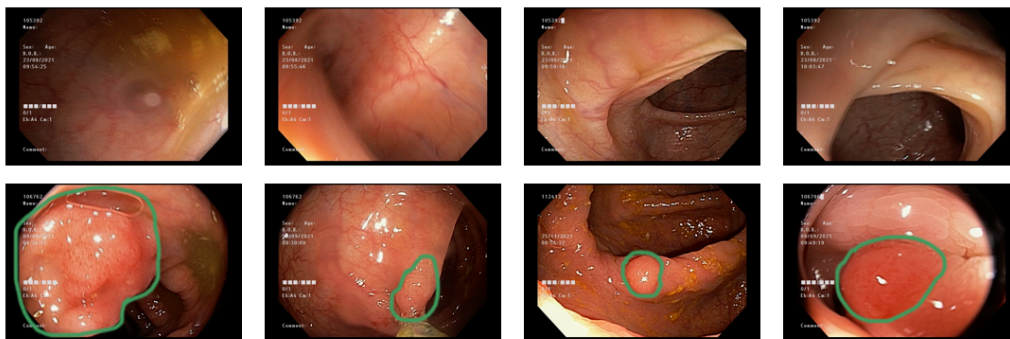
---

<sup>12</sup> Colorectal Canc Collaborative GRP et al. "Surgery for colorectal cancer in elderly patients: a systematic review". In: *The Lancet* 356.9234 (2000), pp. 968–974.

<sup>13</sup> Jasper LA VLEUGELS; Yark HAZEWINDEL, and Evelien DEKKER. "Morphological classifications of gastrointestinal lesions". In: *Best Practice & Research Clinical Gastroenterology* 31.4 (2017), pp. 359–367.

males. En la Figura 1 se observan diferentes hallazgos de pólipos, así como estructuras típicas del tracto intestinal. Durante la colonoscopia, la detección, inspección y análisis de pólipos es desafiante para el especialista, debido a su alto consumo de tiempo (aproximadamente de 30 minutos)<sup>7</sup>, además, el médico debe tener la suficiente experticia para buscar cuidadosamente el tracto intestinal buscando zonas donde pueden haber pólipos<sup>14</sup>. Por otra parte, durante la colonoscopia, los pólipos pueden no detectarse debido a su semejanza con las paredes del intestino, y se estima que la tasa de error del personal puede ser hasta del 26%<sup>15</sup>.

**Figura 1.** Imágenes que ilustran el procedimiento de colonoscopia, la primera fila son imágenes de tracto intestinal, y la segunda fila pólipos



## 1.2. Estrategias de aprendizaje y ajuste de representaciones profundas

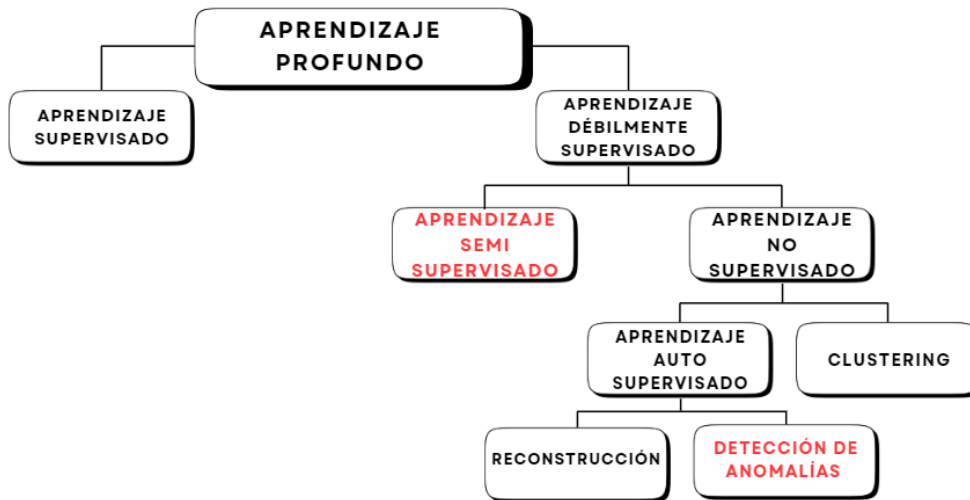
Las estrategias de aprendizaje para arquitecturas de aprendizaje profundas, permiten realizar el ajuste de un modelo respecto a una tarea determinada y considerando diferentes condiciones, como cantidad y calidad de datos etiquetados. Este ajuste es realizado por

<sup>14</sup> Douglas K REX et al. "Quality in the technical performance of colonoscopy and the continuous quality improvement process for colonoscopy: recommendations of the US Multi-Society Task Force on Colorectal Cancer". In: *The American journal of gastroenterology* 97.6 (2002), pp. 1296–1308.

<sup>15</sup> Jeroen C VAN RIJN et al. "Polyp miss rate determined by tandem colonoscopy: a systematic review". In: *Official journal of the American College of Gastroenterology/ ACG* 101.2 (2006), pp. 343–350.

medio de conjuntos de datos para entrenar y validar el modelo. En la Figura 2 se ilustra una taxonomía básica de los esquemas de aprendizaje de interés en este trabajo, los cuales serán discutidos en las siguientes subsecciones.

**Figura 2.** Taxonomía básica de los esquemas de aprendizaje profundo de interés en este trabajo.



**1.2.1. Aprendizaje supervisado.** El aprendizaje supervisado es la técnica más común de entrenamiento en reconocimiento de imágenes<sup>16</sup>, la cual utiliza conjuntos de datos totalmente etiquetados como se puede visualizar en la figura 3. El proceso de entrenamiento se puede ver en el algoritmo 1.

Formalmente, en el aprendizaje supervisado se tienen los conjuntos de datos  $\mathcal{D}_{tr} = \{X_i^{tr}, Y_i^{tr}\}_{i=1}^{n_{tr}}$  para entrenamiento y  $\mathcal{D}_{ts} = \{X_i^{ts}, Y_i^{ts}\}_{i=1}^{n_{ts}}$  para validación<sup>17,18</sup>, donde  $X$  son los datos de en-

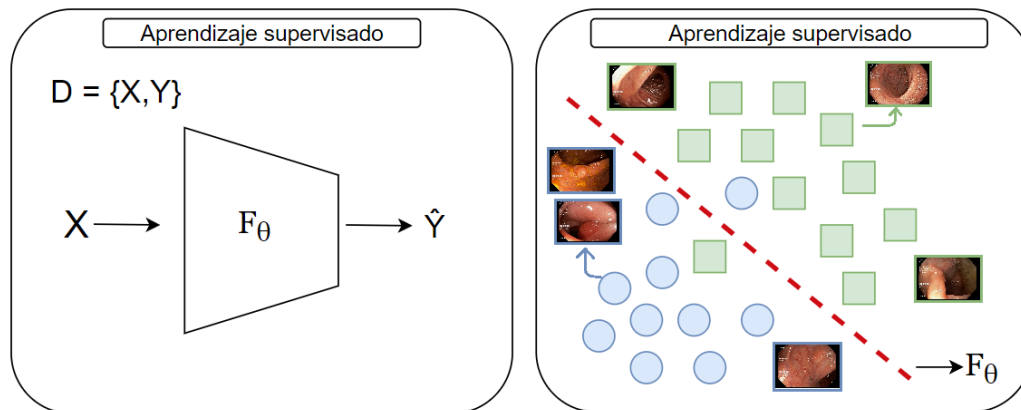
<sup>16</sup> Vladimir NASTESKI. “An overview of the supervised machine learning methods”. In: *Horizons. b* 4 (2017), pp. 51–62.

<sup>17</sup> Trevor HASTIE et al. “Overview of supervised learning”. In: *The elements of statistical learning: Data mining, inference, and prediction* (2009), pp. 9–41.

<sup>18</sup> Linus ERICSSON et al. “Self-supervised representation learning: Introduction, advances, and challenges”. In: *IEEE Signal Processing Magazine* 39.3 (2022), pp. 42–62.

trada y  $Y$  sus respectivas etiquetas. La forma en que un modelo genera una salida se puede definir mediante una función de aprendizaje  $F_\theta : X \rightarrow Y$ ,  $F_\theta(X) = Y$ <sup>17</sup>, donde una entrada  $X$  se asocia con su respectiva salida  $Y$ <sup>19</sup> y  $\theta$  son los parámetros de entrenamiento. Para lograr que el modelo aprenda a resolver la tarea, debe comparar la salida  $\hat{Y}$  con el valor real  $Y$ . Comúnmente, esta comparación se realiza por medio de una diferencia  $\|Y - F_\theta(X)\|_p$ <sup>17</sup>, donde  $p$  es el tipo de error que se desea minimizar. Su objetivo es permitirle al modelo aprender sobre los datos y realizar predicciones confiables. Por ejemplo, en el caso de  $p = 1$  se puede usar el error absoluto medio o norma L1 como función de pérdida para comparar las predicciones  $\hat{Y}$  con los valores reales  $Y$  quedando  $|Y - \hat{Y}|$ .

**Figura 3.** Aprendizaje supervisado: En el módulo de la izquierda se muestra un esquema básico supervisado y en el módulo de la derecha la separación entre las clases, la clase azul representa datos etiquetados como pólipos y la clase verde tracto intestinal.



El aprendizaje supervisado tiene limitaciones como el requerimiento total de las etiquetas, lo que puede conllevar a sesgos asociados a etiquetadores y además la creación de etiquetas es un proceso costoso<sup>20</sup>. Teniendo en cuenta lo anterior, surgen otros tipos de aprendizajes,

<sup>19</sup> Yann LECUN et al. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.

<sup>20</sup> Zhi-Hua ZHOU. "A brief introduction to weakly supervised learning". In: *National science review* 5.1 (2018), pp. 44–53.

---

**Algoritmo 1:** Entrenamiento Supervisado

---

$$D = \{(X, Y)\}$$

**for**  $i = 0$  **to**  $n$  epochs **do**

$$F_{\theta}^i(X) \rightarrow \hat{Y}$$

$$L_y = \|Y - F_{\theta}^i(X)\|_p$$

$F_{\theta}^i(X) \rightarrow F_{\theta}^{i-1}(X) - \frac{\partial L_y}{\partial \theta}$ , Ajuste de  $\theta$  mediante propagación hacia atrás.

**end for**

---

que se enfrentan a conjunto de datos que tienen pocas etiquetas o no tienen ninguna, con respecto a la tarea de interés.

**1.2.2. Aprendizaje débilmente supervisado.** El aprendizaje débilmente supervisado surge de la necesidad de abordar desafíos de reconocimiento cuando se tiene información parcial sobre las etiquetas <sup>21</sup>. De hecho es común que en entornos reales se cuente con pocos datos etiquetados, o con etiquetas débiles<sup>21</sup>. Una etiqueta débil corresponde a anotaciones parciales, mascarar, etiquetas a nivel semántico, ruidosas que además, se traducen como información parcial de los datos<sup>20 21</sup>. Hoy en día, muchos modelos utilizan el aprendizaje débilmente supervisado para mejorar y obtener mas etiquetas confiables sobre los datos.

**1.2.3. Aprendizaje semisupervisado.** El aprendizaje semisupervisado utiliza los datos etiquetados junto a los no etiquetados para entrenar un modelo<sup>22</sup>. Este esquema genera pseudo etiquetas de los datos no etiquetados<sup>22</sup>, teniendo en cuenta la distribución de los datos, considerando que los datos cercanos tienen la misma predicción<sup>22</sup>. Formalmente se tiene el conjunto de entrenamiento  $\mathcal{D}_{tr} = \{\mathcal{D}_l, \mathcal{D}_u\}$ , compuesto por el subconjunto de

---

<sup>21</sup> Dingwen ZHANG et al. "Weakly supervised object localization and detection: A survey". In: *IEEE transactions on pattern analysis and machine intelligence* 44.9 (2021), pp. 5866–5885.

<sup>22</sup> Jesper E VAN ENGELEN and Holger H HOOS. "A survey on semi-supervised learning". In: *Machine learning* 109.2 (2020), pp. 373–440.

datos etiquetados  $\mathcal{D}_l = \{X_i^{tr}, Y_i^{tr}\}_{i=1}^{n_{tr}}$  y el subconjunto  $\mathcal{D}_u = \{X_i^{tr}\}_{i=1}^{n_{tr}}$  que presenta datos con etiquetas desconocidas<sup>22 23</sup>. La forma típica en la que el modelo genera salidas es la siguiente: primero se entrena el modelo con los datos etiquetados  $F_\theta : X_l \rightarrow \hat{Y}_l$ , luego se propagan las etiquetas en los datos sin etiquetas  $F_\theta : X_u \rightarrow \hat{Y}_u$ , generando pseudo etiquetas<sup>22 24</sup>. Así mismo, hay que tener en cuenta que las  $\hat{Y}_u$  es muy probable que pertenezcan al conjunto de las etiquetas  $Y$  conocidas, debido a que las entradas  $X_u$  pueden ser cercanas a las entradas  $X_l$ <sup>24</sup> y/o sean de alta confiabilidad<sup>25</sup>. Esto quiere decir que las pseudo etiquetas generadas por el modelo están en el mismo dominio de las etiquetas conocidas y pueden contribuir como nuevas observaciones para el modelo. El modelo se va ir ajustando y aprendiendo mediante una comparación entre las etiquetas conocidas  $Y_l$  y las salidas generadas de los datos etiquetados  $\hat{Y}_l$ . Normalmente esta comparación se realiza de igual manera a lo supervisado  $\|Y_l - \hat{Y}_l\|_p$ , para hacer el ajuste de los parámetros  $\theta$  mediante propagación hacia atrás. Cuando se tengan las pseudo etiquetas  $\hat{Y}_u$ , se incluyen con su respectivo  $X_u$  en el conjunto de  $\mathcal{D}_l$ , con estos nuevos datos se puede utilizar el modelo para generar salidas y obtener buenos resultados, sin necesidad de trabajar con datos totalmente etiquetados desde el inicio, lo que puede ser costoso y laborioso<sup>26</sup>. Lo anterior se puede observar en la figura 4 y el algoritmo 2.

