



UNIVERSITAT DE
BARCELONA

Inteligencia artificial como herramienta de soporte para la caracterización de pólipos colorrectales en imágenes de luz blanca de alta definición

Ana García-Rodríguez

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

**INTELIGENCIA ARTIFICIAL COMO HERRAMIENTA DE
SOPORTE PARA LA CARACTERIZACIÓN DE PÓLIPOS
COLORRECTALES EN IMÁGENES DE LUZ BLANCA DE
ALTA DEFINICIÓN**

TESIS DOCTORAL

Ana García-Rodríguez

DEPARTAMENTO DE MEDICIINA

DIRIGIDA POR

María Glòria Fernández Esparrach

Jorge Bernal del Nozal

Barcelona, noviembre 2022



**UNIVERSITAT DE
BARCELONA**



UNIVERSITAT DE
BARCELONA

CLÍNIC
BARCELONA
Hospital Universitari

Inteligencia artificial como herramienta de soporte para la
caracterización de pólipos colorrectales en imágenes de luz blanca
de alta definición

Memoria de tesis doctoral presentada por

Ana García-Rodríguez

para optar al grado de doctora por la Universitat de Barcelona

Dirigida por

María Glòria Fernández Esparrach, Departamento de Medicina

**Jorge Bernal del Nozal, Departamento de Ciencias de la Computación, Centro
de Visión por Computador, Universidad Autónoma de Barcelona**

Programa de Doctorado de Medicina e Investigación Traslacional. Facultad de
Medicina y Ciencias de la Salud. Universitat de Barcelona

Barcelona, julio 2022

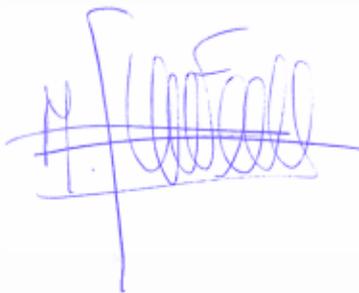
AUTORIZACIÓN DE LA DIRECTORA Y TUTORA DE LA TESIS

LA DRA. MARÍA GLORIA FERNANDEZ ESPARRACH, INVESTIGADORA SENIOR DEL INSTITUT D'INVESTIGACIONS BIOMÈDIQUES AUGUST PI I SUNYER (IDIBAPS) DE BARCELONA,

CERTIFICA:

Que la memoria que lleva por título *"Inteligencia artificial como herramienta de soporte para la caracterización de pólipos colorrectales en imágenes de luz blanca de alta definición"*, presentada por Ana García-Rodríguez para optar al grado de Doctora en Medicina, ha sido realizada bajo mi dirección y tutela. Una vez finalizada autorizo su presentación para ser juzgada por el tribunal correspondiente.

Y para que quede constancia a los efectos oportunos, firmo la presente en 13 de julio de 2022.



Dra. María Glòria Fernández Esparrach

AUTORIZACIÓN DEL DIRECTOR DE LA TESIS

EL DR. JORGE BERNAL DEL NOZAL, PROFESSOR AGREGAT DEL DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN DE LA UNIVERSIDAD AUTÓNOMA DE BARCELONA E INVESTIGADOR ADSCRITO AL CENTRO DE VISIÓN POR COMPUTADOR,

CERTIFICA:

Que la memoria que lleva por título *"Inteligencia artificial como herramienta de soporte para la caracterización de pólipos colorrectales en imágenes de luz blanca de alta definición"*, presentada por Ana García-Rodríguez para optar al grado de Doctora en Medicina, ha sido realizada bajo mi dirección. Una vez finalizada autorizo su presentación para ser juzgada por el tribunal correspondiente.

Y para que quede constancia a los efectos oportunos, firmo la presente en 13 de julio de 2022.



Dr. Jorge Bernal del Nozal

AGRADECIMIENTOS

Gracias en primer lugar a los directores de esta tesis, Glòria Fernández Esparrach y Jorge Bernal del Nozal, por incluirme en vuestro equipo de investigación y haber hecho posible este trabajo.

Glòria, excelente médico y persona. Gracias por contar conmigo, ser fuente de inspiración y aprendizaje continuo. Admiro tu capacidad de trabajo y organización, tu motivación constante, tu adaptabilidad y fortaleza, así como tu tolerancia y respeto. Eres ejemplo a seguir.

