DOI: https://doi.org/10.22516/25007440.471

Detección automática de pólipos colorrectales con técnicas de inteligencia artificial

Artificial intelligence techniques for the automatic detection of colorectal polyps

Martín Alonso Gómez-Zuleta,[1] Diego Fernando Cano-Rosales,[2*] Diego Fernando Bravo-Higuera, MSc,[3] Josué André Ruano-Balseca, MSc,[4] Eduardo Romero-Castro, PhD.[5]

ACCESO ABIERTO

Citación:
Gómez-Zuleta MA, Cano-Rosales DF, Bravo-Higuera DF, Ruano-Balseca JA, Romero-Castro E. Detección automática de pólipos colorrectales con técnicas de inteligencia artificial. Rev Colomb Gastroenterol. 2021;36(1):7-17. https://doi.org/10.22516/25007440.471

[1] Médico Internista y Gastroenterólogo. Unidad de Gastroenterología y Ecoendoscopia (UGEC). Hospital Universitario Nacional. Profesor asociado de Medicina, Universidad Nacional de Colombia.
[2] Médico Internista. *Fellow* de Gastroenterología, Universidad Nacional de Colombia. Hospital Universitario Nacional.
[3] Magister en Ingeniería Biomédica. Universidad Militar Nueva Granada. Docente cátedra. Universidad Nacional de Colombia. Bogotá D. C., Colombia.
[4] Magister en Ingeniería Biomédica, Ingeniero Biomédico, Estudiante de Doctorado en Ingeniería – Sistemas y computación. Universidad Nacional de Colombia. Bogotá D. C., Colombia.
[5] Doctor en Ciencias Biomédicas, Magister en Ingeniería Eléctrica, Médico cirujano. Profesor titular. Universidad Nacional de Colombia. Bogotá D. C., Colombia.

***Correspondencia:**
Diego Fernando Cano-Rosales.
dfcanor@unal.edu.co

Fecha recibido: 24/10/19
Fecha aceptado: 11/11/20

Resumen

El cáncer colorrectal (CCR) es uno de los tumores malignos con mayor prevalencia en Colombia y el mundo. Estas neoplasias se originan en lesiones adenomatosas o pólipos que deben resecarse para prevenir la enfermedad, lo cual se puede realizar con una colonoscopia. Se ha reportado que durante una colonoscopia se detectan pólipos en el 40 % de los hombres y en el 30 % de las mujeres (hiperplásicos, adenomatosos, serrados, entre otros), y, en promedio, un 25 % de pólipos adenomatosos (principal indicador de calidad en colonoscopia). Sin embargo, estas lesiones no son fáciles de observar por la multiplicidad de puntos ciegos en el colon y por el error humano asociado con el examen. Diferentes investigaciones han reportado que alrededor del 25 % de pólipos colorrectales no son detectados o se pasan por alto durante la colonoscopia y, como consecuencia, el paciente puede tener un cáncer de intervalo. Estas cifras muestran la necesidad de contar con un segundo observador (sistema de inteligencia artificial) que reduzca al mínimo la posibilidad de no detectar estos pólipos y, de este modo, sea posible prevenir al máximo el cáncer de colon. **Objetivo:** crear un método computacional para la detección automática de pólipos colorrectales usando inteligencia artificial en videos grabados de procedimientos reales de colonoscopia. **Metodología:** se usaron bases de datos públicas con pólipos colorrectales y una colección de datos construida en un Hospital Universitario. Inicialmente, se normalizan todos los cuadros de los videos para disminuir la alta variabilidad entre bases de datos. Posteriormente, la tarea de detección de pólipos se hace con un método de aprendizaje profundo usando una red neuronal convolucional. Esta red se inicia con pesos aprendidos en millones de imágenes naturales de la base de datos ImageNet. Los pesos de la red se actualizan usando imágenes de colonoscopia, siguiendo la técnica de ajuste fino. Finalmente, la detección de pólipos se realiza asignando a cada cuadro una probabilidad de contener un pólipo y determinando el umbral que define cuando el pólipo se encuentra presente en un cuadro. **Resultados:** este enfoque fue entrenado y evaluado con 1875 casos recopilados de 5 bases de datos públicas y de la construida en el hospital universitario, que suman aproximadamente 123 046 cuadros. Los resultados obtenidos se compararon con las marcaciones de diferentes expertos en colonoscopia y se obtuvo 0,77 de exactitud, 0,89 de sensibilidad, 0,71 de especificidad y una curva ROC (*receiver operating characteristic*) de 0,87. **Conclusión:** este método logra detectar pólipos de manera sobresaliente, superando la alta variabilidad dada por los distintos tipos de lesiones, condiciones diferentes de la luz del colon (asas, pliegues o retracciones) con una sensibilidad muy alta, comparada con un gastroenterólogo experimentado, lo que podría hacer que se disminuya el error humano, el cual es uno de los principales factores que hacen que no se detecte o se escapen los pólipos durante un examen de colonoscopia.

Palabras clave

Colonoscopia, cáncer colorrectal, pólipos, detección, inteligencia artificial.

Abstract

Colorectal cancer (CRC) is one of the most prevalent malignant tumors worldwide. These neoplasms originate from adenomatous lesions or polyps that must be resected to prevent the development of the disease, and that can be done through a colonoscopy. Polyps are reported during colonoscopy in 40% of men and 30% of women (hyperplastic, adenomatous, serrated, among others), and, on average 25% are adenomatous polyps (the main indicator of quality in colonoscopy). However, these lesions are not easy to visualize because of the multiplicity of blind spots in the colon and human errors associated with the performance of the procedure. Several research works have reported that about 25% of colorectal polyps are overlooked or undetected during colonoscopy, and as a result, the patient may have interval cancer. These figures show the need for a second observer (artificial intelligence system) to reduce the possibility of not detecting polyps and prevent colon cancer as much as possible. **Objective:** To create a computational method for the automatic detection of colorectal polyps using artificial intelligence using recorded videos of colonoscopy procedures. **Methodology:** Public databases of colorectal polyps and a data collection constructed in a university hospital were used. Initially, all the frames in the videos were normalized to reduce the high variability between databases. Subsequently, polyps were detected using a deep learning method with a convolutional neural network. This network starts with weights learned from millions of natural images taken from the ImageNET database. Network weights are updated using colonoscopy images, following the fine-tuning technique. Finally, polyps are detected by assigning each box a probability of polyp presence and determining the threshold that defines when the polyp is present in a box. **Results**: This approach was trained and evaluated with 1 875 cases collected from 5 public databases and the one built in the university hospital, which total approximately 123 046 frames. The results obtained were compared with the markings of different experts in colonoscopy, obtaining 0.77 accuracy, 0.89 sensitivity, 0.71 specificity, and a receiver operating characteristic curve of 0.87. **Conclusion:** This method detected polyps in an outstanding way, overcoming the high variability caused by the types of lesions and bowel lumen condition (loops, folds or retractions) and obtaining a very high sensitivity compared with an experienced gastroenterologist. This may help reduce the incidence of human error, as it is one of the main factors that cause polyps to not be detected or overlooked during a colonoscopy.