**1.2.4. Aprendizaje auto-supervisado.** El aprendizaje auto-supervisado permite ajustar modelos sin dependencia explícita de etiquetas semánticas, usando típicamente la informa-

---

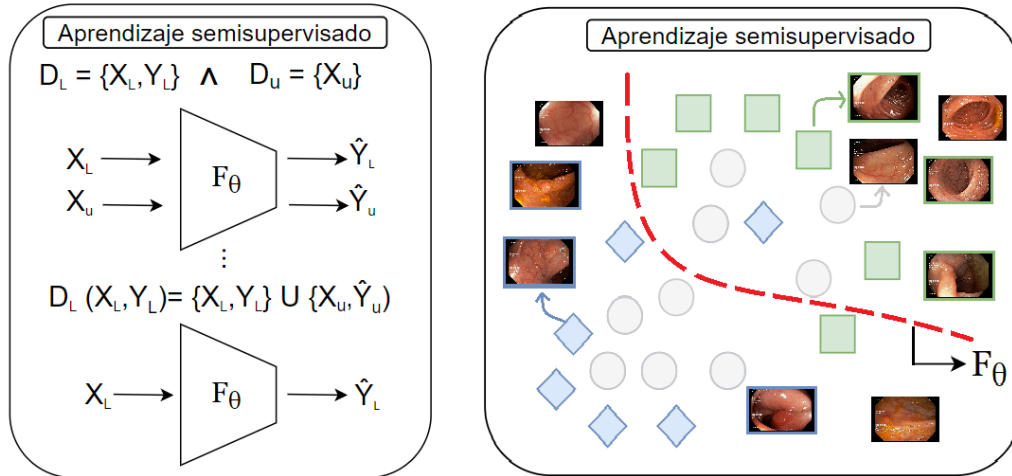
<sup>23</sup> Lars SCHMARJE et al. "A survey on semi-, self-and unsupervised learning for image classification". In: *IEEE Access* 9 (2021), pp. 82146–82168.

<sup>24</sup> Zhi-Hua ZHOU and Zhi-Hua ZHOU. "Semi-supervised learning". In: *Machine Learning* (2021), pp. 315–341.

<sup>25</sup> Kihyuk SOHN et al. "Fixmatch: Simplifying semi-supervised learning with consistency and confidence". In: *Advances in neural information processing systems* 33 (2020), pp. 596–608.

<sup>26</sup> Xiaojin Jerry ZHU. "Semi-supervised learning literature survey". In: (2005).

**Figura 4.** Aprendizaje semisupervisado: En el módulo de la izquierda se muestra un esquema básico semisupervisado y en el módulo de la derecha la separación entre las clases, la clase azul representa datos etiquetados como pólipos y la clase verde tracto intestinal, los datos con color gris representan aquellos datos que no están etiquetados.




---

### Algoritmo 2: Semisupervisado

---

Conjunto de datos  $D_l = \{(X_l, Y_l)\}$

Conjunto de datos  $D_u = \{X_u\}$

**for**  $i = 0$  **to**  $n$  epochs **do**

$F_\theta(X_l) \rightarrow \hat{Y}_l$

$L_y = ||Y_l - \hat{Y}_l||_p$

$F_\theta^i(X_l) \rightarrow F_\theta^{i-1}(X_l) - \frac{\partial L_y}{\partial \theta}$ , Ajuste de  $\theta$  mediante propagación hacia atrás.

$F_\theta(X_u) \rightarrow \hat{Y}_u$

$D_l\{(X_l, Y_l)\} = \{(X_l, Y_l) \cup (X_u, \hat{Y}_u)\}$

**end for**

$F_\theta(X_l)$ ,  $X_l$  es el nuevo conjunto etiquetado.

---

ción de las características de los datos<sup>18</sup>, es decir sin el uso de información de alto nivel<sup>27</sup>. El aprendizaje auto-supervisado trabaja mediante el uso de tareas de pretexto<sup>18</sup>, por ejemplo, la reconstrucción de una imagen con ayuda de un codificador  $F_\theta$  y un decodificador  $G_\theta$ . Cuando la entrada pasa por  $F_\theta$  se genera un vector embebido  $Z$  que es la representación a

---

<sup>27</sup> Longlong JING and Yingli TIAN. "Self-supervised visual feature learning with deep neural networks: A survey". In: *IEEE transactions on pattern analysis and machine intelligence* 43.11 (2020), pp. 4037–4058.

bajo nivel del dato de entrada, este vector pasa por  $G_\theta$  y reconstruye la entrada  $\hat{X}$ . Cuando se tiene entrenado el modelo de tarea de pretexto, se puede usar los vectores  $Z$  para agrupar los que estén cercanos y separar los que estén alejados. Cuando los datos están agrupados, se pueden asignar pseudo-etiquetas a los grupos<sup>1828</sup>. Formalmente se tiene  $\mathcal{D}_{tr} = \{X_i^{tr}\}_{i=1}^{n_{tr}}$ , también se tiene un conjunto de validación  $\mathcal{D}_{ts} = \{X_i^{ts}\}$ , de ambos conjuntos no se cuenta con etiquetas para el ajuste de la representación. La manera en que este enfoque trabaja es la siguiente:  $F_\theta : X \rightarrow Z$ ,  $F_\theta(X) = Z$ , el embebido pasa por la función del modelo para generar su la reconstrucción de la entrada:  $G_\theta : Z \rightarrow \hat{X}$ ,  $G_\theta(Z) = \hat{X}$ <sup>18</sup>. La tarea de pretexto se va ir ajustando y aprendiendo mediante una comparación entre la entrada  $X_u$  y la reconstrucción  $\hat{X}$ . Normalmente esta comparación se realiza de la siguiente manera  $\|X - \hat{X}\|_p$ , donde  $p$  representa el tipo de error que se desea minimizar y, así mismo, se ajustan los parámetros  $\theta$  mediante propagación hacia atrás. Las pseudo-etiquetas se generan mediante el agrupamiento de los embebidos de la siguiente forma: Se selecciona un valor  $c$  del conjunto de centros de clusters  $C$ , luego se realiza la siguiente operación  $\bar{Y} = \|Z - C\|$ , donde  $\bar{Y}$  son las pseudo-etiquetas<sup>18 28</sup>. Lo mencionado se ve en la figura 5 y el algoritmo 3.

---

### Algoritmo 3: Auto-supervisado

---

```

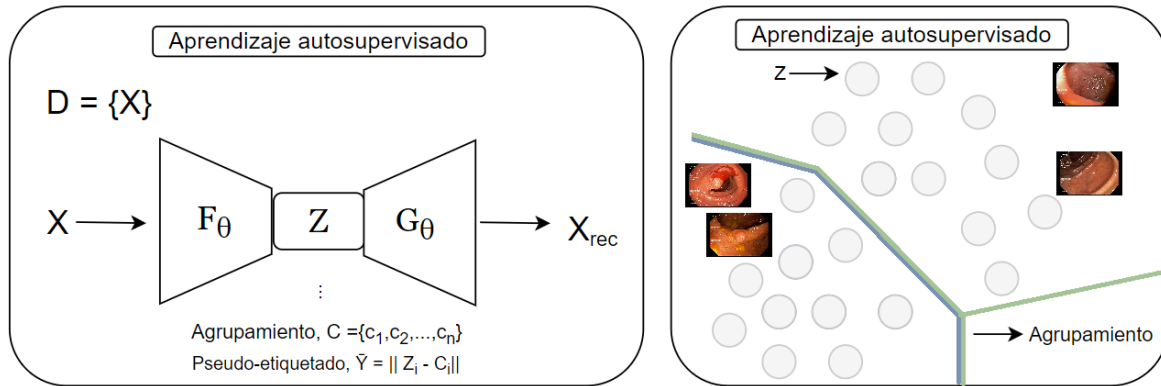
Conjunto de datos  $D = \{X\}$ 
for  $i = 0$  to  $n$  epochs do
     $F_\theta(X) \rightarrow Z$ 
     $G_\theta(Z) \rightarrow \hat{X}$ 
     $L_x = \|X - \hat{X}\|_p$ 
     $F_\theta^i(X) \rightarrow F_\theta^{i-1}(X) - \frac{\partial L_x}{\partial \theta}$ , ajuste de  $\theta$  mediante propagación hacia atrás.
     $G_\theta^i(Z) \rightarrow G_\theta^{i-1}(Z) - \frac{\partial L_x}{\partial \theta}$ , ajuste de  $\theta$  mediante propagación hacia atrás.
     $\bar{Y} = \|Z_i - C_i\|$ , asignación de pseudo-etiqueta mediante agrupamiento
end for

```

---

<sup>28</sup> Mathilde CARON et al. "Deep clustering for unsupervised learning of visual features". In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 132–149.

**Figura 5.** Aprendizaje auto-supervisado: En el módulo de la izquierda se muestra un esquema básico auto-supervisado y en el módulo de la derecha la agrupación de los datos, los que están en agrupados en el área azul representan pólipos y los que están agrupados en el área verde representan tracto intestinal.

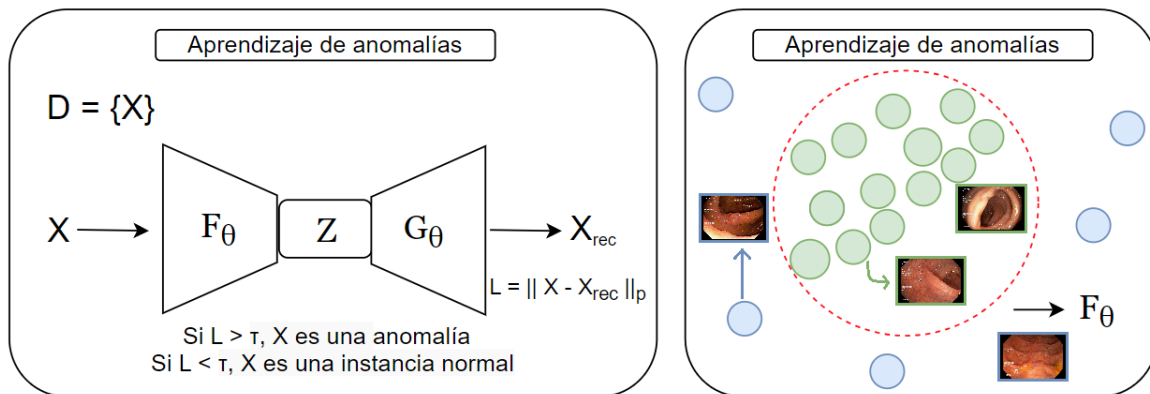


**1.2.5. Detección de anomalías.** Dentro del esquema de aprendizaje auto-supervisado se encuentra la detección de anomalías como un esquema de entrenamiento que busca identificar muestras que se desvían del comportamiento normal<sup>29</sup>. Los modelos que trabajan con este enfoque se centran en aprender muy bien los datos de una clase en específico (*One-class learning*). La detección de anomalías es una técnica útil para nuestro problema de detección de pólipos, ya que podemos aprovechar la gran cantidad de imágenes de tracto intestinal para aprender a identificar los datos de pólipos como anomalías. Una estrategia común para la detección de anomalías es establecer una región que incluya todos los datos conocidos, considerando cualquier punto fuera de esta región como una anomalía. Sin embargo, en la práctica esta tarea no es tan simple, ya que definir con precisión los límites de esta región puede ser desafiante. Puede ocurrir que los casos que están cerca de esos límites se clasifiquen como anómalos cuando en realidad son datos normales o viceversa<sup>29</sup>. Una alternativa es definir un umbral  $\tau$ , para crear una regla que permita la separación de los datos que pertenecen a la clase entrenada<sup>29</sup>. Formalmente en este esquema se tiene

<sup>29</sup> Varun CHANDOLA; Arindam BANERJEE, and Vipin KUMAR. "Anomaly detection: A survey". In: *ACM computing surveys (CSUR)* 41.3 (2009), pp. 1–58.

los conjuntos de datos  $\mathcal{D}_{tr} = \{X_i^{tr}\}_{i=1}^{n_{tr}}$  para entrenamiento y  $\mathcal{D}_{ts} = \{X_i^{ts}\}_{i=1}^{n_{ts}}$  para validación. La forma en la que el modelo va a generar salidas, siguiendo un esquema de detección de anomalías, se realiza con una tarea de pretexto como puede ser la reconstrucción de las datos de entrenamiento de una sola clase ( $F_\theta(X) = Z$  y  $G_\theta(Z) = \hat{Y}$ ). Para separar las anomalías de los datos normales, se realiza un calculo de error  $L = \|X - \hat{X}\|_p$  y proponiendo un umbral  $\tau$ , tal que  $L > \tau$  es una anomalía y en el caso contrario sera un dato normal<sup>30</sup>. Lo anterior se puede observar en la Figura 6 y el algoritmo 4.