Jorge, gracias por tu apoyo y paciencia, por abrirme a tu mundo e instruirme de una forma fácil y divertida. Gracias también al resto de compañeros del Centro de Visión por Computador de la UAB, F. Javier Sánchez, impulsor inicial de este trabajo y Yael Tudela, por su enorme colaboración en la parte más técnica de este estudio.

Gracias a la predecesora de este proyecto, Cristina Sánchez-Montes, por ayudarme a coger las riendas de tu trabajo y a mi cobecario y amigo Rodrigo Garcés, por ser mi confidente dentro y fuera del hospital. A mi compañera de andanzas desde el inicio de nuestra etapa en el Hospital Clínic, mi amiga Ariadna Sánchez, por todo lo que implica ser una coR más.

A toda la Unidad de Endoscopia del Hospital Clínic (médic@s, enfermer@s, auxiliares, secretari@s y anestesiastas), con especial mención a los que habéis formado parte del estudio y habéis colaborado en la recogida de datos: Sabela Carballal, Leticia Moreira, Liseth Rivero, Ingrid Ordás, Oswaldo Ortiz, Álex Fernández-Simón, Eva Vaquero, María Daca, los que habéis participado activamente en mi trayectoria como doctoranda, Henry Córdova, María Pellisé, Àngels Ginès y Oriol Sendino, por vuestro enorme esfuerzo y dedicación docente, a mis compañeros de despacho, Begoña González, Isis Araujo, Cristina de Miguel y Antonio Giordano, por amenizarme las mañanas de ordenador y a los jefes, Josep Llach y Francesc Balaguer, por hacerme parte integrante de la Unidad. También a la Unidad de Anatomía Patológica del H. Clínic, con Miriam Cuatrecasas como principal componente.

A todo el Servicio de Digestivo del Hospital de Viladecans, con alusión particular a mis 2 jefas, Mercè Barenys y Joana Villaverde, por vuestra flexibilidad horaria y ayuda para compaginar ambos trabajos, así como por vuestro apoyo moral de inicio a fin, pero también al resto de componentes, por ser los mejores compañeros y amigos (Martha, Pau, Eva, Luisa, Inês, Carlos, Gerard, Antonia y Ana Belén, enfermería y auxiliares).

Por último, a mi familia, por estar a mi lado todo este tiempo,

Barcelona, julio 2022

FINANCIACIÓN

Los estudios que constituyen la presente tesis doctoral se han realizado bajo la siguiente financiación:

- “Aplicación de nuevas tecnologías en endoscopia para mejorar la detección de pólipos de colon: navegación electromagnética y algoritmos de localización automática”. Fundación de la Sociedad Española de Endoscopia Digestiva. 2015
- “Intelligent in-Vivo Endoscopic Diagnosis and Intervention Support Systems (iVENDIS)”. Ministerio de Ciencia e Innovación. Plan Nacional I+D+I. DPI2015-65286-R. 2015
- “Validación clínica del primer sistema in-vivo de predicción histológica automática de los pólipos de colon integrando la visión del endoscopista con la visión por computador (HISINVIA)”. Ministerio de Economía, Industria y competitividad. Instituto de Salud Carlos III. PI17/00894. 2017
- “Predicció histològica in-vivo dels pòlips colorectals utilitzant un mètode d’intel·ligència artificial basat en llum blanca”. Beca “Iniciació a la Recerca”. Societat Catalana de Digestologia. 2019
- “Controlando el olvido en redes neuronales: aprendizaje continuo a través de la detección de anomalías en secuencias de imágenes”. Ministerio de Ciencia e Innovación. Proyectos de I+D+i. PID2020-120611RB-I00. 2020

