



Detección de pólipos gastrointestinales utilizando un sistema de diagnóstico asistido por computador (DAC) – Fase I

Natalia López Grisales

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Asesores

María Bernarda Salazar Sánchez, Doctora (PhD) en Ingeniería Electrónica

Elkin Manuel Romero Camargo, Médico especialista en Gastroenterología y Hepatología

Universidad de Antioquia
Facultad de Ingeniería
Especialización en Analítica y Ciencia de Datos
Medellín, Antioquia, Colombia
2024

Cita	(López Grisales, 2024)
Referencia	López Grisales, N. (2024). <i>Desarrollo de un sistema de Diagnóstico Asistido por Computador (DAC) para la detección de pólipos gastrointestinales – Fase I</i> [Trabajo de grado especialización]. Universidad de Antioquia, Medellín, Colombia.
Estilo APA 7 (2020)	



Especialización en Analítica y Ciencia de Datos, Cohorte VI.

Grupo de Investigación Intelligent Information Systems Lab – In2Lab.

Centro de Investigación Ambientales y de Ingeniería (CIA).



Centro de Documentación Ingeniería (CENDOI)

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Rector: John Jairo Arboleda Céspedes.

Decano: Julio Cesar Saldarriaga Molina

Jefe departamento: Danny Alexandro Múnica Ramírez

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Dedicatoria

A...

Agradecimientos

Tabla de contenido

1.	Introducción	7
2.	Metodos y materiales	7
2.1.	Descripción delconjuntos de datos.....	7
2.2.	Procesamiento de los datos	8
2.3.	Entrenamiento de modelos.....	8
2.3.1.	Red U-Net.....	9
2.3.2.	Autoencoder Variacional (VAE).....	9
3.	Resultados y Discusión	10
4.	Conclusiones	12
	Referencias.....	12

Lista de Figuras

<i>Figura 1. Muestra original de los datos contenidos en las tres bases de datos. Columna izquierda: Imágenes RGB de cada dataset. Columna derecha: Las máscaras de segmentación (ground truth).....</i>	<i>8</i>
<i>Figura 2. Conteo de pólipos por posición dentro de la imagen (se divide cada imagen en 9 celdas de 3x3) para cada dataset.....</i>	<i>8</i>
<i>Figura 3. Arquitectura de la red neuronal U-Net.....</i>	<i>9</i>
<i>Figura 4. Arquitectura de Autoencoder Variacional - VAE.....</i>	<i>9</i>
<i>Figura 5. Historial de métricas de entrenamiento de la red U-Net.....</i>	<i>10</i>
<i>Figura 6. Historial de métricas de entrenamiento y validación del modelo Autoencoder Variacional - VAE.</i>	<i>10</i>
<i>Figura 7. Matriz de confusión para (Superior) el modelo U-Net y (Inferior) el modelo VAE.</i>	<i>11</i>
<i>Figura 8. Predicciones realizadas por la U- Net.....</i>	<i>12</i>
<i>Figura 9. Predicciones por el VAE.....</i>	<i>12</i>

Lista de Tablas

<i>Tabla 1. Reporte de clasificación para la red U-Net.....</i>	<i>11</i>
<i>Tabla 2. Reporte de clasificación para el VAE.....</i>	<i>11</i>
<i>Tabla 3. Media y desviación estándar de la intersección sobre la unión (IoU) y el coeficiente de Dice para la segmentación de la red U-Net y VAE.</i>	<i>12</i>

Detección de pólipos gastrointestinales utilizando un sistema de diagnóstico asistido por computador (DAC) – Fase I

Resumen— El cáncer colorrectal es un desafío significativo para la salud global, ocupando el tercer lugar tanto en incidencia como en tasas de mortalidad. Los métodos de detección temprana, como la colonoscopia, han mostrado ser prometedores en la reducción de la incidencia y mortalidad del de este tipo de cáncer. Sin embargo, la efectividad de la colonoscopia se ve obstaculizada por la variabilidad y la posibilidad de pólipos no detectados. Para abordar esto, este estudio propone un sistema de diagnóstico asistido por computadora para la identificación de pólipos colorrectales utilizando dos técnicas de aprendizaje profundo entrenadas en tres conjuntos de datos: una red neuronal U-Net y una estructura autoencoder variacional. El mejor modelo fue la red U-Net, la cual alcanzó una precisión del 94% para la detección temprana de pólipos, resultados superiores comparados con el autoencoder variacional en la segmentación de imágenes.

Keywords—Colorectal cancer, colonoscopy, computer-aided diagnosis, deep learning, polyp detection.