**Figura 6.** Aprendizaje de anomalías. En el módulo de la izquierda se muestra un esquema básico de aprendizaje de anomalías y en el módulo de la derecha la separación de las clases, las que están de color verde son las instancias normales que en este caso representan al tracto intestinal, y las anomalías que están en azul son los pólipos.



<sup>30</sup> Jinwon AN and Sungzoon CHO. "Variational autoencoder based anomaly detection using reconstruction probability". In: *Special lecture on IE 2.1* (2015), pp. 1–18.

---

**Algoritmo 4: Entrenamiento de Anomalías**

---

Conjunto de datos  $D = \{X\}$

**for**  $i = 0$  **to**  $n$  epochs **do**

$F_{\theta}^i(X) \rightarrow Z$

$G_{\theta}^i(Z) \rightarrow X_{rec}$

$L = \|X - X_{rec}\|_p$

**if**  $L > \tau$  **then**

x es una anomalía

**else**

x es una instancia normal

**end if**

$F_{\theta}^i(X) \rightarrow F_{\theta}^{i-1}(X) - \frac{\partial L}{\partial \theta}$ , Ajuste de  $\theta$  mediante propagación hacia atrás.

$G_{\theta}^i(Z) \rightarrow G_{\theta}^{i-1}(Z) - \frac{\partial L}{\partial \theta}$ , Ajuste de  $\theta$  mediante propagación hacia atrás.

**end for**

---

### 1.3. Aproximaciones computacionales para la localización y caracterización de pólipos

Las aproximaciones computacionales para la localización y caracterización de pólipos han emergido como una alternativa para reducir la subjetividad del diagnóstico del cáncer colorrectal. Las primeras aproximaciones se centraron en identificar y caracterizar los bordes de los pólipos para aproximar su forma mediante un descriptores geométricos<sup>31</sup>. También se propusieron estrategias basadas en mapas de energía, que delimitan los pólipos medi-

---

<sup>31</sup> Sae HWANG et al. "Polyp detection in colonoscopy video using elliptical shape feature". In: *2007 IEEE International Conference on Image Processing*. Vol. 2. IEEE. 2007, pp. II-465.

ante los valles generados por la incidencia de la luz<sup>32</sup>. Sin embargo, estas estrategias se basan en el diseño de características ajustadas manualmente, que resultan en variaciones sensibles de escenarios, que puede dificultar su aplicación en escenarios clínicos reales. Las representaciones profundas han demostrado su capacidad para caracterizar y localizar pólipos a partir de la forma y características texturales, lo que permite lidiar con las variaciones de escenas capturadas durante colonoscopias. Las aproximaciones basadas en arquitecturas codificador-decodificador permiten obtener la caracterización de los pólipos<sup>3334</sup>. También se han implementado aproximaciones basadas en arquitecturas U-net, que añaden convoluciones dilatadas. Estas convoluciones permiten capturar características de diferentes escalas<sup>35</sup>. Otra arquitectura que ha demostrado buenos resultados es la ResUnet++. Esta arquitectura combina bloques convolucionales que permiten la captura de características con un campo aleatorio condicional (CRF) que captura información contextual a partir de modelos estadísticos<sup>36</sup>. Sin embargo, todas estas aproximaciones tienen una limitación común: su capacidad para caracterizar pólipos de manera local. Esto se debe a que las convoluciones tienen un campo receptivo limitado, lo que dificulta la captura de caracterís-

---

<sup>32</sup> Jorge BERNAL et al. "Polyp segmentation method in colonoscopy videos by means of MSA-DOVA energy maps calculation". In: *Clinical Image-Based Procedures. Translational Research in Medical Imaging: Third International Workshop, CLIP 2014, Held in Conjunction with MICCAI 2014, Boston, MA, USA, September 14, 2014, Revised Selected Papers 3*. Springer. 2014, pp. 41–49.

<sup>33</sup> Patrick BRANDAO et al. "Fully convolutional neural networks for polyp segmentation in colonoscopy". In: *Medical Imaging 2017: Computer-Aided Diagnosis*. Vol. 10134. Spie. 2017, pp. 101–107.

<sup>34</sup> Qiaoliang LI et al. "Colorectal polyp segmentation using a fully convolutional neural network". In: *2017 10th international congress on image and signal processing, biomedical engineering and informatics (CISP-BMEI)*. IEEE. 2017, pp. 1–5.

<sup>35</sup> Xinzi SUN et al. "Colorectal polyp segmentation by u-net with dilation convolution". In: *2019 18th IEEE international conference on machine learning and applications (ICMLA)*. IEEE. 2019, pp. 851–858.

<sup>36</sup> Debesh JHA et al. "A comprehensive study on colorectal polyp segmentation with ResUNet++, conditional random field and test-time augmentation". In: *IEEE journal of biomedical and health informatics* 25.6 (2021), pp. 2029–2040.

ticas globales. Por otro lado, la arquitectura HarDNet-MSEG<sup>37</sup> basada en representaciones convolucionales, que consiste en una arquitectura encoder-decoder con bloques de campo receptivo y que trabaja con múltiples ramas que contienen convoluciones normales y dilatadas para capturar mejor la información espacial obtuvieron buenos resultados en la segmentación de pólipos<sup>37</sup>.

Los mecanismos de atención se han propuesto para segmentar pólipos utilizando relaciones no lineales entre características. Estos mecanismos han mostrado la capacidad de discriminar entre pólipos y tracto intestinal. Por ejemplo, la arquitectura PraNet<sup>38</sup> utiliza un decodificador parcial para generar mapas semánticos globales y mecanismos de atención en reversa para mejorar la segmentación. Los mecanismos de atención en reversa se basan en la idea de que un médico primero localiza el pólipo y luego lo examina cuidadosamente para etiquetarlo. Estos mecanismos extraen regiones y establecen relaciones entre los límites del pólipo y el área que este tiene<sup>38</sup>. Sin embargo, esta arquitectura es sensible y puede descartar regiones relevantes. En un trabajo posterior, se propuso la arquitectura Polyp-PVT (Pyramid Vision Transformers)<sup>39</sup>, que mejora la robustez de la extracción de características y que combina información de alto y bajo nivel en un módulo de atención global para obtener las relaciones más importantes entre características. Así mismo, Ruiz *et. al.*<sup>10</sup> propuso una estrategia que combina los mecanismos de atención, bloques de campo receptivo y un entrenamiento débilmente supervisado, segmentando y clasificando pólipos. El modelo captura características del pólipo que pasan al mecanismo de atención donde se aprenden

---

<sup>37</sup> Chien-Hsiang HUANG; Hung-Yu WU, and Youn-Long LIN. "Hardnet-mseg: A simple encoder-decoder polyp segmentation neural network that achieves over 0.9 mean dice and 86 fps". In: *arXiv preprint arXiv:2101.07172* (2021).

<sup>38</sup> Deng-Ping FAN et al. "Pranet: Parallel reverse attention network for polyp segmentation". In: *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VI*. vol. 23. Springer International Publishing. 2020, pp. 263–273.

<sup>39</sup> Bo DONG et al. "Polyp-pvt: Polyp segmentation with pyramid vision transformers". In: *arXiv preprint arXiv:2108.06932* (2021).

relaciones entre las dos clases, para así caracterizar el pólipo<sup>10</sup>. Este enfoque es realizado bajo un esquema débilmente supervisado, que entrena con etiquetas de bajo coste que son las mascararas. De igual manera, en la literatura se propuso ColonSegNet<sup>40</sup>, que es una arquitectura codificador-decodificador para la detección, localización y segmentación de pólipos mediante bloques residuales, pero cuyos resultados fueron obtenidos en conjuntos de datos controlados, lo que podría no funcionar en ambientes reales donde se presentan movimientos de cámara, problemas de iluminación.

Así mismo, hoy en día han aparecido aproximaciones basadas en enfoques semisupervisados para la segmentación que hacen uso de la arquitectura U-Net<sup>41</sup>, que aunque trata de combatir la escasez de etiquetas, esta limitado a trabajar con datos en ambientes controlados, es decir, con datos de pólipos etiquetados<sup>42</sup>. En la literatura, también aparecen aproximaciones que permiten el modelamiento de datos desbalanceados (observaciones con pólipos) con respecto al tracto intestinal. Por ejemplo, Tian *et al.*<sup>43</sup> propuso un modelo generativo que reconstruye la imagen, que se entrena con ambas clases, utilizando principalmente imágenes de tracto y solo unas pocas imágenes de pólipo<sup>43</sup> etiquetadas. Aunque trata de abordar el problema de los datos desbalanceados, utilizan conjuntos de datos privados con fotogramas controlados y necesita mas conjuntos de datos de pólipos para evaluar la generalización.

---

<sup>40</sup> Debesh JHA et al. "Real-time polyp detection, localization and segmentation in colonoscopy using deep learning". In: *Ieee Access* 9 (2021), pp. 40496–40510.

<sup>41</sup> Olaf RONNEBERGER; Philipp FISCHER, and Thomas BROX. "U-net: Convolutional networks for biomedical image segmentation". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* 18. Springer. 2015, pp. 234–241.

<sup>42</sup> Siwei CHEN; Gregor URBAN, and Pierre BALDI. "Weakly Supervised Polyp Segmentation in Colonoscopy Images Using Deep Neural Networks". In: *Journal of Imaging* 8.5 (2022), p. 121.

<sup>43</sup> Yu TIAN et al. "Few-shot anomaly detection for polyp frames from colonoscopy". In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VI* 23. Springer. 2020, pp. 274–284.

## **2. Problema de Investigación**

Las colonoscopias son procedimientos tediosos, que toman aproximadamente 30 minutos, lo que conlleva a que la detección de pólipos sea una tarea desafiante incurriendo en un alto índice de detecciones incorrectas. Es por ello, que en la literatura existen estrategias computacionales para soportar tareas de segmentación, localización y clasificación de pólipos. Sin embargo, estas estrategias están acotadas para trabajar en escenarios controlados con pocos datos donde algunos de datos solo presentan fotogramas con pólipos, otros pocos presentan poca cantidad de fondo pólipos y otras estrategias son susceptibles al ruido. De hecho, estas estrategias no consideran que la gran mayoría de fotogramas en una colonoscopia no contiene masas sospechosas. Esto hace que herramientas computacionales puedan tener limitaciones al ser transferidas a escenarios reales. Sumado a esto, la adaptación de estas herramientas, debido a su carácter estrictamente supervisado, requiere el etiquetado exhaustivo de nuevas colonoscopias lo cual puede incurrir en un alto costo de adaptación, así como en la dedicación de un tiempo de los especialistas.

### **3. Objetivos**

#### **Objetivo General**

- Implementar una representación profunda, siguiendo un entrenamiento no supervisado, para detectar pólipos observados en colonoscopias.

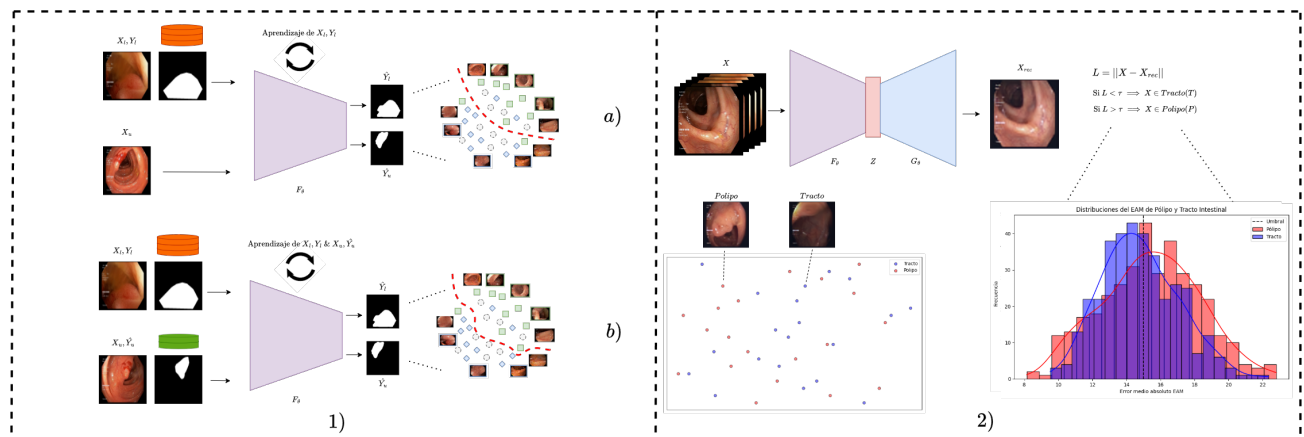
#### **Objetivo Específicos**

- Definir un conjunto de datos a partir de secuencias de colonoscopia reales que contenga un desbalance de pólipos.
- Implementar una arquitectura computacional de aprendizaje profundo, utilizando una estrategia de aprendizaje no supervisado para la detección de pólipos.
- Desarrollar un esquema de detección de anomalías para capturar patrones relevantes en la caracterización de pólipos sobre secuencias de colonoscopia.
- Validar la metodología propuesta con respecto a la capacidad de capturar patrones anormales asociados a los pólipos.