ÍNDICE

GLOSARIO	11
ENUMERACIÓN DE LOS ARTÍCULOS QUE COMPONEN LA TESIS.....	15
RESUMEN DE LA TESIS	17
TÍTULO	17
INTRODUCCIÓN	17
HIPÓTESIS	18
OBJETIVOS	18
MÉTODOS	18
PRINCIPALES RESULTADOS	19
CONCLUSIONES.....	19
INTRODUCCIÓN	21
1. ANTECEDENTES	21
1.1. EPIDEMIOLOGIA DEL CÁNCER COLORRECTAL	21
1.2. DEFINICIÓN DE CÁNCER COLORRECTAL	21
2. PÓLIPOS COLORRECTALES	23
2.1. CLASIFICACIÓN HISTOLÓGICA.....	23
2.2. CLASIFICACIÓN MORFOLÓGICA	25
3. SISTEMAS DE IMAGEN ENDOSCÓPICA Y CLASIFICACIONES UTILIZADAS PARA EL DIAGNÓSTICO ÓPTICO	27
4. MANEJO DE LOS PÓLIPOS COLORRECTALES.....	32
5. INTELIGENCIA ARTIFICIAL	36
5.1. CONCEPTOS GENERALES.....	36
5.2. MACHINE LEARNING.....	37
5.2.1. Tipos de Machine Learning.....	38
5.2.2. Etapas de un experimento con Machine Learning	38
5.3. SISTEMAS AUTOMÁTICOS DE CLASIFICACIÓN DE IMÁGENES	40
5.3.1. Bag of Words	40
5.3.2. Deep Learning	41
5.4. INTELIGENCIA ARTIFICIAL PARA ASISTIR A LA COLONOSCOPIA.....	44
5.5. ESTUDIOS DE SISTEMAS CADX YA PUBLICADOS:	44
5.5.1. CADx con NBI (con y sin magnificación)	45
5.5.2. CADx con Blue Laser/Light Imaging	46
5.5.3. CADx con endocitoscopia	46
5.5.4. CADx con otras modalidades ópticas complejas	48
5.5.5. CADx con luz blanca	48
5.6. IMPLEMENTACIÓN DE LA INTELIGENCIA ARTIFICIAL EN LA VIDA REAL	52
5.7. BASES DE DATOS.....	52
5.8. LIMITACIONES DE LOS ESTUDIOS CADX EXISTENTES.....	54
5.9. DISPOSITIVOS INTELIGENTES COMERCIALIZADOS Y REGULACIÓN.....	55
HIPÓTESIS	59
OBJETIVOS	61
MATERIAL Y MÉTODOS Y RESULTADOS.....	63

Artículo 1:	65
Título.....	65
Objetivo	65
Resumen	65
Artículo 2:	79
Título.....	79
Objetivo	79
Resumen	79
Artículo 3:	89
Título.....	89
Objetivo	89
Resumen	89
DISCUSIÓN	95
FUTURAS LÍNEAS DE TRABAJO	103
CONCLUSIONES	105
BIBLIOGRAFÍA	107
DECLARACIÓN DE CÓDIGO ÉTICO Y BUENAS PRÁCTICAS	119

ÍNDICE DE FIGURAS

Figura 1: Profundidad del cáncer y estadificación TNM del AJCC.....	23
Figura 2: Representación de 2 ejemplos de adenomas (imagen endoscópica e histológica correspondiente).....	24
Figura 3: Clasificación de las lesiones serradas.....	25
Figura 4: Clasificación morfológica de Paris (adaptación de la original de Axon).....	26
Figura 5: Modelos de LST granular y no granular, con la clasificación correspondiente según las categorías de la clasificación de París	27
Figura 6: Patrón de criptas de Kudo y tipo de lesión	28
Figura 7: Clasificación de NICE	29
Figura 8: Clasificación JNET	30
Figura 9: Clasificación WASP.	31
Figura 10: Pieza de DSE	32
Figura 11: Algoritmo de tratamiento en función de la probabilidad de invasión profunda	34
Figura 12: Ramas de la IA	37
Figura 13: Gráfico explicativo de los componentes y las posibles aplicaciones de la IA	39
Figura 14: Ejemplo de cómo funciona el BoW.	41
Figura 15: Esquema básico de una RN	42
Figura 16: Flujo de trabajo del enfoque tradicional (a) y en la era DL (b).....	43
Figura 17: Representación gráfica del modelo CADx de Sánchez-Montes et al	49
Figura 18: Posibilidades de interacción entre una IA y endoscopistas	58