Keywords

Colonoscopy, Colorectal cancer, Polyps, Detection, Artificial intelligence.

INTRODUCCIÓN

El cáncer colorrectal (CCR) es el tercer cáncer más frecuente en el mundo y la segunda causa de muerte por cáncer. En Colombia es la cuarta neoplasia más frecuente en hombres y mujeres, con tasas de incidencia que aumentan cada año (1, 2). Muchos estudios concluyen que la tamización del CCR es costo-efectiva en población de riesgo medio (población sin antecedentes familiares y sin un historial médico que muestre predisposición). Se sabe que la edad (≥ 50 años), los hábitos alimentarios y el tabaco son factores de riesgo que aumentan la incidencia de padecer esta enfermedad. En la población general, el riesgo es del 5 %-6 % y esta incidencia aumenta de forma sustancial a partir de los 50 años, por lo cual se considera que las personas de 50 años o más son población en riesgo medio, para la cual se debería iniciar un programa de tamización (3, 4).

En cuanto al grado de supervivencia en los pacientes con CCR, está directamente relacionado con la extensión de la enfermedad en el momento del diagnóstico. Los individuos diagnosticados en estado avanzado tienen una tasa de supervivencia del 7 % a los 5 años, mientras que para sujetos con CCR detectado en un estado inicial se ha reportado una tasa del 92 % (5); por esta razón, es de gran importancia detectar el tumor en estadios tempranos o, más aún, detectar el pólipo en estado adenomatoso (premaligno), con lo cual se previene la enfermedad. Se sabe que con las técnicas de tamización disponibles (sangre oculta, colonoscopia), el CCR es altamente prevenible en más del 90 % de los casos.

Múltiples trabajos han demostrado que la colonoscopia es el examen de elección para la prevención y detección temprana del CCR porque, como se mencionó previamente, es capaz de detectar el origen principal del CCR como son los pólipos adenomatosos (6-9).

Además de detectar el cáncer en estados tempranos, el cual si se trata a tiempo es completamente curable, la detección de pólipos es un indicador de calidad en la colonoscopia y se considera que durante el examen se encuentren pólipos adenomatosos (los cuales tienen alto riesgo de cáncer) en un 20 % de mujeres y en un 30 % de hombres; es decir que, en promedio, se deberían encontrar pólipos adenomatosos en un 25 % de todas las colonoscopias que se realizan. Infortunadamente, diferentes estudios han reportado que alrededor del 26 % de los pólipos que están presentes en una colonoscopia no se detectan, una tasa de error muy alta explicada básicamente por dos factores:

la cantidad de puntos ciegos durante una colonoscopia (pólipos ubicados detrás de los pliegues, asas del colon, la preparación, entre otros) y el error humano (se pasaron por alto) asociado con el procedimiento (10-12). Se han realizado múltiples trabajos que buscan atacar estos dos factores para disminuir esta tasa de pólipos perdidos al máximo, es así como se han diseñado accesorios que permiten encontrar los pólipos ocultos detrás de los pliegues como son Cap, Endocuff o incluso un miniendoscopio denominado *tercer ojo*, que busca aplanar los pliegues o ver detrás de ellos. Adicionalmente, recientemente se ha considerado que el factor asociado con el error humano es al menos mitigable con la introducción de segundos lectores (computadores), un escenario en el cual la tecnología y la inteligencia artificial empiezan a mostrar resultados que pueden mejorar drásticamente la tasa de detección de los pólipos y permitir bajar el número de pólipos no detectados en una unidad de gastroenterología.

El desarrollo de estrategias computacionales para la extracción de patrones y la detección automática de pólipos colorrectales en videos de colonoscopia es un problema muy complejo. Los videos de una colonoscopia se registran en medio de una gran cantidad de fuentes de ruido que fácilmente ocultan lesiones; por ejemplo, los brillos en la pared intestinal producidos por la fuente de luz o reflexión especular, la motilidad de los órganos y la secreción intestinal que ocluyen el campo de visión del colonoscopio, y la experiencia del especialista que influye en la suavidad de la exploración del colon. Actualmente, varias estrategias han abordado este reto como una tarea de clasificación, utilizando técnicas automáticas de aprendizaje de máquina.

Por una parte, algunos autores han intentado una selección de características de bajo nivel para obtener límites de pólipos candidatos. Bernal y colaboradores (13) presentaron un modelo de apariencia de pólipos que caracteriza los valles de los pólipos como límites cóncavos y continuos. Esta caracterización se usa para entrenar un clasificador que en un conjunto de prueba (test) obtuvo 0,89 de sensibilidad en la tarea de detección de pólipos. Shin y colaboradores (14) presentaron una estrategia basada en una clasificación por parches, usando una combinación de características de forma y color, y obtuvieron una sensibilidad de 0,86. Por otra parte, varios trabajos han utilizado redes neuronales convolucionales (CNN) profundas, un conjunto de algoritmos agrupados bajo el término de *aprendizaje profundo*. Urban y colaboradores (15) presentaron una red convolucional que detecta pólipos de diferentes tamaños en tiempo real con una sensibilidad de 0,95. Sin embargo, Taha y colaboradores (16) analizaron algunas de las limitaciones de estos trabajos, una de ellas es el hecho de que estos métodos requieren una gran cantidad de datos para ser entrenados. Además, estas bases de datos son adquiridas en condiciones clínicas específicas; en particular, el dispositivo de captura, el protocolo de exploración realizado por el experto y la extracción de las secuencias con lesiones fácilmente visualizables. Aunque se han presentado varios avances, aún existe el reto de formular modelos generalizables para detectar lesiones de manera precisa, independientemente del tipo de lesión, forma de exploración del experto o de la unidad de colonoscopia usada.