## 4. Metodología Propuesta

En este trabajo se exploraron dos estrategias de entrenamiento débilmente supervisado para la detección de pólipos en secuencias de colonoscopia explorando el ajuste de modelos computacionales sin requerimientos estrictamente supervisados. En un primer enfoque se propuso un esquema de aprendizaje semisupervisado para realizar una tarea de segmentación binaria de pólipos ajustando una arquitectura con múltiples mecanismos de atención. En un segundo enfoque se propuso un esquema no supervisado, ajustando una arquitectura de múltiples mecanismos de atención, por medio de una tarea de reconstrucción de imágenes. Desde esta perspectiva, la arquitectura fue entrenada únicamente con datos del tracto intestinal (fondo) y detectando los pólipos desde una perspectiva de anomalías. En la Figura 7 se observa de manera general los dos enfoques propuestos.

**Figura 7.** Metodologías propuestas para la detección de pólipos en secuencias de colonoscopia. A la izquierda se observa la metodología de entrenamiento semisupervisado. En la sección 1-a la arquitectura realizó el aprendizaje con los datos limitados etiquetados  $X_l, Y_l$ . Luego el aprendizaje se complementa con pseudoetiquetas. A la derecha se observa la estrategia no supervisada siguiendo una tarea de reconstrucción de imágenes. La detección se realizó mediante una estrategia de detección de anomalías, entrenando con clases de solo fondo



#### 4.1. Arquitectura convolucional con múltiples mecanismos de atención.

El objetivo principal de este trabajo fue explorar y proponer alternativas de aprendizaje y ajuste de arquitecturas de aprendizaje profundo, permitiendo así ser adaptadas al problema de detección de pólipos en escenarios reales, *i.e.*, datos poco etiquetados y con redundancia de imágenes de fondo (tracto intestinal). Es por ello, que en este trabajo decidimos adoptar una arquitectura del estado del arte para la segmentación y caracterización de pólipos<sup>10</sup>. Sin embargo, la metodología propuesta de aprendizaje puede ser flexible para incluir otras representaciones profundas.

La arquitectura adoptada ha demostrado ser capaz de segmentar pólipos, siendo robusta a cambios de apariencia, forma y artefactos presentes en fotogramas de colonoscopia. La Figura 8 ilustra la arquitectura base implementada en este trabajo. Esta arquitectura se fundamenta principalmente en mecanismos de atención y bloques convolucionales de campo receptivo (RFB, por sus siglas en inglés) para capturar características del tracto intestinal y particularmente para caracterizar pólipos, conjugando información local (desde bloques convolucionales), intermedia (desde bloques receptivos) y no-local (desde mecanismos de atención).

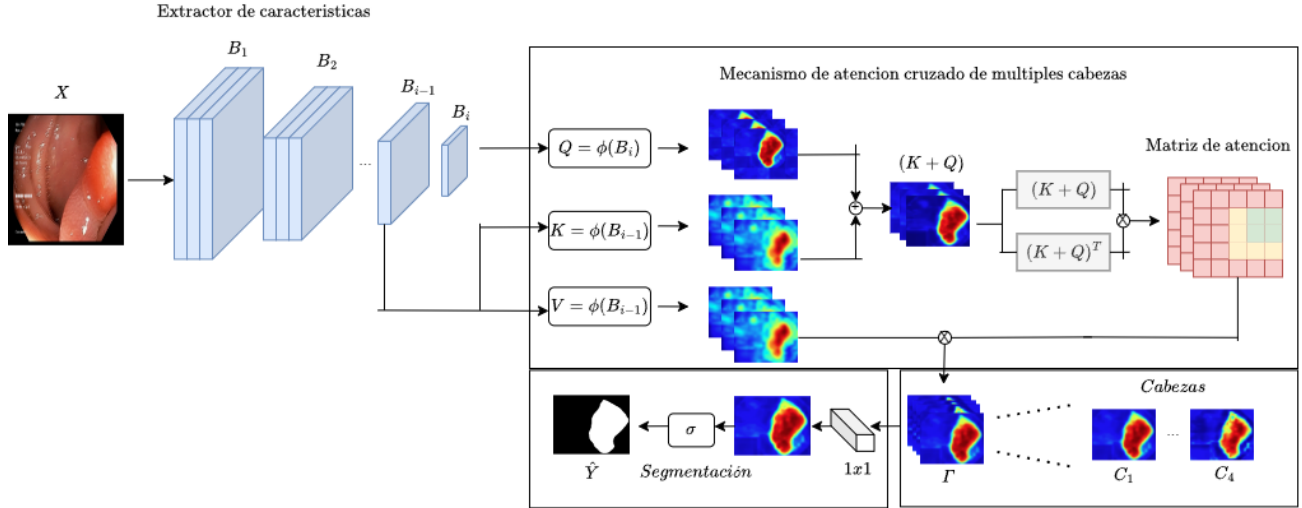
Particularmente, la arquitectura recibe como entrada una imagen de la colonoscopia  $X$  que se proyecta a un bloque convolucional para extraer un banco de características  $\mathcal{B} = \{B_1, B_2, \dots, B_{i-1}, B_i\}$ . En este trabajo, el bloque convolucional corresponde a un modelo Res2Net-50 preentrenado<sup>44</sup>.

A partir de esto, tomamos las dos últimas salidas ( $B_{i-1}(X)$  y  $B_i(X)$ ) para ser la entrada del mecanismo de atención cruzada de múltiples cabezas. Las ramas de procesamiento de entrada a cada mecanismo de atención están conformadas por bloques RFB  $\phi$  que implementan convoluciones  $1 \times 1$ ,  $3 \times 3$ ,  $5 \times 5$  y  $7 \times 7$  de manera local y dilatada para codificar

---

<sup>44</sup> Shang-Hua GAO et al. "Res2net: A new multi-scale backbone architecture". In: *IEEE transactions on pattern analysis and machine intelligence* 43.2 (2019), pp. 652–662.

**Figura 8.** Arquitectura propuesta por Ruiz *et. al.* para la segmentación de cáncer colorrectal.



los patrones geométricos relevantes de la entrada  $X$ . Para cada mecanismo de atención se definen tres tensores  $Q$ ,  $K$  y  $V$ , que aprovechan una representación intermedia RFB y sirven como entrada para conformar cada mecanismo de atención. Es decir, cada rama de procesamiento es mapeada como:  $Q = \phi(B_i(X))$ ,  $K = \phi(B_i(X))$  y  $V = \phi(B_i(X))$ .

En cuanto a los mecanismos de atención, los tensores  $Q$  y  $K$  son concatenados en un bloque  $((Q + K))$ , se transpone  $((Q + K)^T)$  y multiplica con si misma para obtener la matriz de atención que correlaciona la información de forma no local, como:  $att = \sigma((Q + K) \times (Q + K)^T)$ , donde estas operaciones ocurren respecto a la dimensión de los canales y contiene la similitud del coseno entre cada uno de los píxeles de cada canal. A partir de la matriz de atención se ponderan los píxeles y canales que son más importantes. Entonces, la entrada  $V$  se multiplica con la matriz  $att$ , obteniendo por módulo de atención (cada cabeza):  $C_i = V * att$ . Esta operación se puede hacer en paralelo para múltiples cabezas en un tensor  $\Gamma = \{C_1, C_2, \dots, C_i\}$ , para finalmente ser mapeados con una convolución  $1 \times 1$  con la función de activación sigmoide dando como resultado la segmentación del pólipo  $\hat{Y} = \sigma(Conv_{1 \times 1}(\Gamma))$ .

## 4.2. Entrenamiento semisupervisado para la detección de cáncer colorrectal

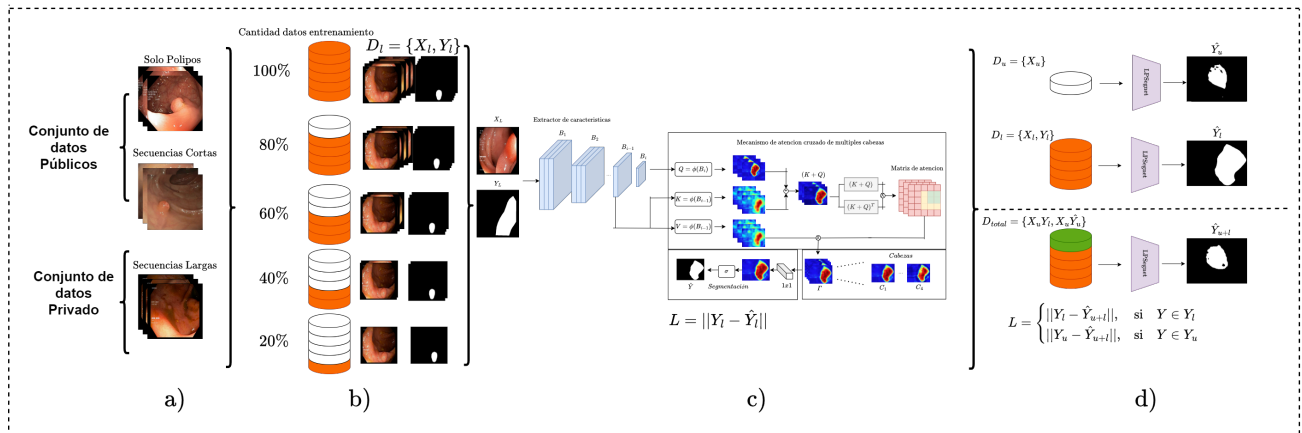
La detección de pólipos en la literatura se ha limitado a esquemas completamente supervisados, limitándose a pocos datos, el uso de observaciones de la misma institución y sin tener en cuenta mecanismos para refinar las estrategias con nuevos datos. En este trabajo se propuso, implementó y evaluó una estrategia de aprendizaje semisupervisado que utiliza dos particiones de datos: un conjunto de datos etiquetados ( $D_l = \{X_l, Y_l\}$ ) y un conjunto de datos sin etiquetar ( $D_u = \{X_u\}$ ).

Particularmente, la estrategia semisupervisada se entrenó con el conjunto  $D_l$  la representación profunda  $F_\theta(X_l) \rightarrow Y_l$ . Con este ajuste se generan pseudoetiquetas para el conjunto  $D_u$ , como,  $F_\theta(X_u) = \hat{Y}_u$ . Posteriormente estos nuevos datos etiquetados se unen al conjunto  $D_l$  para refinar el modelo de aprendizaje profundo mediante un reentrenamiento (*fine-tuning*) a partir de las etiquetas conocidas y pseudoetiquetas. En esta metodología se ajustó la arquitectura LPSegnet ( $F_\theta$ ), usando en el conjunto de entrenamiento, únicamente un subconjunto etiquetado (de cinco folds), y propagando la representación en los otros subconjuntos.

Entrenando la arquitectura base ( $F_\theta$ ) se hace el ajuste de parámetros del modelo con los datos etiquetados  $D_l$ , aprendiendo patrones relacionados con características generales de los pólipos (Figura 9-c). Luego, este modelo entrenado es utilizado para generar pseudoetiquetas  $F_\theta(X_u) = \hat{Y}_u$  confiables, siendo seleccionadas a partir de una regla definida a partir de las segmentaciones generadas  $\hat{Y}_l$  (Figura 9-d). Para este trabajo, se utilizó como regla de confianza el valor promedio de píxeles segmentados para determinar si la clase es pólipo, o de lo contrario se deja la máscara totalmente oscura para indicar que es fondo. Las pseudoetiquetas confiables generadas se concatenan al conjunto de datos etiquetado  $D_{total} = \{X_l, Y_l\} \cup \{X_u, \hat{Y}_u\}$  para realizar un reentrenamiento (*fine-tuning*) del modelo y refinar los parámetros aprendidos con nuevos pólipos que no poseen una máscara hecha por un experto. La motivación de este enfoque se debe a utilizar información parcial de los datos, permitiendo generar un modelo de segmentación generalizable en el mismo dominio

para datos no vistos y reduciendo costos ya que no es necesario el etiquetado por completo de los vídeos por parte de un experto gastroenterólogo.