1. INTRODUCCIÓN

El cáncer colorrectal (CCR) representa un desafío de salud significativo a nivel mundial, siendo el tercero más diagnosticado y la tercera causa principal de muerte relacionada con el cáncer en los Estados Unidos (Siegel Mph et al., 2023). En Colombia, en el 2020, fue el tercero en incidencia y mortalidad en ambos sexos de los casos reportados para el mismo año (9.4% - 10.0%). Según reportes del *Instituto Nacional de Cancerología*, de 4471 nuevos casos reportados el 8,3% corresponden a cáncer colorrectal, y es el segundo más frecuente de los tumores malignos del tracto gastrointestinal para los hombres después del cáncer gástrico (Márquez-Ustariz A et al., 2023). Según la Organización Mundial de la Salud (OMS), los cambios en el estilo de vida y las pruebas de detección periódicas pueden ayudar a prevenir el cáncer colorrectal Click or tap here to enter text.. Cuyos principales métodos de diagnóstico son el examen físico, el examen del interior del colon mediante colonoscopia o sigmoidoscopia, e imágenes de ecografía abdominal, tomografía computarizada y resonancia magnética (World Health Organization, 2023). La realización de una colonoscopia se asocia con una reducción del 20% en la incidencia de cáncer colorrectal y una disminución del 26% en la mortalidad por el mismo en comparación con la atención estándar basada en pruebas inmunoquímicas fecales (Han et al., 2024).

La colonoscopia es un examen por imágenes, mínimamente invasivo, donde un profesional de la salud analiza las paredes del intestino grueso mediante imágenes transmitidas por una pequeña cámara que se encuentra en el extremo de un tubo flexible llamado colonoscopio o endoscopio (The American Cancer Society, 2023). Sin embargo, la detección de cáncer con este método tiene tasas de error variables, en los rangos del 17 % al 28 % para pólipos de cualquier tamaño, del 2,1 %

al 12 % para pólipos grandes y del 5 % al 11 % para adenomas avanzados. Además, las tasas de detección de adenomas y lesiones serradas sésiles tienen una mayor dificultad porque reporten mayor variabilidad entre colonoscopistas (Park & Cha, 2022).

La experiencia y habilidades del profesional de la salud son imprescindibles para el diagnóstico, sin embargo, hay casos de algunos pólipos que pueden pasar desapercibidos durante la colonoscopia por el tamaño y morfología de los pólipos, el entorno y mucosidades dentro del intestino, la duración del procedimiento y el estado de agotamiento del profesional (Ahmad et al., 2019). Estas dificultades hacen relevante el uso de un sistema de diagnóstico asistido por computador (DAC) para apoyar el proceso de detección de los y disminuir errores en la caracterización de los pólipos (Tomar, 2021) (Sánchez-Montes et al., 2020).

Recientemente se han propuesto diferentes sistemas DAC para las colonoscopias usando procesamiento digital de imágenes, algoritmos de aprendizaje automático y aprendizaje profundo (Tharwat et al., 2022). Si bien existe dificultad en acceder a datos de calidad en esta área, los últimos años la capacidad de predicción de los modelos ha aumentado gracias al acceso a grandes volúmenes de datos disponibles de manera abierta (Hsu et al., 2021).

Por lo tanto, en este trabajo se propone un sistema DAC para la identificación de pólipos colorrectales mediante técnicas de aprendizaje profundo (deep learning), entrenadas utilizando tres (3) bases de datos de acceso libre: CVC-ColonDB (Bernal et al., 2012), CVC-ClinicDB (Bernal et al., 2015) y Kvasir-SEG (Jha et al., 2020). El mejor modelo encontrado alcanzó una precisión del 94% en la identificación de pólipos colorrectales en colonoscopias.

2. METODOS Y MATERIALES

El desarrollo del sistema DAC propuesto en este trabajo se dividió en dos etapas: La primera consistió en un análisis exploratorio, preprocesamiento de las imágenes y la posterior preparación de los datos de entrenamiento, validación y prueba; y una segunda etapa, en la cual se realizó el entrenamiento de dos modelos reportados en la literatura para la clasificación y segmentación de pólipos gastrointestinales: una red U-Net (Ronneberger et al., 2015) y un Autoencoder variacional (VAE) (Kingma & Welling, 2013).

2.1. DESCRIPCIÓN DEL CONJUNTO DE DATOS

La información utilizada en este desarrollo corresponde a tres conjuntos de datos de libre acceso que contienen imágenes de pólipos gastrointestinales y sus correspondientes máscaras de segmentación (ver **Figura 1**). El número de imágenes total en cada base de datos es 380, 612 y 1000 para CVC-ColonDB, CVC-ClinicDB y Kvasir-SEG, respectivamente. Las tres

fuentes de información tienen imágenes con dimensiones diferentes lo cual requiere de un proceso de estandarización previo al entrenamiento de los modelos.