El principal objetivo del presente trabajo es crear una estrategia automática de detección de pólipos colorrectales con el propósito de construir un segundo lector que soporte el proceso de exploración del colon y de disminuir el número de lesiones no detectadas durante una colonoscopia. En este documento se presenta una estrategia de clasificación automática de pólipos en secuencias de videos de colonoscopia. Esta investigación se apoya en un algoritmo de aprendizaje profundo y evalúa diferentes arquitecturas de redes convolucionales. Este artículo está organizado de la siguiente manera: inicialmente, se presenta la metodología para la detección automática de pólipos; luego, se describen las consideraciones éticas alrededor de este trabajo; posteriormente, se muestra la configuración experimental junto con los resultados del método detectando pólipos comparados con las anotaciones de un experto; después, se presenta la discusión de este trabajo; y, finalmente, se encuentran las conclusiones y trabajo futuro.

METODOLOGÍA

Este trabajo presenta una metodología de aprendizaje profundo para modelar la alta variabilidad en un procedimiento de colonoscopia, con el propósito de realizar una detección automática de pólipos en procedimientos de colonoscopia. Esta tarea se divide en dos etapas: entrenamiento y clasificación. En primer lugar, se realiza un preprocesamiento cuadro a cuadro, común para las dos etapas. Después, se entrena una red neuronal convolucional usando una gran cantidad de imágenes de colonoscopia anotadas por un gastroenterólogo experto en colonoscopia (con alrededor de 20 años de experiencia y más de 50 mil colonoscopias realizadas) en dos clases: clase negativa o *no contiene pólipo*, y clase positiva o *contiene pólipo*. El modelo obtenido del proceso de aprendizaje es utilizado para clasificar imágenes nuevas (o no usadas en el proceso de entrenamiento) como perteneciente a alguna de las dos clases. El flujo de este trabajo se visualiza en la **Figura 1** y se explica a continuación.

PROTOCOLO DE ADQUISICIÓN Y PREPROCESAMIENTO

Para disminuir el efecto de las numerosas fuentes de ruido en el proceso de adquisición de diferentes colonoscopios

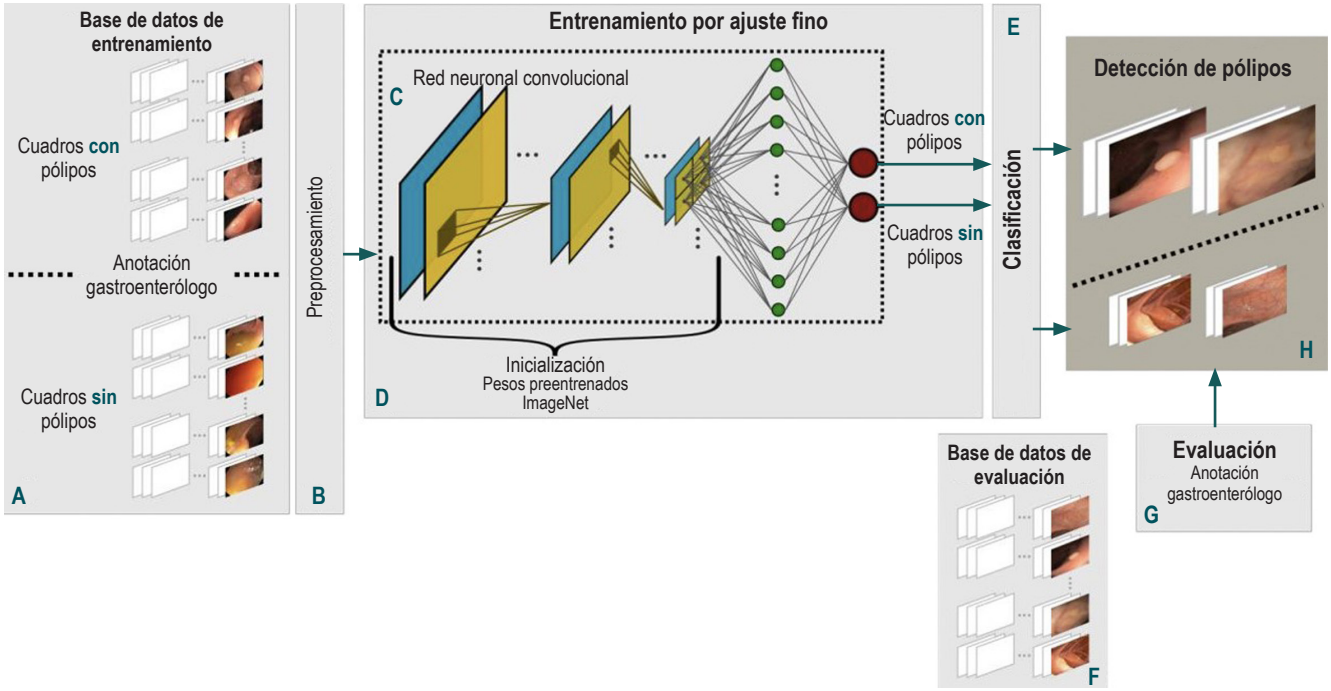

Figura 1. Flujo del método propuesto para detectar pólipos automáticamente. Primero, se consolidó una base de datos anotada cuadro a cuadro de videos de colonoscopia (**A**). Cada uno de estos cuadros son preprocesados (**B**) para alimentar unos modelos basados en CNN (**C**). Este modelo se entrena con un ajuste fino de unos pesos preentrenados con millones de imágenes naturales (**D**). Con la red entrenada, se evalúa su rendimiento para detectar pólipos (**H**) con una base de datos de prueba (**F**) y se comparan los resultados obtenidos con las anotaciones de un experto (**G**).

y las condiciones fisiológicas de colon y el recto, es necesario realizar un preprocesamiento cuadro por cuadro del video. Primero, se normaliza cada cuadro con media 0 y desviación estándar (DE) de 1, con el fin de que las características extraídas entre cuadros sean comparables. Luego, dependiendo del dispositivo de captura, los cuadros presentan distintas resoluciones espaciales, por lo cual cada cuadro es escalado recortado a 300 x 300 pixeles, de manera que todos tengan la misma malla de captura.