**Figura 9.** Metodología semisupervisada propuesta.



### 4.3. Entrenamiento autosupervisado y detección de pólipos como anomalías

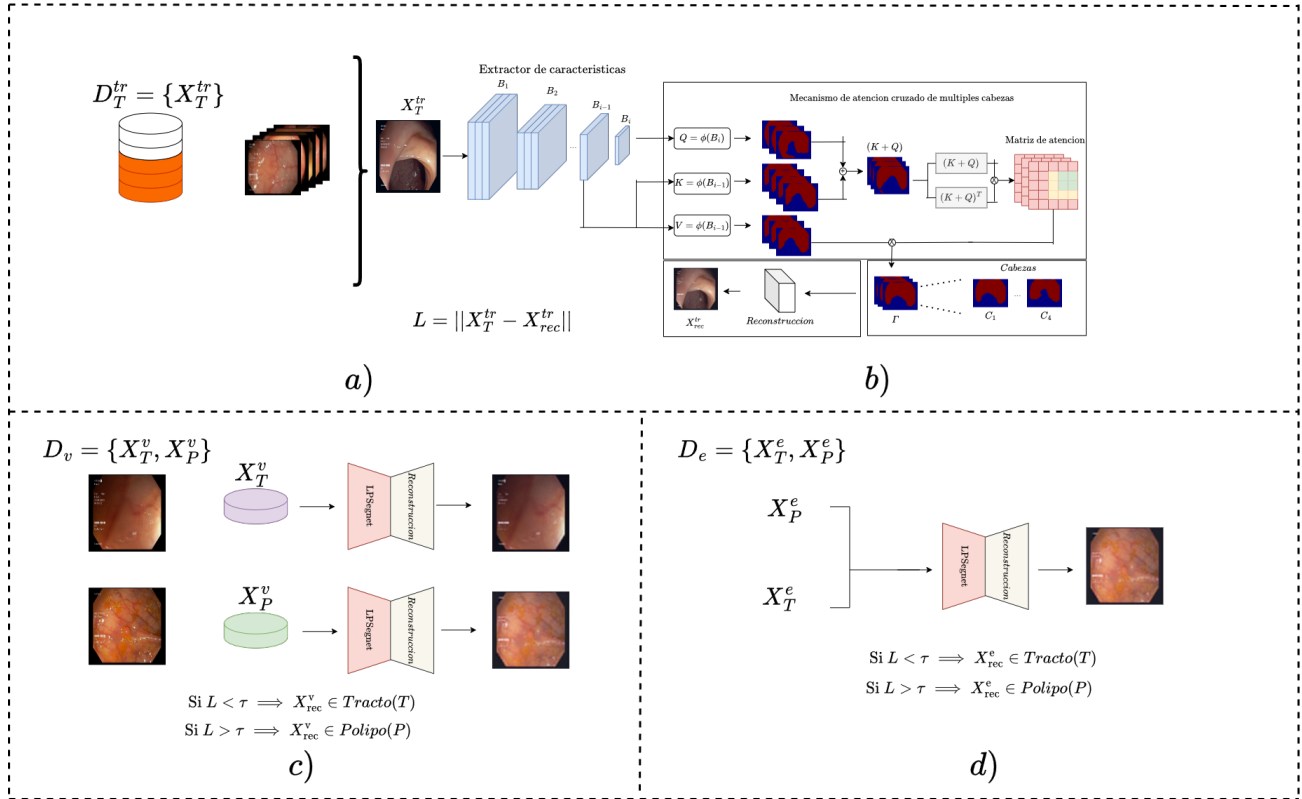
En este trabajo, también se exploró, propuso e implemento una estrategia autosupervisada para la caracterización de pólipos como anomalías, que no han sido observados durante el entrenamiento. De hecho, durante los procedimientos de colonoscopia, la mayoría de los fotogramas (mas del 90%) corresponden a tracto intestinal, en comparación con los fotogramas con pólipos. Computacionalmente, este hecho implica un desbalance natural de clases para la detección de pólipos. Entonces resulta conveniente asumir que una arquitectura puede aprender la variabilidad del tracto intestinal, a lo largo de la navegación colonoscópica, considerando el aprendizaje de una sola clase. Así, de este modo, durante la evaluación y puesta en funcionamiento de la arquitectura, cuando se observa un fotograma con un pólipo, se debe generar una alerta en el funcionamiento de la arquitectura, por una observación anormal, lo cual puede ser un indicador de detección de las masas. Este enfoque permite la valoración de pólipos con una amplia variabilidad en forma, margen y representación textural.

Para lograr este enfoque, primero se extendió la arquitectura base para que sea operable en un esquema autosupervisado, *i.e.*, sin etiquetas relacionadas con el pólipo. Particularmente, a la arquitectura LPSegnet se le acondicionó en la salida un bloque convolucional que se encarga de resumir los canales de  $\Gamma$  en una imagen de salida  $\hat{X}$ . Así, la arquitectura implementada retorna en la salida una imagen de tres canales, permitiendo realizar ahora una tarea autosupervisada de reconstrucción del tracto observado. Esta arquitectura es entonces ajustada con etiquetas de una sola clase (fotogramas del tracto intestinal sin pólipos), siguiendo una función de pérdida  $L = \|X - \hat{X}\|_1$  donde  $F_\theta(X) = \hat{X}$  y  $\hat{X} \rightarrow X$ . En este caso, la red puede aprender una alta variabilidad de observaciones control del tracto intestinal, incluyendo variaciones de luz y artefactos encontrados durante la colonoscopia. Además, este entrenamiento puede aprovechar del amplio número de ejemplos disponibles de navegación en las colonoscopias convencionales.

Una vez entrenada la arquitectura, entonces los pólipos proyectados deben tener reconstrucciones erróneas y su cuantificación conduce a la detección de los mismos. Específicamente, se utilizó un conjunto de datos de entrenamiento  $D_T^{tr} = \{X_T\}^{n_{tr}}$  que contiene únicamente imágenes con tracto intestinal (Figura 10-a), el cual permite el ajuste de la arquitectura. Este conjunto de datos se utilizó para entrenar al modelo en reconstruir imágenes y patrones de un tracto intestinal sano (Figura 10-b). Luego, utilizando un conjunto de validación  $D_v = \{X_T^v, X_P^v\}^{n_v}$  que corresponde al conjunto de datos de tractos ( $T$ ) y pólipos ( $P$ ), se cuantifico el nivel de reconstrucción admisible para determinar un umbral de reconstrucción  $\tau$ . este valor de umbral es el límite a partir de la cual una reconstrucción se empieza a considerar anómala (Figura 10-c).

El paso final de esta red corresponde a utilizar el conjunto de datos  $D_e = \{X_T^e, X_P^e\}^{n_e}$  para evaluar con imágenes nunca vistas en entrenamiento ni validación, el modelo y su viabilidad en escenarios clínicos de colonoscopias (Figura 10-d). La motivación de este enfoque se debe a explorar una tarea de aprendizaje totalmente no supervisada, permitiendo generar un modelo capaz de detectar pólipos sin la intervención de un experto gastroenterólogo en

**Figura 10.** Metodología auto-supervisada para la detección de pólipos como anomalías



el procedimiento ahorrando en costos de etiquetados. Este enfoque de entrenamiento utiliza el conjunto de datos COLON únicamente ya que posee secuencias largas de colonoscopia y una gran cantidad de datos de tracto intestinal.

## 5. Diseño Experimental

### 5.1. Conjunto de datos de colonoscopias

**5.1.1. Conjunto de datos publico** Las arquitecturas consideradas en este trabajo se entrenaron y validaron en secuencias de imágenes de pólipos y fondo utilizando diversos conjuntos de datos públicos. La selección de estos conjuntos públicos permitió la comparación con otros métodos del estado del arte y la evaluación de la competitividad de los enfoques propuestos. Los conjuntos de datos utilizados fueron los siguientes:

- **ASU-Mayo Clinic Colonoscopy Video:** Este conjunto de datos liberado en 2015 en el desafío Automatic Polyp Detection. Este conjunto de datos contiene vídeos de 10 secuencias de colonoscopia con pólipos y 10 vídeos con solamente tracto intestinal, ambos grupos contiene su correspondiente anotación. Adicionalmente este conjunto de datos contiene 18 vídeos para testeo sin anotaciones<sup>45</sup>.
- **ETIS-Larib:** Este conjunto de datos también fue liberado en el desafío Automatic Polyp Detection, este conjunto de datos contiene 196 imágenes de pólipos con anotaciones que fueron extraídos de 34 colonoscopias diferentes<sup>46</sup>.
- **CVC-Clinic DB:** De igual forma fue liberado en el desafío Automatic Polyp Detection, contiene 612 fotogramas con pólipo con anotación que fueron extraídos de 31 colono-

---

<sup>45</sup> Nima TAJBAKHSI; Suryakanth R GURUDU, and Jianming LIANG. “Automated polyp detection in colonoscopy videos using shape and context information”. In: *IEEE transactions on medical imaging* 35.2 (2015), pp. 630–644.

<sup>46</sup> Juan SILVA et al. “Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer”. In: *International journal of computer assisted radiology and surgery* 9 (2014), pp. 283–293.

scopias diferentes<sup>47</sup>.

- **CVC-Video Clinic DB:** Este conjunto de datos fue liberado en el desafío Endoscopic Vision Challenge en el 2018, este conjunto de datos contiene 18 colonoscopias diferentes con una duración de alrededor de 1 minuto, todas las secuencias contienen pólipos y están con su respectiva anotación, adicionalmente este conjunto de datos contiene 18 vídeos para validación sin anotación<sup>4849</sup>.
- **Kvasir-SEG:** Este conjunto de datos fue liberado en el 2020, contiene alrededor de 1000 fotogramas con pólipos anotados que fueron seleccionados de diferentes procedimientos de colonoscopia<sup>50</sup>.

**5.1.2. Conjunto de secuencias largas de colonoscopia: COLON** En este trabajo también se utilizó un conjunto de secuencias largas de colonoscopia (denominado COLON) . El conjunto de datos COLON esta compuesto por secuencias largas de colonoscopia de 35 vídeos grabados con la cámara Olympus Evis Excera III. Cada vídeo tiene una resolución espacial de  $480 \times 720$  píxeles y una resolución temporal de 30 *fps*. En total, se cuenta con alrededor de 30.000 pólipos anotados en estas secuencias, mientras que el resto de los fo-

---

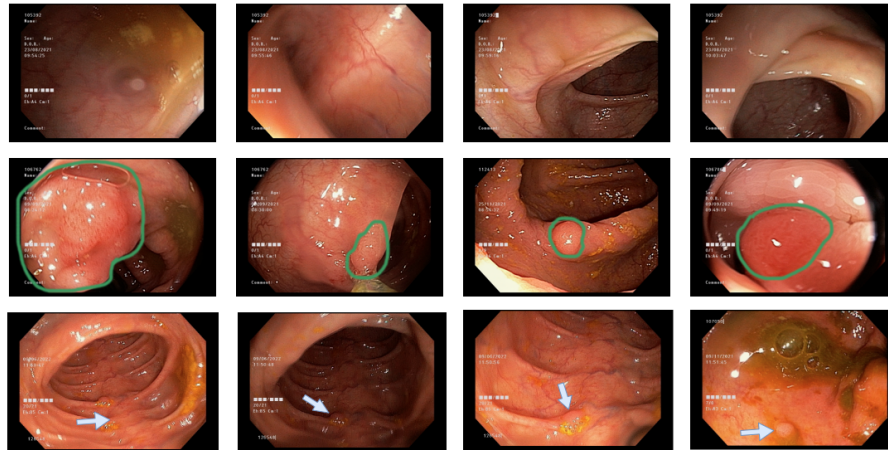
<sup>47</sup> Jorge BERNAL et al. "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians". In: *Computerized medical imaging and graphics* 43 (2015), pp. 99–111.

<sup>48</sup> Quentin ANGERMANN et al. "Towards real-time polyp detection in colonoscopy videos: Adapting still frame-based methodologies for video sequences analysis". In: *Computer Assisted and Robotic Endoscopy and Clinical Image-Based Procedures: 4th International Workshop, CARE 2017, and 6th International Workshop, CLIP 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, 2017, Proceedings 4*. Springer. 2017, pp. 29–41.

<sup>49</sup> Jorge J BERNAL et al. "Polyp detection benchmark in colonoscopy videos using gcreator: A novel fully configurable tool for easy and fast annotation of image databases". In: *Proceedings of 32nd CARS conference*. 2018.

<sup>50</sup> Konstantin POGORELOV et al. "Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection". In: *Proceedings of the 8th ACM on Multimedia Systems Conference*. 2017, pp. 164–169.

**Figura 11.** Conjunto de secuencias largas de colonoscopia COLON, la primer fila son fotogramas de tracto intestinal, la segunda fila pólipos con su respectiva delineación y la tercer fila pólipos que pueden llegar a ser confundidos con tracto intestinal



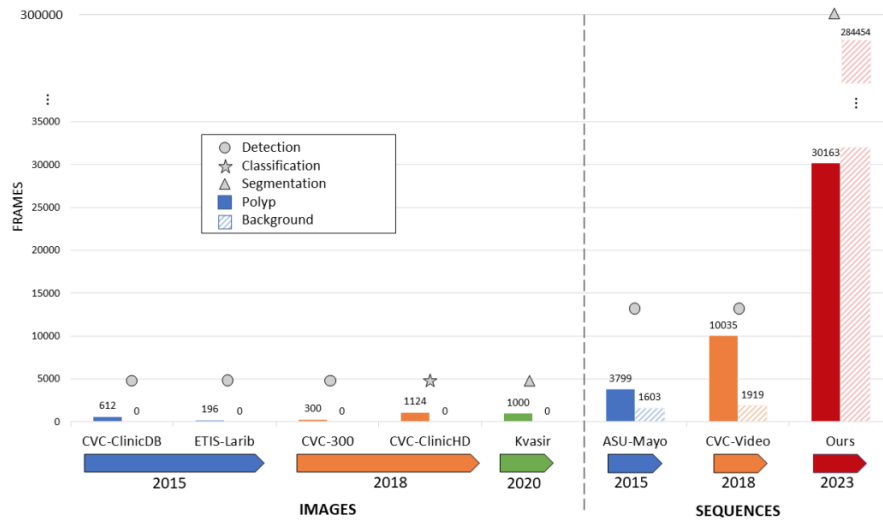
togramas corresponden al tracto intestinal (fondo). El conjunto de datos COLON fue creado por el grupo de investigación BIVL<sup>2</sup>ab en colaboración con el Instituto de Gastroenterología y Hepatología del Oriente IGHO S.A.S <sup>51</sup>. Este conjunto de datos simula un procedimiento de colonoscopia real, ya que incluye variaciones como movimientos de cámara, preparación intestinal deficiente y fatiga del gastroenterólogo durante el proceso. La Figura 11 muestra diferentes fotogramas de secuencias de colonoscopia, en la primera fila se muestran fotogramas de tracto intestinal, en la segunda fila fotogramas de pólipos y en la tercera fotogramas de pólipos que pueden llegar a confundirse como fondo<sup>51</sup>.