ÍNDICE DE TABLAS

Tabla 1: Clasificación de Viena revisada, OMS y TNM.	22
Tabla 2: Recomendaciones de tratamiento endoscópico del Grupo de Resección Endoscópica Español	35
Tabla 3: Resumen estudios CADx publicados.....	51
Tabla 4: Bases de datos públicas disponibles hasta la fecha	53
Tabla 5: Descripción general de los sistemas CAD comercializados.	57

GLOSARIO

AFI = Autofluorescencia

AJCC = American Joint Committee on Cancer

ASGE = American Society of Gastrointestinal Endoscopy

AST = Adenoma serrado tradicional

BoW = Bag of Words

BASIC = BLI Adenoma Serrated International Classification

BLI = Blue Laser/Light Imaging

CAD = Diagnóstico asistido por computador

CADe = Diagnóstico asistido por computador para detección

CADx = Diagnóstico asistido por computador para clasificación

CBIR = Content-based Image Retrieval

CCR = Cáncer colorrectal

CE = Cromoendoscopia

CLE = Endomicroscopía láser confocal

CNN: Convolutional Neural Network

DL = Aprendizaje profundo

DNN = Deep Neural Network

E = Especificidad

EC = Endocitoscopia

ESGE = European Society of Gastrointestinal Endoscopy

FICE = Fuji Intelligent Color Enhancement

GT = Ground Truth

HD = High Definition

HE = Hematoxilina-Eosina

IA = Inteligencia artificial

IC = Intervalo de confianza

JNET = Japanese NBI Expert Team

LB = Luz blanca

LS = Lesión serrada

LSS = Lesión serrada sésil

LST-G/NG = Laterally Spreading Tumor-granular/nongranular

ML = Aprendizaje computacional

NBI = Narrow Band Imaging

NICE = NBI International Colorectal Endoscopic

Nx = número de aumentos

OMS = Organización Mundial de la Salud

PH = Pólipo hiperplásico

PIVI = Preservation and Incorporation of Valuable Endoscopic Innovations

RN = Red neuronal

ROC: Receiver Operating Characteristic

S = Sensibilidad

SD= Standard Definition

SVM = Support Vector Machine

TBIR = Text-based Image Retrieval

VPN = Valor predictivo negativo

VPP = Valor predictivo positivo

WASP = Workgroup serrated polyps and Polyposis

ENUMERACIÓN DE LOS ARTÍCULOS QUE COMPONEN LA TESIS

Tesis en formato de compendio de artículos.

La tesis consta de 3 objetivos y 3 artículos:

- Jorge Bernal, Aymeric Histace, Marc Masana, Quentin Angermann, Cristina Sánchez-Montes, Cristina Rodríguez de Miguel, Maroua Hammami, **Ana García-Rodríguez**, Henry Córdova, Olivier Romain, Gloria Fernández-Esparrach, Xavier Dray, F. Javier Sánchez
«GTCreator: a flexible annotation tool for image-based datasets» Int J Comput Assist Radiol Surg. 2019 Feb 1;14(2):191–201.
 - **Factor de impacto : 3.421**
 - **Cuartil : Q1**
- **Ana García-Rodríguez**, Yael Tudela, Henry Córdova, Sabela Carballal, Ingrid Ordás, Leticia Moreira, Eva Vaquero, Oswaldo Ortiz, Liseth Rivero, F Javier Sánchez, Miriam Cuatrecasas, Maria Pellisé, Jorge Bernal, Glòria Fernández-Esparrach
«In vivo computer-aided diagnosis of colorectal polyps using white light endoscopy» Endosc Int open. 2022
***Revista no indexada en JCR o Scimago Journal Rank**
- **Ana García-Rodríguez**, Jorge Bernal, F. Javier Sánchez, Henry Córdova, Rodrigo Garcés Durán, Cristina Rodríguez de Miguel, Gloria Fernández-Esparrach
«Polyp fingerprint : automatic recognition of colorectal polyps' unique features» Surg Endosc. 2020;34(4):1887–9
 - **Factor de impacto : 3.453**
 - **Cuartil : Q1**

RESUMEN DE LA TESIS

TÍTULO

Inteligencia artificial como herramienta de soporte para la caracterización de pólipos colorrectales en imágenes de luz blanca de alta definición.