ARQUITECTURA DE LAS CNN

La unidad principal de estas arquitecturas es la neurona, que proporciona una salida como función de las entradas a ella. Un arreglo de neuronas forma una capa o bloque y una red está compuesta por varios bloques elementales que se ordenan de la siguiente manera: varios pares de capas convolucionales (**Figura 1 C**, cuadro azul) y de agrupamiento (**Figura 1 C**, cuadro amarillo) que entregan un vector de características de la imagen, seguidos por un conjunto de capas completamente conectadas (**Figura 1 C**, círculos verdes) que se encargan de calcular la probabilidad de que un conjunto de características pertenezca a una cierta clase, y se finaliza con una capa de activación (**Figura 1 C**, círculos rojos), en la cual se normalizan las probabilidades obtenidas y se logra la clasificación binaria deseada. La función de estos bloques es:

- Capas convolucionales (*convolutional layers*): identifica las características locales en toda la imagen como patrones de forma, bordes y textura, vitales en la descripción de pólipos. Esta capa conecta un subconjunto de pixeles vecinos de la imagen o neuronas con todos los nodos de la primera capa convolucional. Una de estas capas o kernel convolucional se distingue por los pesos específicos de cada nodo; al ser operado sobre una región específica de la imagen, proporciona un mapa de característica de la región.
- Capas de agrupamiento (*pooling layers*): reduce la complejidad computacional, que a su vez disminuye el tamaño de las características en las capas convolucionales y se obtiene un conjunto jerárquico de los mapas de características de la imagen.
- Capas completamente conectadas (*fully-connected layers*): esta capa conecta cada una de las neuronas de la capa previa a cada una de las neuronas en la siguiente capa. La capa previa es una representación plana o vec-

tor de los mapas de características obtenidos. El número de neuronas de la siguiente capa es determinado por el número de clases que se requiere clasificar. Finalmente, la capa completamente conectada provee una votación para determinar si una imagen pertenece a una clase específica.

- Función de activación (*activation function*): normaliza las probabilidades obtenidas de las capas completamente conectadas según una función específica, en las que se obtiene una probabilidad de 0 a 1.

Una arquitectura en particular se compone por un arreglo de módulos que contienen diferentes configuraciones y órdenes de bloques fundamentales explicados anteriormente, y se conoce como *gradiente* al resultado obtenido por cada neurona. En este trabajo se utilizaron tres arquitecturas altamente evaluadas y validadas en el estado del arte: InceptionV3, Vgg16 y ResNet50. A continuación se describe cada una de ellas.

- InceptionV3: se compone de 48 capas con 24 millones de parámetros. En gran parte, estas capas están agrupadas en 11 módulos, en los cuales se extraen características a múltiples niveles. Cada módulo se compone de una configuración determinada de capas convolucionales y de agrupamiento, rectificadas por la función *unidad lineal rectificadora* (ReLu). Finaliza con una función de activación llamada *exponencial normalizada* (*softmax*) (17).
- Vgg16: se organiza en 16 capas para un total de 138 000 parámetros. 13 de las capas son convolucionales, con una de agrupamiento (en algunas), 2 capas completamente conectadas y finaliza con una función de activación exponencial normalizada. Esta arquitectura se destaca por usar pequeños filtros de tamaño 3 x 3 en las capas convolucionales. En comparación con la mayoría de arquitecturas, el costo computacional de esta es menor (18).
- ResNet50: se compone de 50 capas con 26 millones de parámetros. Esta arquitectura se construye bajo el concepto de redes residuales. Es común que en arquitecturas muy profundas como la mencionada, el gradiente propagado se desvanezca en las últimas capas. Para evitar esto, ciertas capas son entrenadas con el residuo del gradiente obtenido en esta y el gradiente de una capa dos posiciones antes. Esta arquitectura finaliza con una función de activación exponencial normalizada (19).

ENTRENAMIENTO POR AJUSTE FINO

Un alto rendimiento en la clasificación de las clases depende en su gran mayoría de la cantidad de imágenes anotadas y la forma de iniciar los pesos para entrenar los CNN. Una colonoscopia tiene aproximadamente 12 000 cuadros por video, por lo cual la disponibilidad de bases de datos con imágenes anotadas es limitada. Entonces, entrenar con un número limitado de datos e iniciar los pesos de la red de manera aleatoria, como se hace generalmente, resulta en un proceso de entrenamiento fallido. Para evitar este inconveniente, se usan pesos (*transfer learning*) de redes del mismo tipo, que han sido previamente entrenados para otro problema de clasificación en imágenes naturales, con bases de datos que contienen grandes cantidades de imágenes anotadas. La razón por la cual se hace de esta manera es que, aun cuando las imágenes naturales y las de colonoscopia sean diferentes, su estructura estadística es similar y, asimismo, la construcción de primitivas que representan los objetos. En estas circunstancias, las redes entrenadas para reconocer objetos en imágenes naturales se usan como condición inicial para entrenar estas redes en la tarea de reconocer pólipos.

El uso de estos pesos se realiza por medio de un proceso llamado ajuste fino (*fine tuning*), para el cual se toma toda la red preentrenada y se retira la última capa completamente conectada. Esta capa es reemplazada por una nueva, que tiene el mismo número de neuronas que el número de clases en la tarea de clasificación (*pólipo-no pólipo*) y se inicia con los pesos de la red preentrenada. Entonces, primero se entrena la última capa y, posteriormente, se actualizan los pesos del resto de capas de la red en un proceso iterativo; esta metodología se conoce como *propagación hacia atrás*. Cada iteración de este entrenamiento se realiza usando un cierto número de muestras o lotes (*batch*) de las imágenes de entrenamiento. Este proceso termina cuando la red fue entrenada con todas las muestras del conjunto, conocido como una época (*epoch*) de entrenamiento. El número de épocas se determina según la complejidad de las muestras a clasificar. Finalmente, el entrenamiento culmina cuando la probabilidad de una imagen de entrenamiento sea alta y concuerde con la etiqueta anotada.

DETECCIÓN DE PÓLIPOS

Usando el modelo de la red entrenada, este se aplica a un conjunto de videos de evaluación en el que se clasifica y asigna una etiqueta: (1) cuadros con y (0) sin presencia de pólipos. Sin embargo, hay cuadros con estructuras que se asemejan a la apariencia de un pólipo, como son las burbujas producidas por los fluidos intestinales. En estos cuadros, el modelo presenta un error de clasificación, tomando este cuadro como si tuviera una lesión presente. Analizando temporalmente estos errores, es notable que se presentan como valores atípicos (de 3 a 10 cuadros) en una ventana de tiempo pequeña (60 cuadros o 2 segundos). Por tanto, la clasificación realizada por la red es filtrada temporalmente y

determina que, si al menos el 50 % de 60 cuadros contiguos son clasificados sin presencia de pólipos, el resto de los cuadros son filtrados y se les asigna una nueva etiqueta, como cuadros que no contienen pólipo. Finalmente, un pólipo es detectado cuando el método propuesto clasifica una imagen como cuadro con pólipo presente o clase positiva.