Este conjunto de datos contiene la mayor cantidad de datos etiquetados entre las clases de tracto intestinal y pólipo, en comparación con los conjuntos de datos públicos. La Figura 12 muestra una comparación entre los diferentes conjuntos de datos utilizados en este trabajo, donde el conjunto de datos COLON supera significativamente a los demás<sup>51</sup>. Adicionalmente, COLON presenta un desbalanceo natural de etiquetas, donde solo alrededor del

---

<sup>51</sup> Lina RUIZ et al. "COLON: The largest COLonoscopy LONG sequence public database". In: *arXiv preprint arXiv:2403.00663*. 2024.

**Figura 12.** Comparación entre los conjuntos de datos públicos y el conjunto de datos de secuencias largas de colonoscopia COLON, tomado de: *COLON: The largest COLonoscopy LONG sequence public database*



10% de los fotogramas contienen pólipos. Este desbalanceo es realista a lo observado en la rutina clínica y además es propicio para entrenar y validar las estrategias de aprendizaje semisupervisado y de anomalías propuestas y emular el comportamiento en un ambiente real.

## 5.2. Hiperparámetros de la arquitectura propuesta

El entrenamiento de la arquitectura propuesta, tanto para el enfoque semisupervisado como para el de detección de anomalías, se realizó utilizando los hiperparámetros de la Tabla 1. En cuanto a la arquitectura, se establecieron 4 cabezas de atención, con un total de 16 bloques RFB (4 por cabeza). Se utilizó una Res2Net-50<sup>44</sup> con 50 bloques convolucionales junto con bloques residuales para extraer las características locales de la imagen de entrada. Adicionalmente, cada bloque RFB incluye convoluciones independientes (una normal y otra dilatada) con *kernels* de  $1 \times 1$ ,  $3 \times 3$ ,  $5 \times 5$  y  $7 \times 7$ , respectivamente. Las ramas de las convoluciones se integran en una sola mediante una convolución  $1 \times 1$ .

**Tabla 1.** Hiperparámetros de entrenamiento de arquitectura

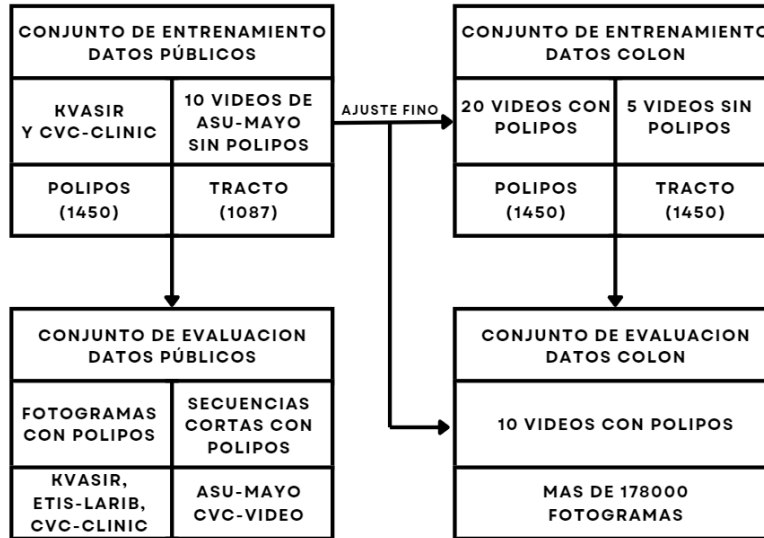
<b>Parámetros de entrenamiento</b>	<b>Semi-supervisado</b>	<b>Detección de anomalías</b>
Época	80	80
Tasa de aprendizaje	1e-4 / 1e-6	1e-4 / 1e-6
Tamaño del lote	8	8
Tamaño del entrenamiento	352	352
Tasa de decaimiento	0.1	0.1
Época de decaimiento	50	50

### 5.3. Validación

**5.3.1. Enfoque semi-supervisado** Este enfoque de aprendizaje semi-supervisado se validó en dos etapas. En la primera etapa, se entrenó el modelo con 2.900 fotogramas balanceados de pólipos y tracto intestinal de los conjuntos de datos públicos Kvasir-SEG, CVC-Clinic y ASU-Mayo. En la segunda etapa, se evaluó el rendimiento del enfoque en dos aspectos. Primero, se validó el rendimiento del modelo en los datasets Kvasir-SEG, ETIS-Larib, CVC-Clinic, ASU-Mayo y CVC-Video. Segundo, se comparó la capacidad de generalización del modelo entrenado con datos públicos y la misma arquitectura después de un reentrenamiento (*fine-tuning*) con 25 vídeos de entrenamiento del conjunto de datos COLON. El proceso de validación se ilustra en la Figura 13. Para el entrenamiento, se realizaron 5 particiones aleatorias de muestras etiquetadas y no etiquetadas (20%, 40%, 60%, 80% y 100% de datos etiquetados). Durante el proceso, se generaron pseudoetiquetas confiables a partir de una regla de confianza basada en el número promedio de píxeles detectados como pólipo y tracto intestinal para generar muestras con y sin pólipo. En ningún caso se utilizó la etiqueta real asociada a la muestra.

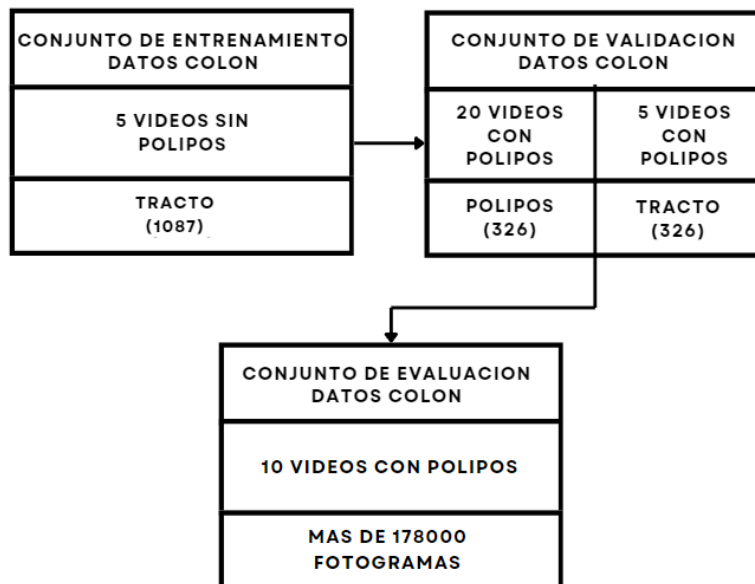
**5.3.2. Enfoque de detección de anomalías** La validación del enfoque de detección de anomalías se realizó en tres etapas, utilizando únicamente el conjunto de datos COLON. En la primera etapa, se entrenó al modelo con 5 vídeos que no contienen pólipos para que aprendiera a modelar el tracto intestinal. En la segunda etapa, se utilizaron 20 vídeos con

**Figura 13.** Entrenamiento y validación con los conjunto de datos para el enfoque semi-supervisado



pólipos y 5 de fondo para determinar un umbral que separa ambas clases. Este umbral se calculó a partir de las curvas de precisión y sensibilidad, buscando un punto intermedio que sirviera como límite en el error del modelo para determinar qué fotogramas corresponden a pólipos (anómalos) y cuáles a fondo. En los resultados se reporta el comportamiento ante estos cambios. En la tercera etapa, se validó el modelo con 10 vídeos del conjunto de prueba de COLON, que contienen pólipos y fondo, con el objetivo de evaluar el rendimiento del umbral y del modelo para distinguir entre fotogramas con y sin pólipos. El proceso de validación se ilustra en la Figura 14.

**Figura 14.** Entrenamiento y validación con los conjunto de datos para el enfoque anómalo



## 6. Evaluación y Resultados

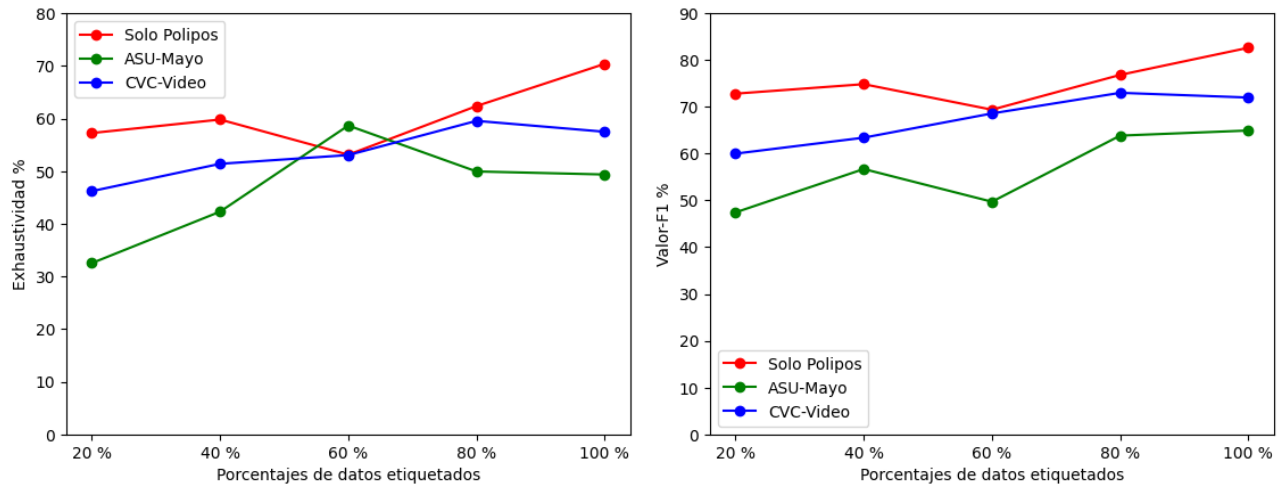
En este trabajo se exploraron dos alternativas semisupervisadas para detectar y localizar pólipos en secuencias colonoscópicas, priorizando principalmente el trabajo en escenarios reales, con secuencias que contienen fotogramas sin este biomarcador y que reportan alta variabilidad textural para representar el tracto rectal. Las dos estrategias de aprendizaje propuestas en este trabajo fueron validadas utilizando conjuntos de datos públicos compuestos por secuencias cortas de colonoscopia y fotogramas de pólipos únicamente. También se validó con un conjunto de datos COLON compuesto por secuencias largas de colonoscopia para validar los métodos en entornos reales. A continuación, se detallan los resultados obtenidos sobre las dos estrategias de aprendizaje que se utilizaron.

### 6.1. Detección de pólipos siguiendo un esquema semi-supervisado

En esta aproximación, fue incluido una implementación de aprendizaje semisupervisado sobre una arquitectura que incluye modelos de atención para localizar y segmentar los pólipos. Las métricas empleadas fueron la Sensibilidad y el Valor F1, ya que nos indican si el modelo genera falsos negativos y un promedio entre falsos positivos y negativos, respectivamente. En este contexto, estamos interesados en desarrollar un soporte computacional que permita generar alertas sobre masas sospechosas, durante la colonoscopia, hecho crítico en la valoración de cáncer colorrectal y una de las principales razones de la incidencia de hallazgos en estados tardíos.

Como experimento inicial se validó con tres diferentes conjuntos de datos públicos (descripción de los datos son expuestos en la Figura 13). Para ello se siguió un experimento que toma diferentes conjuntos de datos con etiquetas y propaga el resto de datos, dando un carácter semisupervisado y permitiendo generalizar la representación ante escenarios dinámicos, donde las etiquetas puede irse agregando continuamente. En la Figura 15, se

**Figura 15.** Métricas obtenidas con los conjuntos de datos públicos.

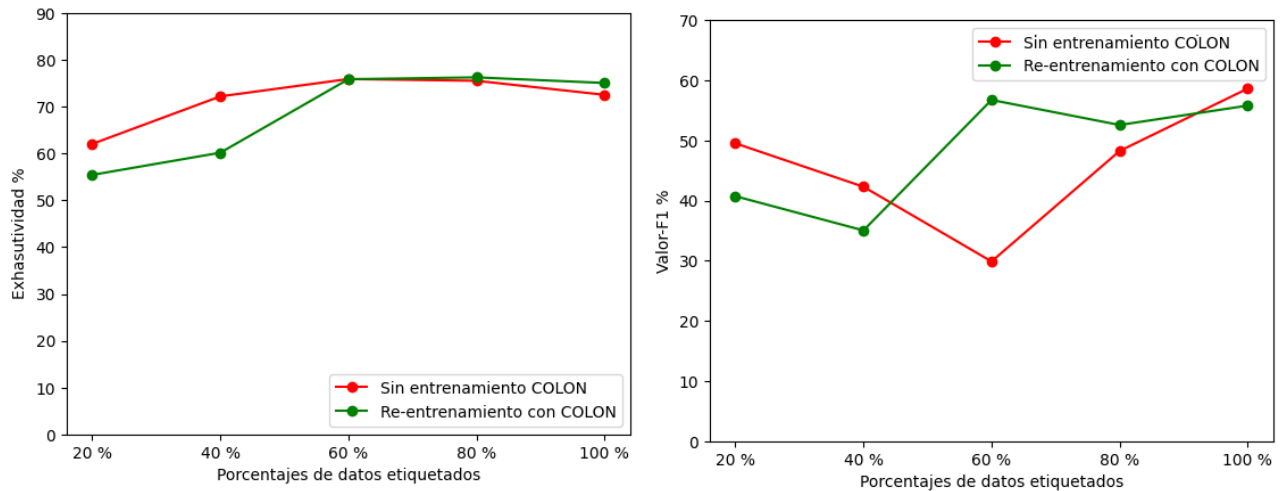


muestran los resultados obtenidos por el modelo en cada una de las particiones de datos etiquetados. Se observa que una mayor cantidad de datos etiquetados incluidos en el entrenamiento logran mejorar el modelo propuesto.