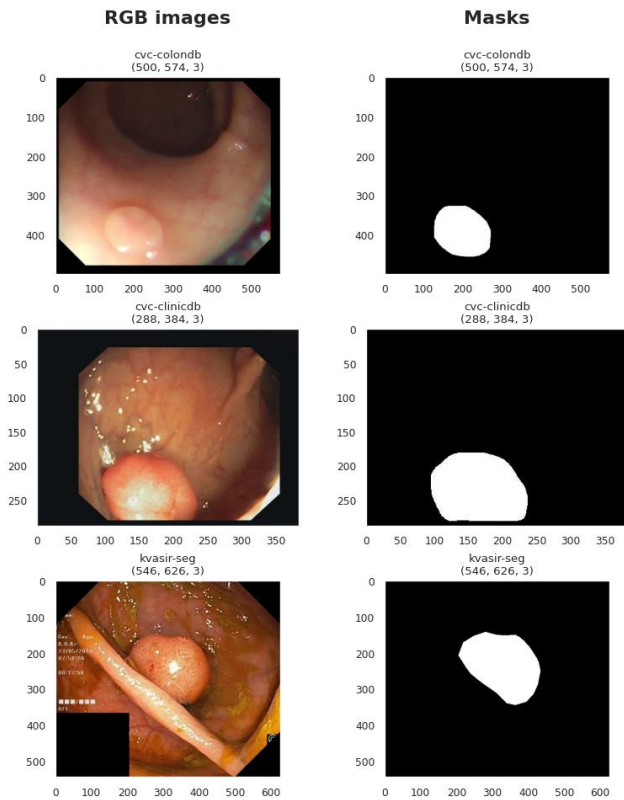


Figura 1. Muestra original de los datos contenidos en las tres bases de datos. Columna izquierda: Imágenes RGB de cada dataset. Columna derecha: Las máscaras de segmentación (ground truth).

La base de datos CVC-ColonDB (Bernal et al., 2012) corresponde a muestras generadas a partir del análisis 15 casos aleatorios, en los que expertos gastroenterólogos anotaron todas las secuencias que mostraban pólipos. Cada secuencia contenía 20 fotogramas con un tamaño de 500×574 píxeles. La parte central de las imágenes se selecciona eliminando los bordes negros no funcionales. Los expertos validaron posteriormente que los 20 fotogramas evidenciaron información diferente dentro de la secuencia, eliminando así fotogramas similares.

La segunda base de datos utilizada es CVC-ClinicDB (Bernal et al., 2015) publicada por el Hospital Clínic de Barcelona, España. Se generó a partir de 23 estudios de vídeo diferentes de intervenciones de colonoscopia estándar con luz blanca. Para cada caso se extrajeron todas las secuencias que contenían un pólipo, proporcionando al final un número total de 31 secuencias diferentes. Los expertos crearon un *ground truth* definido manualmente como la máscara en la región cubierta por el pólipo.

Por último, el tercer conjunto de datos empleado en este proyecto corresponde a KVASIR-SEG (Jha et al., 2020), el cual contiene el mayor número de imágenes en comparación con las otras dos bases de datos. Tiene 1000 imágenes de pólipos gastrointestinales y sus correspondientes máscaras de segmentación, anotadas manualmente por un médico y verificadas posteriormente por un gastroenterólogo

experimentado. La resolución de las imágenes contenidas en Kvasir-SEG varía entre 332x487 y 1920x1072 píxeles (Jha et al., 2020).

2.2. PROCESAMIENTO DE LOS DATOS

Durante la exploración de datos se utilizaron las bibliotecas NumPy, OpenCV y Matplotlib de Python. En esta etapa, se verifica el formato y la cantidad de archivos, se renombraron y convierten los archivos a un formato estándar, se generaron cajas delimitadoras para los pólipos y se exploraron las características básicas de las imágenes, como el ancho, el alto, la posición del pólipo, el tamaño y el recuento de pólipos, y el matiz, el brillo y la saturación de las imágenes. Todo el desarrollo se ha alojado en el siguiente repositorio de GitHub: <https://github.com/natalialopezg/Monografia-EACD/tree/master>.

Se utilizaron librerías como *os* y *OpenCV* para cargar y verificar imágenes y máscaras, asegurando su coherencia en cantidad y formato. Todas las imágenes se convirtieron de BGR a RGB, se unificaron al mismo formato (.jpg) y se normalizaron escalando los números de píxeles por 255. Así mismo, las máscaras se redimensionaron a 256x256 píxeles para reducir la dimensionalidad de los datos y mejorar el rendimiento del entrenamiento. Antes de aplicar el redimensionamiento, se llevó a cabo un recorte previo, llevando cada imagen y su correspondiente máscara a una relación de aspecto 1:1, ya que como se evidencia en la **Figura 2**, la mayoría de los pólipos se encuentran en el centro de la imagen, por lo cual, no se verían afectados por el recorte en los extremos de la imagen sobre el eje de mayor dimensión disminuyendo las distorsiones durante el redimensionamiento.

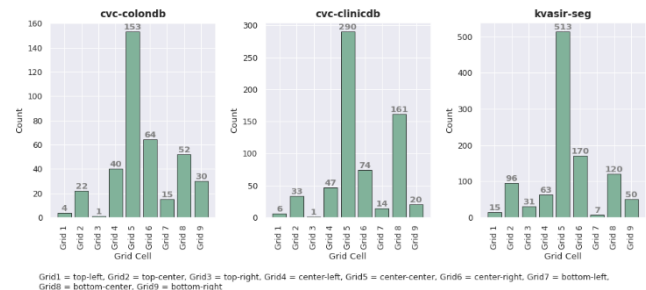


Figura 2. Conteo de pólipos por posición dentro de la imagen (se divide cada imagen en 9 celdas de 3x3) para cada dataset.