BASES DE DATOS

La construcción de la base de datos en este trabajo tuvo como propósito capturar la mayor variabilidad de un procedimiento de colonoscopia. Para entrenar y evaluar el enfoque propuesto se reunieron secuencias de diferentes centros de gastroenterología que contienen lesiones polipoides y no polipoides de tamaños variados (morfología y ubicación en el colon), exploraciones hechas por distintos expertos y equipos de captura. A continuación, se detallan estas bases de datos.

ASU-Mayo Clinic Colonoscopy Video Database

Este conjunto se construyó en el Departamento de Gastroenterología de la Clínica Mayo en Arizona, Estados Unidos. Consta de 20 secuencias de colonoscopia, divididas en 10 con presencia de pólipos y 10 sin presencia de ellos. Las anotaciones fueron realizadas por estudiantes de gastroenterología y validadas por un especialista experto. Esta colección se ha usado con gran frecuencia en el estado del arte y resalta como la base de datos para el evento "2015 ISBI Grand Challenge on Automatic Polyp Detection in Colonoscopy Videos" (20).

CVC-ColonDB

Se compone de 15 secuencias cortas de diferentes lesiones, acumulando un total de 300 cuadros. Las lesiones de esta colección presentan una alta variabilidad y dificultad de detección, ya que son bastante similares a las regiones sanas. Cada cuadro fue anotado por un experto gastroenterólogo. Esta colección fue construida en el Hospital Clínico de Barcelona, España (13).

CVC-ClinicDB

Consiste en 29 secuencias cortas con diferentes lesiones que reúnen 612 cuadros anotados por un experto. Esta base de datos fue utilizada por el conjunto de entrenamiento del evento MICCAI 2015 Sub-Challenge on Automatic Polyp Detection Challenge in Colonoscopy Videos. Esta colección fue construida en el Hospital Clínico de Barcelona, España (21).

ETISLarib Polyp DB

Presenta 196 imágenes con pólipos cada una anotada por un experto. Esta base de datos fue utilizada en el conjunto de prueba para el evento MICCAI 2015 Sub-Challenge on Automatic Polyp Detection Challenge in Colonoscopy Videos (22).

The Kvasir Dataset

Es una base de datos que se recopilaron utilizando equipos endoscópicos en Vestre Viken Health Trust (VV), en Noruega. Las imágenes son anotadas por uno o más expertos médicos de VV y el Registro de Cáncer de Noruega (CRN). El conjunto de datos consta de las imágenes con una resolución diferente de 720 x 576 hasta 1920 x 1072 píxeles (20).

HU-DB

Esta colección fue construida en el Hospital Universitario en Bogotá, que contiene 253 videos de colonoscopias con un total de 233 lesiones. Cada cuadro de los videos fue anotado por un experto en colonoscopia con alrededor de 20 años de experiencia y más de 50 000 colonoscopias realizadas.

Cada uno de estos videos se capturó a 30 cuadros por segundo y a una resolución espacial de 895 x 718, 574 x 480 y 583 x 457. En total, se consolidó una base de datos con 1875 casos y un total de 48 573 cuadros con presencia de pólipos y 74 548 cuadros sin presencia de pólipos. Cada uno de los cuadros de estos videos fue anotado por un experto como positivo si había presencia de pólipo, o negativo cuando no había presencia de pólipo. En la **Tabla 1** se resume el número de videos y cuadros por base de datos utilizados en este trabajo.

Tabla 1. Descripción de la cantidad de videos o casos y cuadros de videos de colonoscopia por cada una de las bases de datos usadas en este trabajo*

Base de datos	Número de videos		Cuadros	
	Pólipo	No pólipo	Pólipo	No pólipo
ASU-Mayo	10	10	4683	13481
CVC-ClinicDB	29	0	612	0
CVC-ColonDB	15	0	379	0
ETIS	28	0	196	0
Kvasir	1000	500	1000	500
HU	**233**	**50**	**41 703**	**60 567**
Total	**1315**	**560**	**48 573**	**74 548**

*La consolidación de varias bases de datos para entrenar y evaluar la metodología propuesta permite abarcar una gran variabilidad de lesiones.

CONSIDERACIONES ÉTICAS

El presente trabajo está acorde a la resolución n.° 008430 de 1993, que establece las normas científicas, técnicas y administrativas para la investigación en humanos (artículo 11). Este proyecto se clasifica como investigación con riesgo mínimo, dado que solo se requiere del uso de imágenes digitales, las cuales se generan a partir de videos de colonoscopias anonimizados; es decir, no existe manera alguna de conocer el nombre o la identificación de los sujetos incluidos en el estudio.

RESULTADOS

Las CNN utilizadas en este trabajo son InceptionV3, Resnet50 y Vgg16. Las etiquetas asignadas por cada una de estas redes fueron comparadas con las anotaciones realizadas por los especialistas en cada uno de los cuadros. La siguiente configuración experimental y la metodología de evaluación fueron aplicadas a cada una de las arquitecturas.

Configuración experimental

Los CNN fueron entrenados previamente con imágenes de la base de datos pública ImageNet, esta contiene un aproximado de 14 millones de imágenes naturales. Los pesos resultantes son utilizados para iniciar un nuevo proceso de entrenamiento de cuadros de colonoscopia por la metodología de ajuste fino. Este método actualiza los pesos, entrenando la red con la base de datos de colonoscopia. La actualización de los pesos fue realizada con 120 épocas sobre la totalidad del conjunto de entrenamiento. Cada época entrenaba el modelo tomando un lote de 32 cuadros hasta abarcar todos los cuadros en su totalidad. Para cada una de las redes, el umbral de decisión fue ajustado manualmente, orientado a mantener un equilibrio en el desempeño de clasificación para ambas clases. El esquema de entrenamiento fue 70 % de la base de datos para entrenar y un 30 % para validar respecto al número de casos; es decir que los datos se separan desde el principio y los datos de entrenamiento, validación y prueba nunca se mezclan. En total, las redes fueron entrenadas y validadas con 213 casos (24 668 cuadros) con pólipos y 36 videos (27 534 cuadros) sin pólipos. La evaluación se realizó con 103 videos (23 831 cuadros) con pólipos y 25 videos (47 013 cuadros) sin pólipos. El detalle de esta colección se presenta en la **Tabla 2**.