INTRODUCCIÓN

El cáncer colorrectal es una de las patologías con mayor prevalencia en la población mundial. La colonoscopia sigue siendo el estándar de referencia para detectar y extirpar las lesiones precursoras, los pólipos, y su calidad depende de la experiencia del endoscopista y de las técnicas y tecnología utilizadas.

La adquisición de datos biomédicos a gran escala y el avance en los modelos de aprendizaje computacional han impulsado el crecimiento de la inteligencia artificial en la medicina. En el campo de la endoscopia y más concretamente de la colonoscopia, la investigación ha ido dirigida a estandarizar las tareas que los endoscopistas realizan a través del desarrollo e integración de sistemas de soporte que, globalmente, buscan mejorar el diagnóstico de los pólipos colorrectales. Sin embargo, todavía existen desafíos en la implementación de dichos dispositivos en la práctica clínica, especialmente en lo que se refiere a clasificación. Esto se debe fundamentalmente a limitaciones en el diseño que dificultan una adecuada integración en el flujo de trabajo clínico, como la incapacidad de trabajar en tiempo real o la necesidad de tecnología adicional como la cromoendoscopia virtual o endocitoscopia y la magnificación como requisito previo para poder funcionar.

Además del diagnóstico óptico, otro aspecto importante y que aún no está resuelto es la identificación inequívoca de los pólipos, de manera que puedan reconocerse en dos instantes temporales distintos. Puesto que los que pertenecen al mismo tipo histológico son muy parecidos entre sí y habitualmente suele haber más de un pólipo por paciente, es necesario evitar errores en la identificación, especialmente cuando deben tomarse decisiones diferidas o los pacientes son remitidos a otro centro más experto para su abordaje.

Por último, existen pocas bases de datos amplias y variadas de imágenes de colonoscopia de luz blanca en alta definición. Para poder construir y validar cualquier sistema inteligente se necesita establecer el patrón oro en la base de datos, así como tenerla correctamente anotada y continuamente actualizada, para lo que se requieren herramientas específicas.

HIPÓTESIS

La inteligencia artificial puede proporcionar apoyo en tiempo real a los endoscopistas mediante el reconocimiento automático de patrones específicos en imágenes de pólipos colorrectales de luz blanca de alta definición que les permita mejorar su diagnóstico óptico y les ayude a su identificación de manera unívoca. La disponibilidad de herramientas que faciliten y uniformicen la anotación de dichas imágenes sería de gran ayuda.

OBJETIVOS

1. Diseñar una herramienta para crear y administrar anotaciones de bases de datos de imágenes de forma cómoda y flexible entre varios usuarios.
2. Desarrollar y evaluar un sistema de clasificación automática de pólipos en tiempo real e in vivo utilizando solo imágenes de luz blanca de alta definición y comparar su rendimiento con el de los endoscopistas.
3. Desarrollar y evaluar un sistema de recuperación de imágenes mediante la extracción de características de pólipos en imágenes de luz blanca de alta definición que permitan su identificación inequívoca (huella digital del pólipo o *polyp fingerprint*) y potencial agrupación en categorías histológicas por analogía.

MÉTODOS

Se han realizado 3 estudios con diferente metodología:

1. GTCreator: herramienta de edición y anotación de imágenes. Estudio comparativo (cualitativo y cuantitativo) con herramientas similares desarrolladas por terceros.

2. ATENEA: sistema de clasificación automática de pólipos de colon basado en redes neuronales e imágenes de luz blanca. Estudio prospectivo de validación en vida real y comparativo con la predicción histológica de varios endoscopistas.

3. Polyp fingerprint: sistema de recuperación de imágenes basado en características de la superficie del pólipo (color, textura y forma). Estudio de validación off-line.

PRINCIPALES RESULTADOS

1. GTCreator permite una navegación rápida y fácil a través de las imágenes para su anotación. Mostró menor tiempo global de anotación de las distintas bases de datos y mayor precisión en la creación de los patrones oro que las otras 6 herramientas con las que se comparó.