2.3. ENTRENAMIENTO DE MODELOS

En esta etapa, se dividieron los datos en conjuntos de entrenamiento, prueba y validación, utilizando una proporción de 80%, 10% y 10%, respectivamente. Este proceso asegura una preparación adecuada de los datos para su uso en el desarrollo y evaluación de modelos de segmentación de imágenes.

En la segunda etapa del proceso de desarrollo del sistema DAC se entrenaron dos modelos de aprendizaje profundo: una red U-Net y un Autoencoder variacional. Ambos modelos son entrenados durante 100 épocas, con una tasa de aprendizaje de 0.0001 y un tamaño de lote de 16. El resultado de predicción se evalúa empleando una matriz de confusión y un informe de clasificación, proporcionando métricas detalladas como precisión, recall, F1-score, IoU (Intersección sobre Unión) y el coeficiente de Dice. Estas dos últimas

métricas permiten evaluar el rendimiento del modelo en la segmentación de imágenes.

La **intersección sobre la unión (IoU)**, también conocida como índice de Jaccard, es el área de la intersección sobre la unión de la segmentación predicha y el área verdadera (Ver ecuación (1))

$$IoU = \frac{TP}{TP + FP + FN} \quad (1)$$

Dónde,

TP (True Positives - Verdaderos Positivos), es el número de píxeles que fueron correctamente clasificados como parte de la región de interés (pólipo). **FP** (False Positives - Falsos Positivos), corresponde al número de píxeles que fueron incorrectamente clasificados como parte de la región de interés. **FN** (False Negatives - Falsos Negativos), son el número de píxeles que pertenecen a la región de interés pero que fueron incorrectamente clasificados como parte del fondo.

Coefficiente de Dice = F1-score, es la media armónica de precisión y recall. En otras palabras, el coeficiente Dice es $2 * \frac{\text{área de solapamiento}}{\text{número total de píxeles de ambas imágenes}}$ (Ver ecuación (2)).

$$Dice = \frac{2 * TP}{2 * TP + FP + FN} \quad (2)$$

2.3.1. RED U-NET

La arquitectura U-Net es una red neuronal convolucional (CNN) diseñada específicamente para la segmentación rápida y precisa de imágenes biomédicas, sin embargo, gracias a su eficacia y flexibilidad en los últimos años se ha utilizado en una amplia gama de tareas de segmentación de imágenes más allá de las aplicaciones biomédicas (Azad et al., 2022). La red U-Net se compone de un *encoder* y un *decoder* (ver **Figura 3**), con bloques de convolución en ambos lados para capturar características a diferentes escalas y permitir una segmentación precisa de la imagen.

Para el aumento de datos durante el entrenamiento, se incluye diversas transformaciones como *flip horizontal* y *vertical*, rotación, zoom, cizallamiento, y desplazamiento en anchura y altura. Además, se ajustaron los rangos de brillo y se utilizó el modo de llenado '*reflect*' para completar los píxeles en las transformaciones. De este modo, se crean variaciones de las imágenes originales que, en última instancia, mejoran la generalización y robustez del modelo de aprendizaje profundo al exponerlo a una gama más amplia de variaciones de imagen. Estos generadores están preparados para alimentar modelos de segmentación de imágenes durante el entrenamiento o la evaluación.

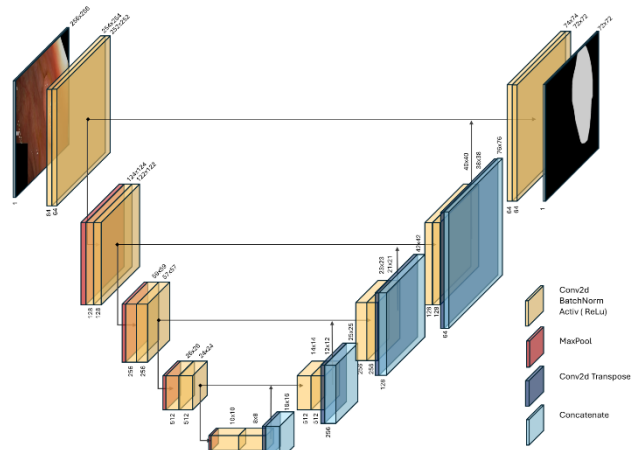


Figura 3. Arquitectura de la red neuronal U-Net.

El *encoder* se encarga de reducir progresivamente la resolución espacial de la entrada, extrayendo características relevantes mediante la aplicación de bloques de convolución seguidos de capas de max-pooling para reducir la dimensionalidad. En cada bloque de convolución se aplican dos capas convolucionales, cada una seguida de normalización por lotes y activación ReLU. Esto ayuda a capturar características complejas de la imagen a diferentes niveles de abstracción. Además, se implementa la regularización L2 para mitigar el sobreajuste durante el entrenamiento.