Evaluación cuantitativa

El enfoque propuesto detecta automáticamente pólipos en videos de colonoscopia; esta tarea está enmarcada como un problema de clasificación binaria. Este método establece una etiqueta a cada cuadro como clase negativa (cuadro que no contiene pólipo) o clase positiva (cuadro que contiene pólipo). Para evaluar el rendimiento de este trabajo, se compara la etiqueta estimada o predicha con la etiqueta anotada por el experto. Esta comparación permite calcular la matriz de confusión, que contabiliza lo siguiente:

- Verdaderos positivos (*true-positives* [TP]): es la cantidad de cuadros que fueron clasificados correctamente como clase positiva por el modelo.
- Verdaderos negativos (*true-negatives* [TN]): es la cantidad de cuadros que fueron clasificados correctamente como clase negativa por el modelo.
- Falsos positivos (*false-positives* [FP]): es la cantidad de cuadros que fueron clasificados incorrectamente como clase positiva por el modelo.
- Falsos negativos (*false-negatives* [FN]): es la cantidad de cuadros que fueron clasificados incorrectamente como clase negativa por el modelo.

Tabla 2. Descripción de la cantidad de secuencias y cuadros escogidos de cada base de datos para evaluar el desempeño de la metodología propuesta*

Base de datos	Número de videos		Cuadros	
	Pólipo	No pólipo	Pólipo	No pólipo
ASU-Mayo	5	2	2124	2553
CVC-ClinicDB	9	0	191	0
CVC-ColonDB	4	0	145	0
ETIS	7	0	45	0
HU	78	23	21 326	44 460
Total	**103**	**25**	**23 831**	**47 013**

*Esto corresponde aproximadamente al 30 % de la base datos en total.

Usando la matriz de confusión, se seleccionaron y calcularon 4 métricas de clasificación que evalúan el desempeño del método para clasificar cuadros con (clase positiva) y sin (clase negativa) pólipo independientemente, y el poder de predicción en ambas clases en general:

- La sensibilidad mide la proporción de cuadros correctamente clasificados que contienen pólipos.
- La especificidad calcula la proporción de cuadros correctamente clasificados que no contienen pólipos.
- La precisión indica el poder predictivo del método para clasificar cuadros con pólipos.
- La exactitud es la tasa de cuadros clasificados correctamente, según el número total de estos.

Los resultados obtenidos se presentan por cada una de las arquitecturas de aprendizaje profundo explicadas en

la sección de metodología. En la **Tabla 3** se presentan los resultados obtenidos por cada una de las arquitecturas.

Tabla 3. Resultados obtenidos por el método propuesto*

Métrica	InceptionV3	Resnet50	Vgg16
Exactitud	0,81	0,77	0,73
Sensibilidad	0,82	0,89	0,81
Especificidad	0,81	0,71	0,70
Precisión	0,67	0,59	0,56
Puntaje F1	0,74	0,71	0,66
ROC (área bajo la curva)	0,85	0,87	0,81

*En las columnas se especifica la arquitectura en evaluación y en las filas, cada una de las métricas utilizadas.

Por una parte, aunque la mayoría de estas arquitecturas muestran un rendimiento sobresaliente en la tarea de clasificación, la arquitectura Resnet50 presenta las mejores métricas en términos de qué tan bien detectó la clase positiva o cuadros con pólipos, y se obtuvo un 0,89 de sensibilidad. Por otra parte, la arquitectura InceptionV3 fue la que mejor detectó la clase negativa o cuadros sin pólipos, y se obtuvo un 0,81 de especificidad. Para evaluar de manera más detallada el rendimiento de estas arquitecturas, se construyeron las curvas ROC (*receiver operating characteristic*) por arquitectura. En esta representación se busca analizar cómo los modelos clasifican las imágenes en términos de especificidad y sensibilidad variando el umbral de decisión sobre las probabilidades entregadas por el modelo. Como se puede apreciar en la **Figura 2**, la arquitectura Resnet50 separa mejor las clases independientemente del umbral de decisión. Esto indica que esta arquitectura logró generalizar mejor la variabilidad intra- e interclases.

DISCUSIÓN

La detección de los pólipos adenomatosos es el principal indicador de calidad en colonoscopia, dado que es un marcador fundamental para la detección y prevención del CCR. En muchos países, la calidad del gastroenterólogo se mide por el número de estos pólipos que detecta en todas sus colonoscopias y en promedio gira alrededor de un 25 % para el experto, pero puede ser tan baja como del 10 % para el gastroenterólogo inexperto, lo cual lleva a que a este último se le escapen más los adenomas.

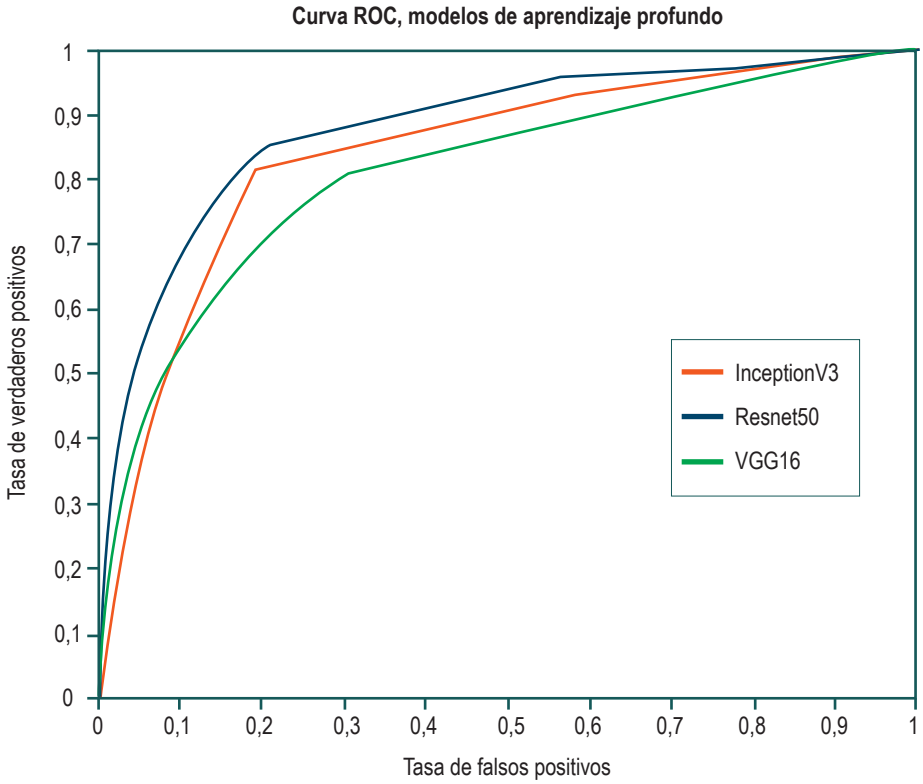

Figura 2. Curvas ROC para cada una de las arquitecturas evaluadas. La línea naranja corresponde a la curva de la arquitectura InceptionV3; la línea azul, a la arquitectura Resnet50; y la línea verde, a la arquitectura Vgg16. La arquitectura Resnet50 presenta un mejor desempeño, con un área bajo la curva de 0,87.