Como era de esperarse, el comportamiento con mejor Exhaustividad y Valor F1, en la mayoría de casos, es logrado con el 100% de los datos etiquetados, logrando así un ajuste completo de la arquitectura a cada conjunto de datos particular. Sin embargo, resulta interesante que el método propuesto con el conjunto de datos de ASU-Mayo tiene el mejor comportamiento en Exhaustividad con el 60% de los datos etiquetados. También se observa una respuesta creciente en el con el conjunto de datos "solo pólipos" a partir de utilizar el 60% de los datos etiquetados junto a pseudoetiquetas. Este hecho resulta interesante, porque el método tiene una eficiencia superior al 70% sin usar la totalidad de los cuadros, y reconociendo que estas son secuencias cortas, podríamos suponer que lograría un comportamiento más generalizable al enfrentarse con escenarios más diversos.

En una segunda etapa de la validación del esquema semisupervisado, se procedió a realizar un entrenamiento y validación en un escenario real, donde se cuentan con extensos videos de colonoscopia, cuyos fotogramas en su mayoría corresponden al tracto intestinal

**Figura 16.** Métricas obtenidas con el conjunto de datos COLON



(ver descripción del conjunto de datos en la Figura 16). En este caso, se exploraron dos alternativas de entrenamiento y evaluación. En la primera configuración de la arquitectura, la arquitectura tomada para evaluar el conjunto de datos de COLON correspondió a la previamente ajustada por los datos públicos. En este caso, la arquitectura no observó durante el entrenamiento los cuadros del conjunto de datos COLON, lo cual resulta provechoso para evaluar la capacidad de generalización de la arquitectura. En una segunda configuración, "Re-entrenamiento con COLON" se utilizó un subconjunto de entrenamiento del conjunto de datos COLON para ajustar finamente la arquitectura.

En la figura 16 se reportan los resultados obtenidos en las secuencias largas de colonoscopia, validado con respecto a los valores de Exhaustividad y Valor F1. Cabe destacar en estos resultados el comportamiento notable de la arquitectura que no observó durante entrenamiento los datos de las nuevas colonoscopias (línea roja, etiqueta "sin entrenamiento COLON"). Como se esperaba, el enfoque semisupervisado (utilizando apenas el 60% de datos de entrenamiento) superó logró un 75,9 % de Exhaustividad y un Valor F1 de 29.87%. Este resultado resulta interesante porque valida la hipótesis de generalización del enfoque propuesto. A validar el modelo después de haber sido entrenado con los conjuntos de datos

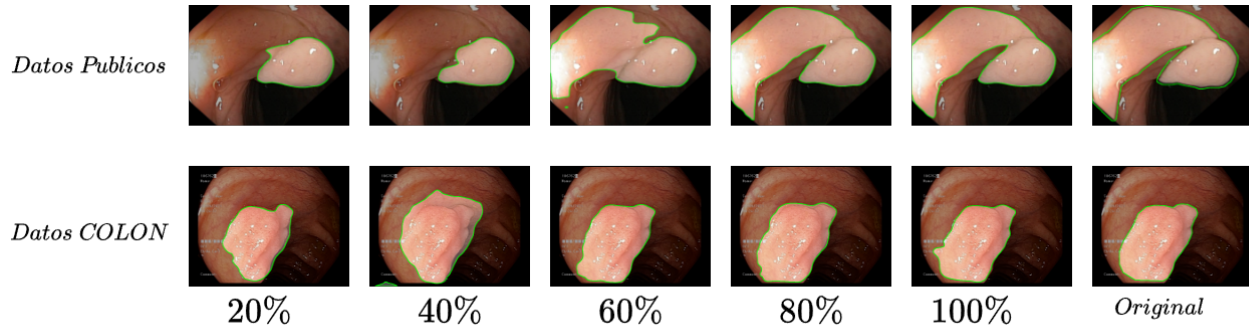
públicos y reentrenado con el conjunto de datos COLON, se observa una Exhaustividad del 75,94% al utilizar el 60% de datos etiquetados. Esto indica que la representación es robusta al cambiar de dominio en los fotogramas, y que es posible utilizar desde un 60% de datos etiquetados para generar nuevas etiquetas fiables que eviten caracterizar pólipos como fondo. Sin embargo, se observa una pérdida de rendimiento en el Valor F1, debido a que el modelo está obteniendo más falsos positivos para este conjunto de datos, es decir, elementos del tracto intestinal detectados como pólipos. También resulta interesante los valores logrados con el reentrenamiento que se van ajustando a la representación y darían pie a en futuros trabajos a definir esquemas alternativos de ajuste, sin que estos conlleven al sobreaprendizaje.

Cómo exploración cualitativa, en la Figura 17 se muestran dos ejemplos de marcación automática de pólipos para los conjuntos de datos públicos (primera fila) y el conjunto de datos COLON (segunda fila). Los fotogramas se evaluaron en las diferentes particiones de datos etiquetados junto con pseudoetiquetas con las que se entrenó el modelo. En el caso del conjunto de datos público se evidencia que las particiones desde 20% hasta 60% de datos etiquetados no termina de segmentar por completo el pólipo, sin embargo, las particiones 80% y 100% muestran una segmentación similar a la original, por lo que una diferencia de 20% no es significativa para detectar pólipos en secuencias cortas de colonoscopia. En comparación, al evaluar las segmentaciones en el conjunto de datos COLON se evidencia que las segmentaciones al usar al menos 60% de los datos es similar a la original, por lo que el modelo propuesto tiene la capacidad de modelar patrones geométricos y texturales que diferencian los pólipos del tracto intestinal.

## **6.2. Detección de pólipos siguiendo un esquema de detección de anomalías**

El modelo propuesto para la detección de anomalías se evaluó utilizando dos arquitecturas clásicas del estado del arte: un autocodificador convencional y una arquitectura típica U-Net. El autocodificador está compuesto de 5 capas convolucionales en el codificador, un

**Figura 17.** Predicciones obtenidas en particiones de datos etiquetados que se utilizaron para entrenar la arquitectura junto con pseudoetiquetas.



vector embebido de 32768 valores escalares y un número de 5 capas convolucionales en el decodificador. Por otra parte, la arquitectura U-Net usada como línea base tiene un total de 4 capas en el codificador, un vector embebido de dimensión (1024, 7, 11) y un total de 4 capas en el decodificador. Estas arquitecturas base fueron ajustadas en la tarea de reconstrucción de fotogramas con el tracto intestinal.

Debido a la gran cantidad de datos y el desbalance natural del conjunto de datos COLON, el entrenamiento se realizó con datos de tracto intestinal con el objetivo de modelar características morfológicas y texturales del intestino. El umbral  $\tau$  se calculó utilizando la curva de Precisión y Exhaustividad, evaluada en un rango de umbrales que abarca el mínimo y máximo de la función de pérdida (MAE). Primeramente se definió el umbral  $\tau_2$  como la media entre la precisión y la sensibilidad de los errores que se obtuvieron en validación. Adicionalmente, se utilizaron dos umbrales equidistantes  $\tau_1$  que disminuyó la detección de falsos negativos y  $\tau_3$  que disminuyó la detección de falsos positivos. Con los umbrales definidos, se realizó la detección de pólipos en el conjunto de evaluación.

En la tabla 2 se reportan los resultados de Especificidad y Exhaustividad (conocido en inglés como *Recall*) obtenidos mediante tres umbrales diferentes tanto para los modelos autosupervisados de la línea base: Autocodificador ( $\tau_1 = 31.62658$ ,  $\tau_2 = 32.62658$ ,  $\tau_3 = 33.62658$ ) y U-Net ( $\tau_1 = 35.0069$ ,  $\tau_2 = 36.0069$ ,  $\tau_3 = 37.0069$ ), como por el método propuesto ( $\tau_1 =$

14.9918,  $\tau_2 = 13.6621$  ,  $\tau_3 = 16.3215$ ). También se ubicó en la tabla los mejores resultados obtenidos en el enfoque semisupervisado con el conjunto de datos COLON, para realizar un análisis comparativo entre estas estrategias débilmente supervisadas. Cabe destacar, que para la selección del umbral  $\tau_1 = 14.9918$ , se logró obtener una Exhaustividad de 84,92%, en cuanto a la detección de masas sospechosas. Este valor resulta relevante, a pesar de la baja Especificidad, teniendo en cuenta la importancia en el contexto de reconocer positivamente pólipos, durante la navegación de la colonoscopia. Además, compensando la arquitectura con otros valores de umbral  $\tau$ , podemos encontrar un mayor balance entre la exhaustividad y la especificidad. El método propuesto, de hecho, supera ampliamente a los enfoques definidos como línea base, logrado una mayor adaptación al problema de detección de pólipos y lidiando mejor con la alta variabilidad observacional del tracto intestinal. También el método tiene un comportamiento alto en exhaustividad comparado con lo logrado por el método supervisado, pero el método semisupervisado tiene un mejor desempeño en la especificidad. En este caso resulta oportuno poder aprovechar estas representaciones y en un trabajo futuro poder explorar las dos condiciones de forma conjunta.

**Tabla 2.** Comparación entre diferentes arquitecturas del estado del arte y las metodologías propuestas.

<b>Arquitectura</b>	<b>Exhaustividad [%]</b>	<b>Especificidad [%]</b>
Semisupervisado 60%	75,94	65,38
Semisupervisado 100%	72,56	92,19
Autocodificador $\tau_1$	18,60	75,89
Autocodificador $\tau_2$	10,97	83,92
Autocodificador $\tau_3$	5,62	90,17
U-Net $\tau_1$	30,93	63,88
U-Net $\tau_2$	24,14	73,18
U-Net $\tau_3$	15,74	81,16
Metodología Propuesta $\tau_1$	84,92	15,54
Metodología Propuesta $\tau_2$	75,80	27,41
Metodología Propuesta $\tau_3$	62,31	41,26

## 7. Conclusiones y trabajo futuro

En este trabajo se exploró esquemas alternativos de entrenamiento y ajuste de representaciones computacionales que soporte la tarea de localización de pólipos, en contextos desbalanceados, con pocas muestras del objeto de interés. Es así, que en este trabajo se presentaron dos estrategias de aprendizaje débilmente supervisado, que ajustaron una arquitectura convolucional, que incluye mecanismos de atención y que demostró ser eficaz en la tarea de segmentación y discriminación de pólipos. En primer lugar, la estrategia semisupervisada exploró la capacidad de una representación profunda, para adaptarse y codificar patrones de pólipos, usando un subconjunto limitado de datos. Esto es razonable y coherente con el contexto clínico, que requiere no solo entrenar representaciones, sino que pueda usar nuevos datos que no estén estrictamente anotados. En este caso, la arquitectura propuesta propaga nuevas etiquetas en datos sin marcaciones, y luego uso las pseudoetiquetas con más confianza para ganar capacidades de generalización. El método demostró tener un comportamiento aceptable usando solo el 60% de los datos. El comportamiento con mejor Exhaustividad y Valor F1 se obtiene utilizando el 100% de las etiquetas, pero de manera interesante en algunos de los conjuntos de datos públicos el modelo obtiene mejores Valores F1 a partir de utilizar 60% de etiquetas junto con pseudoetiquetas, revelando que el modelo tiene una eficacia del 70% sin usar la totalidad de los fotogramas con etiquetas teniendo varios conjuntos de datos, lo cual nos lleva a un comportamiento más generalizable en escenarios más diversos de colonoscopias. De hecho, en una validación sobre un conjunto de datos más cercano a las colonoscopias en la rutina clínica y sin requerir ningún entrenamiento extra, el método propuesto demostró tener capacidades de generalización en la tarea de detección de pólipos.

En un segundo método, en este trabajo se validó un esquema de aprendizaje autosupervisado, siguiendo una tarea de reconstrucción, y detectando pólipos bajo un criterio de muestras anómalas. En esta alternativa de aprendizaje se validó la tarea de aprender sola-

mente de cuadros del tracto intestinal que no contienen, pólipos, los cuales son redundantes en un procedimiento de colonoscopia. Entonces, la arquitectura profunda se adaptó para que pudiera reconstruir los cuadros de entrada y no requiriera una tarea de supervisión. Luego se entrenó la arquitectura y los fotogramas con pólipo, proyectados a la arquitectura, fueron detectados como fotogramas con errores significativos de reconstrucción.