Por otro lado, el *decoder* se encarga de reconstruir la información espacial a partir de las características extraídas por el *encoder*, utilizando bloques de convolución transpuesta para aumentar progresivamente la resolución espacial. Cada bloque del *decoder* también incorpora conexiones de salto (skip connections) desde los bloques correspondientes del *encoder* para proporcionar información detallada a diferentes escalas. Se utilizó la función de pérdida de entropía cruzada binaria y el optimizador Adam de Keras.

2.3.2. AUTOENCODER VARIACIONAL (VAE)

El entrenamiento de un VAE implica la construcción de dos componentes principales: el *encoder* y el *decoder*. El *encoder* toma la entrada de datos y la transforma en una representación latente, mientras que el *decoder* toma esta representación latente y la reconstruye de nuevo en los datos originales (ver **Figura 4**).

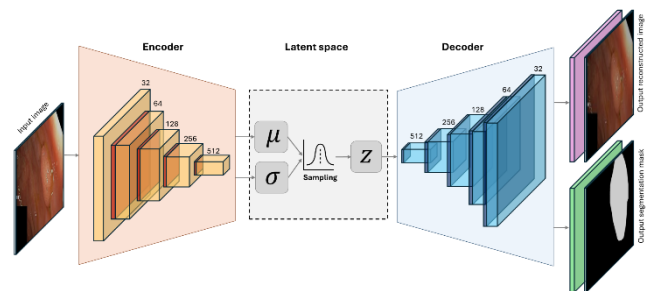


Figura 4. Arquitectura de Autoencoder Variacional - VAE

La arquitectura del *encoder* comienza con capas convolucionales que reducen progresivamente la dimensionalidad de la entrada. Cada una está seguida de una capa de *batch normalization* y activación ReLU para capturar características complejas de la imagen. La última capa convolucional produce dos tensores, z_mean y z_log_var , que representan la media y la varianza del espacio latente. Estos tensores se utilizan para muestrear un punto en el espacio latente utilizando la capa *Sampling*.

El *decoder*, por otro lado, comienza con una capa densa que convierte el punto de la representación latente en un tensor de la forma adecuada para la convolución transpuesta, esto busca aumentar progresivamente la dimensionalidad del tensor hasta que se produce la reconstrucción final de los datos originales. Se incorporan capas de regularización y normalización para mejorar la generalización y estabilidad del modelo.

Durante el entrenamiento, se utiliza una combinación de dos funciones de pérdida: la pérdida de reconstrucción y la divergencia de Kullback-Leibler (T. Wu et al., 2024). La primera mide la discrepancia entre los datos originales y los datos reconstruidos por el *decoder*; y la segunda mide la diferencia entre la distribución latente del modelo y una distribución normal estándar, fomentando que la distribución latente se aproxime a una distribución gaussiana. Tras el entrenamiento del modelo, se realiza un post-procesamiento a las predicciones de las máscaras de segmentación mediante la aplicación de un umbral y una operación de cierre morfológico.

3. RESULTADOS Y DISCUSIÓN

En esta sección se presentan los resultados obtenidos de la evaluación los modelos de segmentación de imágenes: la U-Net y el VAE.

En la **Figura 5** se observa una capacidad de generalización entre 88% - 90% para el modelo U-Net, con una disminución general en la pérdida de validación mínima de 0.13 en la época 100. La precisión con los datos de validación es consistentemente mayor que la de entrenamiento, superando el 94%, lo que sugiere que no hay tendencia de sobreajuste para las distintas épocas analizadas.



Figura 5. Historial de métricas de entrenamiento de la red U-Net

La tendencia en la pérdida global obtenida en la red U-Net es similar a la obtenida en el modelo VAE (ver **Figura 6**), sin embargo, hay una notable diferencia, un desfase estable, en los valores de pérdida reportados tanto para la fase de entrenamiento como la de validación a lo largo de las épocas. La tendencia de menor pérdida global indica una buena capacidad de generalización del modelo (L. Wu et al., 2017). Así mismo, la pérdida de reconstrucción, que mide la capacidad del VAE para reconstruir los datos de entrada, también muestra una disminución constante en ambas líneas, lo que sugiere que las reconstrucciones son cada vez más precisas. Por otro lado, la pérdida de segmentación, que mide la capacidad del modelo para construir la máscara, aunque inicia con una tendencia a la baja en ambas líneas, esta tendencia solo se mantiene durante el entrenamiento, en la validación no logra estabilizarse.



Figura 6. Historial de métricas de entrenamiento y validación del modelo Autoencoder Variacional - VAE.

En la **Figura 7**, se busca analizar por medio de la matriz de confusión la capacidad del modelo para segmentar el objeto de interés del fondo, tratándose como un problema de clasificación binaria donde cada pixel de la imagen corresponde a una de las dos clases siguiente: clase positiva (pólipo) y clase negativa (fondo).