Es así como varios estudios (10-12) reportan que el 26 % de los pólipos no se detectan durante las colonoscopias, lo cual puede contribuir para que se presenten más casos de CCR. Es así como se presentaron 1,8 millones de nuevos casos en el mundo para el 2018 (International Agency for Research on Cancer, 2018) (1). Esta tasa de pérdida se debe a que hay varios factores que afectan una exploración adecuada del colon como la experiencia y el nivel de concentración (asociado con la fatiga) del experto durante toda una jornada laboral, las condiciones fisiológicas del colon como puntos ciegos en las haustras y la dificultad de ubicar el colonoscopio por la motilidad propia del órgano, y la preparación previa del colon por parte del paciente, que determina qué tan observables son las paredes del colon, según el nivel de limpieza de estas (23). La mayoría de estos factores advierte que la colonoscopia es altamente dependiente del factor humano, exhibiendo una necesidad de contar con segundos lectores que no se vean afectados por estos factores. El uso de herramientas computacionales para la detección de pólipos en la práctica clínica ayudaría a corroborar los hallazgos realizados por el experto y, de mayor importancia, alertar sobre posibles lesiones que el experto no detectó. De este modo, estas herramientas ayudarían a disminuir las tasas de pólipos no detectados y, por ende, disminuir la incidencia del CCR.

Para dar soporte al diagnóstico del CCR usando herramientas por visión de computador, este reto se ha abordado de la siguiente manera:

- detección, refiriéndose a la clasificación binaria cuadro a cuadro de un video en clase positiva (con pólipo) y en clase negativa (sin pólipo);
- localización, como la delimitación gruesa (por medio de un recuadro) de la lesión sobre una imagen que contiene pólipo;
- segmentación, como una delimitación fina de la lesión (delineando el borde del pólipo).

La detección de pólipos es la primera tarea y principal que debe afrontar el gastroenterólogo. Las tareas posteriores a la detección (la localización y segmentación) son procesos de utilidad para el experto cuando ya ha detectado la lesión y necesita describirla morfológicamente, tomando como referencia guías médicas como la Clasificación de París (6). Esta clasificación le permite decidir el manejo quirúrgico de la enfermedad a corto y largo plazo. Consecuentemente, estas tareas dependen totalmente de qué tan precisa sea la detección previa; por tanto, la metodología propuesta se enfoca exclusivamente en la tarea principal que requiere el experto: obtener cuadros de colonoscopia con presencia de lesiones. Además, en el estado del arte, los trabajos que han abordado estas tareas (13-15) describen limitaciones para presentar un solo flujo abarcando al menos dos de estas. Estos trabajos usan metodologías diferentes para cada tarea, ya que cada una tiene su propio nivel de complejidad. En general, para detectar cuadros con pólipos se miden relaciones contextuales o globales a nivel de la imagen; mientras que la localización y segmentación analizan a nivel del píxel midiendo relaciones locales.

Este trabajo presenta una estrategia robusta para la detección de pólipos, solventada como un problema de clasificación. Las redes profundas para tareas de clasificación son métodos que fueron formulados décadas atrás, pero no habían sido explotadas ya que la potencia de cómputo y la disponibilidad de bases de datos anotadas era limitada. En los últimos 5 años, el uso de estos modelos ha aumentado drásticamente a causa de desarrollos tecnológicos que permiten una gran cantidad de procesamiento en paralelo y la publicación de bases de datos con millones de imágenes como ImageNet. Esto permitió diseñar redes altamente complejas y entrenarlas exhaustivamente, de modo que se obtuvo un alto rendimiento en tareas de clasificación, ya que es capaz de modelar una alta variabilidad de formas, colores y texturas. Sin embargo, en el ámbito médico, no se dispone de una gran cantidad de datos públicos anotados, por lo que no se contemplaba aplicar estos modelos a problemas de detección o clasificación de enfermedades.

El desarrollo de técnicas de transferencia de aprendizaje (o *transfer learning*) proporcionó una solución a la escasez de datos médicos. Los pesos de las redes entrenadas con millones de imágenes naturales se utilizaron para iniciar una nueva red y entrenarla con una cantidad mucho menor de datos diferentes, como imágenes de colonoscopia. Los trabajos del estado del arte que han utilizado este flujo demuestran que tiene la capacidad de generalizar adecuadamente la alta variabilidad de cuadros con y sin lesiones polipoides en imágenes de colonoscopia extraídas de una base de datos en particular. Sin embargo, los diferentes tipos de lesiones y las condiciones fisiológicas típicas del intestino grueso no son la única fuente de variabilidad. Cuanto menor sea la experticia del especialista, los videos son propensos a tener una mayor cantidad de cuadros ruidosos producidos por oclusiones y movimientos abruptos del colonoscopio. Adicionalmente, los dispositivos de captura varían en las fuentes de luz y los ángulos de visión de las cámaras. Por tanto, entrenar y validar con bases de datos obtenidas de un solo servicio de gastroenterología en específico, como lo hacen los trabajos del estado del arte (13-15) que han presentado rendimientos sobresalientes, no abarca toda la variabilidad que tiene la tarea de clasificación de imágenes de colonoscopia.

Debido a lo anterior, en este trabajo se consolidó un conjunto de videos de entrenamiento con una alta variabilidad que no se ha presentado en el estado del arte al reunir secuencias de distintas bases de datos. El conjunto usado para entre-

nar este enfoque contiene: lesiones de distintos tamaños, posiciones y formas; procedimientos de colonoscopia y anotaciones realizados por distintos expertos gastroenterólogos; y videos capturados usando diferentes unidades de colonoscopia. A pesar de dicha variabilidad, este trabajo obtiene una sensibilidad de 0,89 y una especificidad de 0,71 en la tarea de detección de pólipos en secuencias de colonoscopia.

CONCLUSIONES

Las metodologías de aprendizaje profundo actualmente son una opción prometedora para ser usadas en tareas de clasificación médica. El avance de la tecnología junto al diseño y evaluación constante de las redes ha permitido consolidar un conjunto de métodos y flujos para que tengan un alto desempeño. Con las redes evaluadas en este trabajo, los resultados obtenidos demuestran que pueden ser usadas de rutina como segundos lectores en un servicio de colonoscopia.