Con este enfoque al ser entrenada con una única clase la arquitectura fue capaz de aprender sobre el fondo y cuando se validó con datos de ambas clases mediante los errores de reconstrucción (MAE) se obtuvo un umbral a partir de una curva de Exhaustividad y Precisión que evaluó diferentes valores de estas métricas en diferentes umbrales. Para este enfoque se utilizó tres umbrales, un umbral que aumentaba la Exhaustividad, otro umbral que equilibrio las métricas de Exhaustividad y Precisión y un tercer umbral que disminuyo el umbral, este último para disminuir la tasa de falsos positivos. De esta metodología se observó que el mejor resultado en cuanto a Exhaustividad que es del 84.92% se obtuvo al utilizar la metodología propuesta con un umbral  $\tau = 14.9918$ , que casi triplico lo obtenido en las arquitecturas clásicas como la U-Net o el autocodificador en diferentes umbrales e incluso supero a la metodología semisupervisada que utilizo el 60% de los datos junto pseudoetiquetas. Este enfoque de detección de anomalías propuesto es viable realizar este enfoque de entrenamiento para apoyar las tareas de los radiólogos en la identificación y caracterización de pólipos. sin embargo, los resultados obtenidos en Especificidad del 15.54% nos revela que se generaron una gran cantidad de falsos positivos lo cual puede llegar a aumentar la cantidad de fotogramas que el especialista deberá observar.

El trabajo realizado fue consistente y demostró que alternativas de aprendizaje pueden ser claves para poder transferir estas herramientas en ambientes reales. Sin embargo, como trabajos futuros se deben hacer análisis más detallados con respecto a diferentes grados de agresividad de la enfermedad, expresada en tamaños de los pólipos u otras características de forma y textura. También se requiere definir mecanismos que aprendan umbrales óptimos para la discriminación anómala y poner en prueba estas alternativas en un ambiente

real. Si bien, los métodos logran una apropiada detección de cuadros con pólipo, el cálculo recurrente de falsos positivos puede conducir a ser una herramienta de baja con fiabilidad en la rutina clínica. En trabajos futuros, buscamos utilizar otros conjuntos de datos para observar la capacidad de generalización con datos de colonoscopias obtenidos con otras máquinas. Así mismo, como trabajo futuro se deben explorar otras funciones de pérdidas que estén más asociadas a la tarea de segmentar pólipos, permitiéndole al modelo aprender y representar mejor el tracto intestinal, además de incluir otros módulos que complementen la tarea de reconstrucción para la detección de anomalías.

## BIBLIOGRAFÍA

AN, Jinwon and CHO, Sungzoon. “Variational autoencoder based anomaly detection using reconstruction probability”. In: *Special lecture on IE 2.1* (2015), pp. 1–18 (cit. on p. 23).

ANGERMANN, Quentin; HISTACE, Aymeric, and ROMAIN, Olivier. “Active learning for real time detection of polyps in videocolonoscopy”. In: *Procedia Computer Science* 90 (2016), pp. 182–187 (cit. on p. 12).

ANGERMANN, Quentin et al. “Towards real-time polyp detection in colonoscopy videos: Adapting still frame-based methodologies for video sequences analysis”. In: *Computer Assisted and Robotic Endoscopy and Clinical Image-Based Procedures: 4th International Workshop, CARE 2017, and 6th International Workshop, CLIP 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, 2017, Proceedings 4*. Springer. 2017, pp. 29–41 (cit. on p. 38).

BERNAL, Jorge et al. “Polyp segmentation method in colonoscopy videos by means of MSA-DOVA energy maps calculation”. In: *Clinical Image-Based Procedures. Translational Research in Medical Imaging: Third International Workshop, CLIP 2014, Held in Conjunction with MICCAI 2014, Boston, MA, USA, September 14, 2014, Revised Selected Papers 3*. Springer. 2014, pp. 41–49 (cit. on p. 25).

BERNAL, Jorge et al. “WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians”. In: *Computerized medical imaging and graphics* 43 (2015), pp. 99–111 (cit. on p. 38).

BERNAL, Jorge J et al. “Polyp detection benchmark in colonoscopy videos using gtcreator: A novel fully configurable tool for easy and fast annotation of image databases”. In: *Proceedings of 32nd CARS conference*. 2018 (cit. on p. 38).

BRANDAO, Patrick et al. “Fully convolutional neural networks for polyp segmentation in colonoscopy”. In: *Medical Imaging 2017: Computer-Aided Diagnosis*. Vol. 10134. Spie. 2017, pp. 101–107 (cit. on p. 25).

CARON, Mathilde et al. “Deep clustering for unsupervised learning of visual features”. In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 132–149 (cit. on p. 21).

CHANDOLA, Varun; BANERJEE, Arindam, and KUMAR, Vipin. “Anomaly detection: A survey”. In: *ACM computing surveys (CSUR)* 41.3 (2009), pp. 1–58 (cit. on p. 22).

CHEN, Siwei; URBAN, Gregor, and BALDI, Pierre. “Weakly Supervised Polyp Segmentation in Colonoscopy Images Using Deep Neural Networks”. In: *Journal of Imaging* 8.5 (2022), p. 121 (cit. on p. 27).

COTTON, Peter B et al. “Colonoscopy: practice variation among 69 hospital-based endoscopists”. In: *Gastrointestinal endoscopy* 57.3 (2003), pp. 352–357 (cit. on pp. 12, 15).

COTTON, Peter B et al. “Colonoscopy: practice variation among 69 hospital-based endoscopists”. In: *Gastrointestinal endoscopy* 57.3 (2003), pp. 352–357 (cit. on pp. 12, 14).

DONG, Bo et al. “Polyp-pvt: Polyp segmentation with pyramid vision transformers”. In: *arXiv preprint arXiv:2108.06932* (2021) (cit. on p. 26).

ERICSSON, Linus et al. “Self-supervised representation learning: Introduction, advances, and challenges”. In: *IEEE Signal Processing Magazine* 39.3 (2022), pp. 42–62 (cit. on pp. 16, 20, 21).

FAN, Deng-Ping et al. “Pranet: Parallel reverse attention network for polyp segmentation”. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VI*. Vol. 23. Springer International Publishing. 2020, pp. 263–273 (cit. on p. 26).

GAO, Shang-Hua et al. “Res2net: A new multi-scale backbone architecture”. In: *IEEE transactions on pattern analysis and machine intelligence* 43.2 (2019), pp. 652–662 (cit. on pp. 31, 40).

GRP, Colorectal Canc Collaborative et al. “Surgery for colorectal cancer in elderly patients: a systematic review”. In: *The Lancet* 356.9234 (2000), pp. 968–974 (cit. on p. 14).

GÜINDIC, Luis Charúa. “Cáncer de colon”. In: *Gastroenterología*. Ed. by Eduardo Pérez TORRES et al. New York, NY: McGraw-Hill Education, 2015 (cit. on pp. 11, 14).

HASTIE, Trevor et al. “Overview of supervised learning”. In: *The elements of statistical learning: Data mining, inference, and prediction* (2009), pp. 9–41 (cit. on pp. 16, 17).

HUANG, Chien-Hsiang; WU, Hung-Yu, and LIN, Youn-Long. “Hardnet-mseg: A simple encoder-decoder polyp segmentation neural network that achieves over 0.9 mean dice and 86 fps”. In: *arXiv preprint arXiv:2101.07172* (2021) (cit. on p. 26).

HWANG, Sae et al. “Polyp detection in colonoscopy video using elliptical shape feature”. In: *2007 IEEE International Conference on Image Processing*. Vol. 2. IEEE. 2007, pp. II–465 (cit. on p. 24).

JHA, Debesh et al. “A comprehensive study on colorectal polyp segmentation with ResUNet++, conditional random field and test-time augmentation”. In: *IEEE journal of biomedical and health informatics* 25.6 (2021), pp. 2029–2040 (cit. on p. 25).

JHA, Debesh et al. “Real-time polyp detection, localization and segmentation in colonoscopy using deep learning”. In: *IEEE Access* 9 (2021), pp. 40496–40510 (cit. on p. 27).

JING, Longlong and TIAN, Yingli. “Self-supervised visual feature learning with deep neural networks: A survey”. In: *IEEE transactions on pattern analysis and machine intelligence* 43.11 (2020), pp. 4037–4058 (cit. on p. 20).

LECUN, Yann et al. “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324 (cit. on p. 17).

LEE, Chang Kyun; CHA, Jae Myung, and KIM, Wan Jung. “Endoscopist fatigue may contribute to a decline in the effectiveness of screening colonoscopy”. In: *Journal of clinical gastroenterology* 49.6 (2015), e51–e56 (cit. on p. 12).

LEVIN, Bernard et al. “Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology”. In: *Gastroenterology* 134.5 (2008), pp. 1570–1595 (cit. on pp. 11, 14).

LI, Qiaoliang et al. “Colorectal polyp segmentation using a fully convolutional neural network”. In: *2017 10th international congress on image and signal processing, biomedical engineering and informatics (CISP-BMEI)*. IEEE. 2017, pp. 1–5 (cit. on p. 25).

NASTESKI, Vladimir. “An overview of the supervised machine learning methods”. In: *Horizons*. b 4 (2017), pp. 51–62 (cit. on p. 16).

NOFFSINGER, Amy E. “Serrated polyps and colorectal cancer: new pathway to malignancy”. In: *Annual Review of Pathology: Mechanisms of Disease* 4 (2009), pp. 343–364 (cit. on pp. 11, 14).

ORGANIZACIÓN MUNDIAL DE LA SALUD (OMS). *Cancer Today*. <https://gco.iarc.fr/today/home>. International Agency for Research on Cancer, 2020 (cit. on pp. 11, 14).

POGORELOV, Konstantin et al. “Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection”. In: *Proceedings of the 8th ACM on Multimedia Systems Conference*. 2017, pp. 164–169 (cit. on p. 38).

REX, Douglas K et al. “Quality in the technical performance of colonoscopy and the continuous quality improvement process for colonoscopy: recommendations of the US Multi-Society Task Force on Colorectal Cancer”. In: *The American journal of gastroenterology* 97.6 (2002), pp. 1296–1308 (cit. on p. 15).

RONNEBERGER, Olaf; FISCHER, Philipp, and BROX, Thomas. “U-net: Convolutional networks for biomedical image segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* 18. Springer. 2015, pp. 234–241 (cit. on p. 27).

RUIZ, Lina and MARTINEZ, Fabio. “Weakly supervised polyp segmentation from an attention receptive field mechanism”. In: *2022 44th Annual International Conference of the IEEE Engineering in Medicine And Biology Society (EMBC)*. IEEE. 2022, pp. 3745–3758 (cit. on pp. 12, 26, 27, 31).

RUIZ, Lina et al. “COLON: The largest COLonoscopy LONG sequence public database”. In: *arXiv preprint arXiv:2403.00663*. 2024 (cit. on p. 13).

RUIZ, Lina et al. “COLON: The largest COlonoscopy LONG sequence public database”. In: *arXiv preprint arXiv:2403.00663*. 2024 (cit. on p. 39).

SCHMARJE, Lars et al. “A survey on semi-, self-and unsupervised learning for image classification”. In: *IEEE Access* 9 (2021), pp. 82146–82168 (cit. on p. 19).

SILVA, Juan et al. “Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer”. In: *International journal of computer assisted radiology and surgery* 9 (2014), pp. 283–293 (cit. on p. 37).

SOHN, Kihyuk et al. “Fixmatch: Simplifying semi-supervised learning with consistency and confidence”. In: *Advances in neural information processing systems* 33 (2020), pp. 596–608 (cit. on p. 19).

SUN, Xinzi et al. “Colorectal polyp segmentation by u-net with dilation convolution”. In: *2019 18th IEEE international conference on machine learning and applications (ICMLA)*. IEEE. 2019, pp. 851–858 (cit. on p. 25).

TAJBAKSH, Nima; GURUDU, Suryakanth R, and LIANG, Jianming. “Automated polyp detection in colonoscopy videos using shape and context information”. In: *IEEE transactions on medical imaging* 35.2 (2015), pp. 630–644 (cit. on p. 37).

TAPIA, Oscar et al. “Cáncer de colon y recto: Descripción morfológica y clínica de 322 casos”. In: *International Journal of Morphology* 28.2 (2010), pp. 393–398 (cit. on pp. 11, 14).

TIAN, Yu et al. “Few-shot anomaly detection for polyp frames from colonoscopy”. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VI* 23. Springer. 2020, pp. 274–284 (cit. on p. 27).

VAN ENGELEN, Jesper E and HOOS, Holger H. “A survey on semi-supervised learning”. In: *Machine learning* 109.2 (2020), pp. 373–440 (cit. on pp. 18, 19).

VAN RIJN, Jeroen C et al. “Polyp miss rate determined by tandem colonoscopy: a systematic review”. In: *Official journal of the American College of Gastroenterology/ ACG* 101.2 (2006), pp. 343–350 (cit. on p. 15).

VLEUGELS, Jasper LA; HAZEWINKEL, Yark, and DEKKER, Evelien. “Morphological classifications of gastrointestinal lesions”. In: *Best Practice & Research Clinical Gastroenterology* 31.4 (2017), pp. 359–367 (cit. on p. 14).

ZHANG, Dingwen et al. “Weakly supervised object localization and detection: A survey”. In: *IEEE transactions on pattern analysis and machine intelligence* 44.9 (2021), pp. 5866–5885 (cit. on p. 18).

ZHOU, Zhi-Hua. “A brief introduction to weakly supervised learning”. In: *National science review* 5.1 (2018), pp. 44–53 (cit. on pp. 17, 18).

ZHOU, Zhi-Hua and ZHOU, Zhi-Hua. “Semi-supervised learning”. In: *Machine Learning* (2021), pp. 315–341 (cit. on p. 19).

ZHU, Xiaojin Jerry. “Semi-supervised learning literature survey”. In: (2005) (cit. on p. 19).