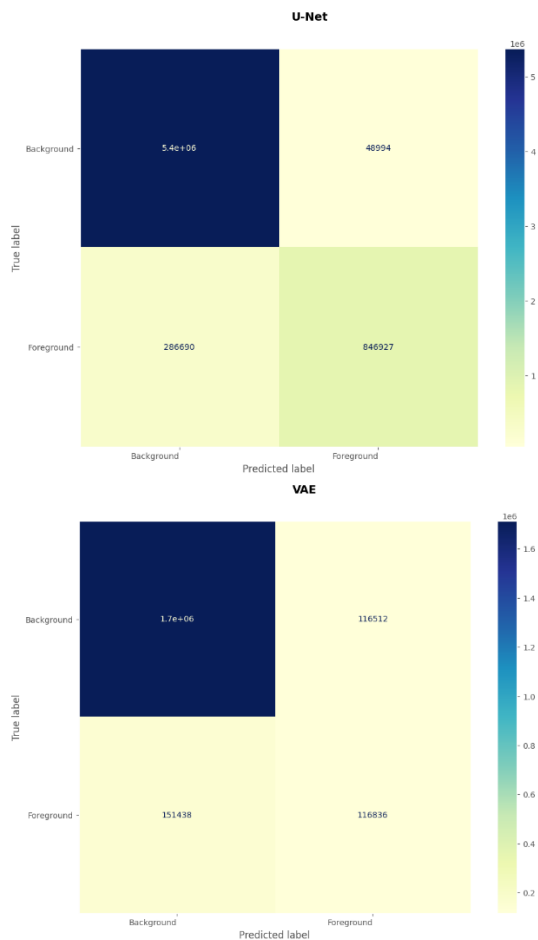


Figura 7. Matriz de confusión para (Superior) el modelo U-Net y (Inferior) el modelo VAE.

Los resultados obtenidos muestran que los modelos tienen una capacidad diferente para distinguir entre el objeto de interés (clase positiva), identificado como *Foreground* en el reporte de clasificación (Ver Tabla 2), y el fondo en las imágenes (clase negativa), identificado como *Background* en el reporte de clasificación. El modelo U-Net tiene altas tasas de verdaderos positivos y verdaderos negativos, es decir que logra distinguir los pixeles correspondientes a cada clase. Sin embargo, el modelo VAE logra una alta tasa de verdaderos negativos, mientras que los verdaderos positivos y falsos positivos alcanzan desempeños muy similares, lo que indica una capacidad limitada del modelo para distinguir con precisión entre el objeto de interés y el fondo en las imágenes.

Estos hallazgos destacan la necesidad de mejorar el proceso de segmentación de las imágenes, para aumentar la calidad de los datos utilizados en el proceso de entrenamiento de la red. Además, es fundamental para tareas de segmentación considerar los desequilibrios entre el número de muestras por clase lo cual pueden afectar significativamente la capacidad del modelo para distinguir cada una (Wang et al., 2022).

En la Tabla 1, el modelo U-Net muestra un rendimiento sólido en la clasificación del fondo, con altas puntuaciones en todas las métricas. Sin embargo, es más conservador al predecir la clase de primer plano, lo que se refleja en una alta

precisión, pero una baja recall, lo cual implica que se debe mejorar en la recuperación de la clase de primer plano. De forma similar, el modelo VAE (ver Tabla 2) presenta una alta precisión en la clasificación del fondo, pero una baja recuperación de la clase de primer plano, lo que se refleja en un bajo puntaje F1. En resumen, el modelo presenta un buen desempeño para identificar el fondo, pero tiene dificultades para detectar objetos de interés, lo que indica la necesidad de mejoras en su capacidad de clasificación. La columna support en el reporte de clasificación representa el número de muestras (o instancias) que pertenecen a cada clase en el conjunto de datos utilizado para la evaluación; en este caso el número de pixeles. Para la red U-Net, el conjunto de datos tiene un total de 6,553,600 muestras, de las cuales 5,419,983 pertenecen a la clase "Background" y 1,133,617 pertenecen a la clase "Foreground". Para el VAE, el total de muestras es de 2,097,152, de las cuales 1,828,878 pertenecen a la clase "Background" y 268,274 pertenecen a la clase "Foreground". La cantidad de muestras es diferente para ambos modelos debido al aumento de datos, al aplicar transformaciones aleatorias a las imágenes, se generan nuevas muestras de datos que el modelo puede utilizar para aprender patrones y características adicionales.

Tabla 1. Reporte de clasificación para la red U-Net.

	precision	recall	f1-score	support
Background (positive class)	0.95	0.99	0.97	5419983
Foreground (negative class)	0.95	0.75	0.83	1133617
accuracy			0.95	6553600
macro avg	0.95	0.87	0.90	6553600
weighted avg	0.95	0.95	0.95	6553600

Tabla 2. Reporte de clasificación para el VAE.

	precision	recall	f1-score	support
Background (positive class)	0.92	0.94	0.96	1828878
Foreground (negative class)	0.50	0.44	0.47	268274
accuracy			0.87	2097152
macro avg	0.71	0.69	0.70	2097152
weighted avg	0.87	0.87	0.87	2097152

El modelo U-Net alcanza un mayor solapamiento (IoU = 0,7166) entre la segmentación real y la reconstruida (ver **Tabla 3**). Además de un mayor coeficiente Dice (0,8712), lo cual se relaciona con un alto grado de solapamiento entre los segmentos predichos y los reales (ver **Figura 8**). Los coeficientes Dice cercanos a 1 indican un rendimiento muy alto del modelo (Eelbode et al., 2020). Los valores obtenidos para el modelo VAE sugieren baja capacidad de capturar la superposición entre las áreas predichas y reales en las imágenes segmentadas.