Es notable que estas redes generalizan adecuadamente la alta variabilidad de los videos de colonoscopia. Los resultados obtenidos demuestran que el método propuesto puede diferenciar sobresalientemente imágenes con y sin presencia de pólipos, independientemente del protocolo clínico particular con el que se grabó el video, refiriéndose al experto que realiza el procedimiento y el dispositivo de captura. Este método podría ser útil para disminuir la brecha entre el gastroenterólogo experto y el principiante en la tasa de detección de adenoma.

Como trabajo futuro, el enfoque propuesto debe ser sometido en procedimientos de colonoscopia completo y evaluar si es posible que sea implementado en tiempo real, además de desarrollar una estrategia que permita no solo detectar, sino también delimitar la lesión dentro del cuadro.

REFERENCIAS

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424. http://dx.doi.org/10.3322/caac.21492
2. Data I, Method L. Globocan Colombia 2018. 2018;380:1-2.
3. Samadder NJ, Curtin K, Tuohy TM, Pappas L, Boucher K, Provenzale D, Rowe KG, Mineau GP, Smith K, Pimentel R, Kirchhoff AC, Burt RW. Characteristics of missed or interval colorectal cancer and patient survival: a population-based study. Gastroenterology. 2014;146(4):950-60. http://dx.doi.org/10.1053/j.gastro.2014.01.013
4. Kaltenbach T, Sano Y, Friedland S, Soetikno R; American Gastroenterological Association. American Gastroenterological Association (AGA) Institute technology assessment on image-enhanced endoscopy. Gastroenterology. 2008;134(1):327-40. http://dx.doi.org/10.1053/j.gastro.2007.10.062
5. Brown SR, Baraza W, Din S, Riley S. Chromoscopy versus conventional endoscopy for the detection of polyps in the colon and rectum. Cochrane Database Syst Rev. 2016;4:CD006439. http://dx.doi.org/10.1002/14651858.CD006439.pub4
6. The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to December 1, 2002. Gastrointest Endosc. 2003;58(6 Suppl):S3-43. http://dx.doi.org/10.1016/s0016-5107(03)02159-x
7. Dinesen L, Chua TJ, Kaffes AJ. Meta-analysis of narrow-band imaging versus conventional colonoscopy for adenoma detection. Gastrointest Endosc. 2012;75(3):604-11. http://dx.doi.org/10.1016/j.gie.2011.10.017
8. Nagorni A, Bjelakovic G, Petrovic B. Narrow band imaging versus conventional white light colonoscopy for the detection of colorectal polyps. Cochrane Database Syst Rev. 2012;1:CD008361. http://dx.doi.org/10.1002/14651858.CD008361.pub2
9. Jin XF, Chai TH, Shi JW, Yang XC, Sun QY. Meta-analysis for evaluating the accuracy of endoscopy with narrow band imaging in detecting colorectal adenomas. J Gastroenterol Hepatol. 2012;27(5):882-7. http://dx.doi.org/10.1111/j.1440-1746.2011.06987.x
10. Komeda Y, Suzuki N, Sarah M, Thomas-Gibson S, Vance M, Fraser C, Patel K, Saunders BP. Factors associated with failed polyp retrieval at screening colonoscopy. Gastrointest Endosc. 2013;77(3):395-400. http://dx.doi.org/10.1016/j.gie.2012.10.007
11. Choi HN, Kim HH, Oh JS, Jang HS, Hwang HS, Kim EY, Kwon JG, Jung JT. [Factors influencing the miss rate of polyps in a tandem colonoscopy study]. Korean J Gastroenterol. 2014;64(1):24-30. http://dx.doi.org/10.4166/kjg.2014.64.1.24
12. van Rijn JC, Reitsma JB, Stoker J, Bossuyt PM, van Deventer SJ, Dekker E. Polyp miss rate determined by tandem colonoscopy: a systematic review. Am J Gastroenterol. 2006;101(2):343-50. http://dx.doi.org/10.1111/j.1572-0241.2006.00390.x
13. Bernal J, Sánchez J, Vilariño F. Towards automatic polyp detection with a polyp appearance model. Pattern Recognition. 2012;45(9):3166-82. https://doi.org/10.1016/j.patcog.2012.03.002
14. Younghak Shin, Balasingham I. Comparison of hand-craft feature based SVM and CNN based deep learning fra-

mework for automatic polyp classification. Annu Int Conf IEEE Eng Med Biol Soc. 2017;2017:3277-3280. http://dx.doi.org/10.1109/EMBC.2017.8037556
15. Urban G, Tripathi P, Alkayali T, Mittal M, Jalali F, Karnes W, Baldi P. Deep Learning Localizes and Identifies Polyps in Real Time With 96% Accuracy in Screening Colonoscopy. Gastroenterology. 2018;155(4):1069-1078. e8. http://dx.doi.org/10.1053/j.gastro.2018.06.037
16. Taha B, Werghi N, Dias J. Automatic Polyp Detection in Endoscopy Videos: A Survey. Biomed Eng. 2017. http://dx.doi.org/10.2316/P.2017.852-031
17. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. 2016;2818-26.
18. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. ICLR. 2015;1-14.
19. Kaiming H, Xiangyu Z, Shaoqing R, Jian S. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 770-778.
20. Tajbakhsh N, Gurudu SR, Liang J. Automated Polyp Detection in Colonoscopy Videos Using Shape and Context Information. IEEE Trans Med Imaging. 2016;35(2):630-44. http://dx.doi.org/10.1109/TMI.2015.2487997
21. Bernal J, Sánchez FJ, Fernández-Esparrach G, Gil D, Rodríguez C, Vilariño F. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Comput Med Imaging Graph. 2015;43:99-111. http://dx.doi.org/10.1016/j.compmedimag.2015.02.007
22. Silva J, Histace A, Romain O, Dray X, Granado B. Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer. Int J Comput Assist Radiol Surg. 2014;9(2):283-93. http://dx.doi.org/10.1007/s11548-013-0926-3
23. Freedman JS, Harari DY, Bamji ND, Bodian CA, Kornacki S, Cohen LB, Miller KM, Aisenberg J. The detection of premalignant colon polyps during colonoscopy is stable throughout the workday. Gastrointest Endosc. 2011;73(6):1197-206. http://dx.doi.org/10.1016/j.gie.2011.01.019