Tabla 3. Media y desviación estándar de la intersección sobre la unión (IoU) y el coeficiente de Dice para la segmentación de la red U-Net y VAE.

	U-Net	VAE
IoU	0.7167 ± 0.0005	0.2443 ± 0.0507
Dice Coefficient	0.8712 ± 0.00007	0.3904 ± 0.0670

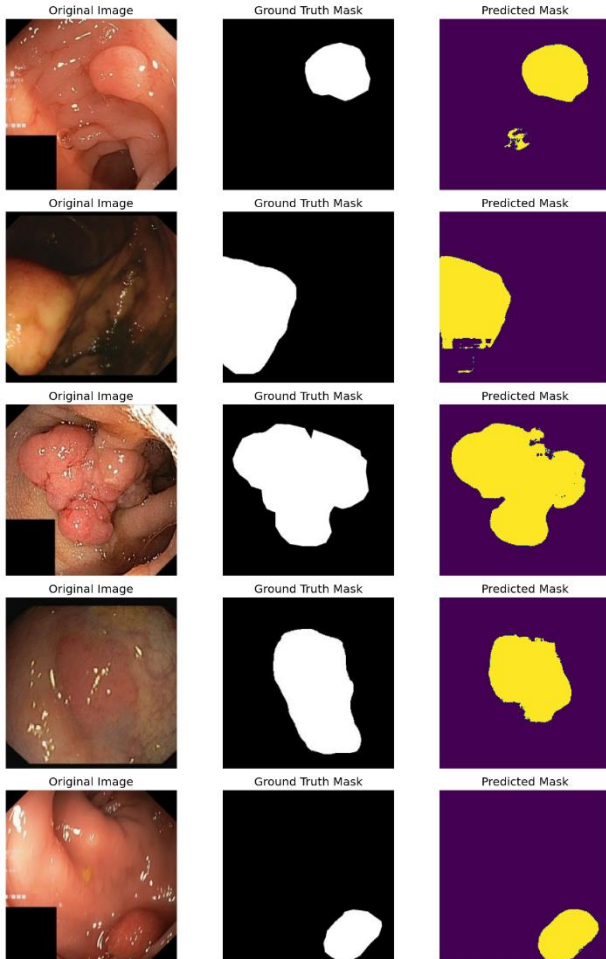


Figura 8. Predicciones realizadas por la U-Net

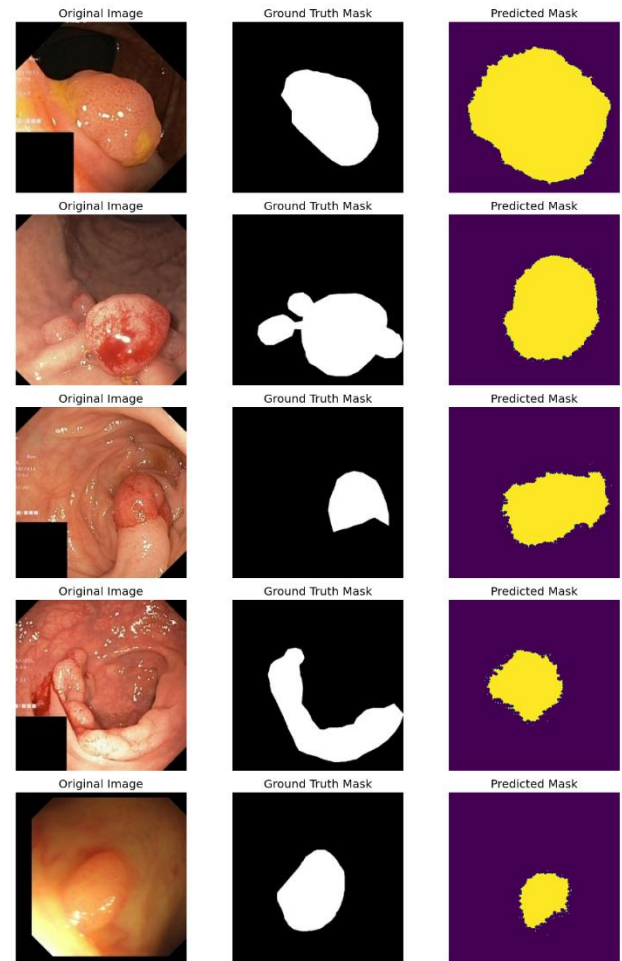


Figura 9. Predicciones por el VAE

4. CONCLUSIONES

La U-Net demostró un sólido rendimiento en la segmentación de imágenes, con una precisión y exactitud general del 95%, mientras que el VAE tuvo un rendimiento notablemente inferior. La alta precisión de la U-Net para el fondo y una buena, aunque menor, precisión para el objeto de interés indican que es efectiva en la identificación del fondo y razonablemente buena en la identificación del objeto de interés. Por otro lado, el VAE mostró una alta precisión para el fondo (0.92), pero su precisión y recall para el objeto de interés fueron bajos, lo que resultó en una puntuación F1 baja (0.47) para el objeto de interés. Las métricas de IoU y el coeficiente de Dice reflejan una segmentación menos efectiva para el VAE en comparación con la U-Net. En conclusión, la U-Net es una opción más robusta y efectiva para tareas de segmentación de imágenes en comparación con el VAE, ya que tiene una excelente capacidad para diferenciar entre fondo y objeto de interés. En la Figura 8 se presentan las predicciones realizadas por el modelo U-Net para algunas imágenes aleatorias y en la Figura 9 las predicciones realizadas por el modelo VAE.

REFERENCIAS

Ahmad, O. F., Soares, A. S., Mazomenos, E., Brandao, P., Vega, R., Seward, E., . . . Lovat, L. B. (2019).

Artificial intelligence and computer-aided diagnosis in colonoscopy: current evidence and future directions. *The lancet. Gastroenterology & hepatology*, 4(1), 71-80. doi:10.1016/s2468-1253(18)30282-6

American Cancer Society. (2023, 10 3). *Colonoscopy*. Retrieved from Cancer.org: <https://www.cancer.org/cancer/diagnosis-staging/tests/endoscopy/colonoscopy.html>

Retrieved from <https://acsjournals.onlinelibrary.wiley.com/doi/10.3322/caac.21772>

Conteduca, V., Sansonno, D., Russi, S., & Dammacco, F. (2013). Precancerous colorectal lesions. *International journal of oncology*, 43(4), 973-984. doi:10.3892/ijo.2013.2041

Doubeni, C. A., Corley, D. A., Quinn, V. P., Jensen, C. D., Zauber, A. G., Goodman, M., . . . Fle. (2018). Effectiveness of screening colonoscopy in reducing the risk of death from right and left colon cancer: a large community-based study. *Gut*, 67(2), 291-298. doi:10.1136/gutjnl-2016-312712

Hsu, C.-M., Hsu, C.-C., Hsu, Z.-M., Shih, F.-Y., Chang, M.-L., & Chen, T.-H. (2021). Colorectal polyp image detection and classification through grayscale images and deep learning. *Sensors (Basel, Switzerland)*, 21(1/), 5995. doi:10.3390/s21185995

Nishihara, R., Wu, K., Lochhead, P., Morikawa, T., Liao, X., Qian, Z. R., . . . Ogino, S. (2013). Long-term colorectal-cancer incidence and mortality after lower endoscopy. *The New England journal of medicine*, 369(12), 1095-1105. doi:10.1056/nejmoa1301969

Park, S. B., & Cha, J. M. (2022). Quality indicators in colonoscopy: the chasm between ideal and reality. *Clinical endoscopy*, 55(3), 332-338. doi:10.5946/ce.2022.037

Sánchez-Montes, C., Bernal, J., García-Rodríguez, A., Córdova, H., & Fernández-Esparrach, G. (2020). Revisión de métodos computacionales de detección y clasificación de pólipos en imagen de colonoscopia. *Gastroenterología y hepatología*, 43(4), 222-232. doi:10.1016/j.gastrohep.2019.11.004

Sánchez-Montes, C., García-Rodríguez, A., Córdova, H., Pellisé, M., & Fernández-Esparrach, G. (2020). Tecnologías de endoscopia avanzada para mejorar la detección y caracterización de los pólipos colorrectales. *Gastroenterología y hepatología*, 43(1), 46-56. doi:10.1016/j.gastrohep.2019.09.008

Shine, R., Bui, A., & Burgess, A. (2020). Quality indicators in colonoscopy: an evolving paradigm: Quality indicators in colonoscopy. *ANZ journal of surgery*, 90(3), 215-221. doi:10.1111/ans.15775

Tomar, N. K. (2021, 01 11). *Automatic polyp segmentation using fully convolutional neural network*. Retrieved from arXiv [eess.IV]: <http://arxiv.org/abs/2101.04001><https://arxiv.org/abs/2101.04001>

Williams, J. G., Pullan, R. D., Hill, J., Horgan, P. G., Salmo, E., Buchanan, G. N., . . . Haboubi, N. (2013). Management of the malignant colorectal polyp: ACPGBI position statement. *Colorectal disease: the official journal of the Association of Coloproctology of Great Britain and Ireland*, 15(s2), 1-38. doi:10.1111/codi.12262

World Health Organization. (2023, 07 11). *Colorectal cancer*. Retrieved from Who.int: <https://www.who.int/news-room/fact-sheets/detail/colorectal-cancer>