2. ATENEA, de los 90 pólipos analizados en el experimento realizado in vivo, predijo correctamente la histología de 63/69 adenomas (91.3%, IC del 95%: 82–97%) y 12/21 no adenomas (57.1%, IC del 95%: 34–78%) mientras los endoscopistas acertaron en 52/69 (75.4%, IC del 95%: 60–85%) y 20/21 (95,2%, IC del 95%: 76–100%), respectivamente. La exactitud diagnóstica global fue del 83,3% (IC del 95%: 74–90%) para ATENEA y del 80% (IC del 95%: 70–88%) para los endoscopistas.

3. Polyp fingerprint hizo coincidir otra imagen del mismo pólipo en 221/243 casos (exactitud diagnóstica 91%). No se observaron diferencias significativas en el número de coincidencias correctas según la clasificación de París (protruidas: 90,7% vs no protruidas: 91,3%) ni el tamaño (<10 mm: 91,6% vs >10 mm: 90%).

CONCLUSIONES

1. GTCreator permite la creación y manipulación de anotaciones de forma rápida y precisa entre varios profesionales.

2. ATENEA obtiene una predicción precisa de la histología de los pólipos in vivo utilizando solo endoscopia de luz blanca y ofrece un rendimiento similar a los endoscopistas a pesar de clasificar peor los no adenomas.

3. Polyp fingerprint permite reconocer diferentes instantáneas de un mismo pólipo, lo cual confirma que el contenido de la región del pólipo es clave para el reconocimiento inequívoco y eventual clasificación histológica de los mismos.

INTRODUCCIÓN

1. ANTECEDENTES

1.1. EPIDEMIOLOGIA DEL CÁNCER COLORRECTAL

El cáncer colorrectal (CCR) es el tercer cáncer más común en ambos sexos y la segunda causa de muerte en el mundo(1). En España es la segunda causa más frecuente de cáncer, en hombres seguido del cáncer de próstata y en mujeres del de mama(2). En la mayoría de los países europeos la incidencia de CCR en la última década presenta una tendencia al alza. El incremento anual ha oscilado entre el 0,4% y el 3,6%(3). Se ha visto también un aumento de la incidencia en adultos jóvenes, de 20 a 49 años (el mayor aumento en sujetos de 20 a 39 años)(4). Se desconoce la causa exacta de dicha tendencia y podría estar relacionada con el aumento de la prevalencia de la obesidad, falta de ejercicio y factores dietéticos como el alcohol y la carne procesada(5). También podría deberse en parte a síndromes hereditarios, pero la mayoría de los CCR en gente joven son esporádicos(6). La mortalidad sin embargo presenta una tendencia a la baja. El estudio EUROCARE-5 sitúa la supervivencia media en Europa del CCR a 5 años del 57%(7). La colonoscopia juega un papel clave en cualquiera de las estrategias dirigidas a reducir la incidencia y mortalidad por CCR, ya que permite la detección y resección de las lesiones precursoras, los pólipos(8).

1.2. DEFINICIÓN DE CÁNCER COLORRECTAL

Casi todos los CCR se originan a partir de un pólipo. La prevalencia de cáncer en los pólipos colorrectales oscila entre el 0,2% y el 5%(9,10). La secuencia adenoma-carcinoma es el proceso por el que aparecen la mayoría de los CCR. En este sentido la **Organización Mundial de la Salud** (OMS) define el adenocarcinoma colorrectal o simplemente CCR como: “ la invasión de células neoplásicas a través de la *muscularis mucosae* en la submucosa”(11,12).

El sistema de **clasificación de Viena** es un consenso entre patólogos occidentales y japoneses que divide las neoplasias epiteliales del tracto gastrointestinal en 5 categorías(13). Según esta clasificación el CCR se incluiría en la categoría 5. En la 8ª edición del **sistema de estadificación TNM** del American Joint Committee on Cancer

(AJCC)(14) estas lesiones son denominadas pT1 o “precoces” y se distinguen de las que invaden más allá de la submucosa (pT2-4), debido a que el grado de infiltración en profundidad tiene una serie de connotaciones en el tratamiento y pronóstico de la enfermedad. La detección de tumores en estadio T1 ha aumentado significativamente desde la implementación de los programas de cribado(15) y el abanico de posibilidades terapéuticas es cada vez mayor. En la siguiente tabla se muestra la clasificación de Viena revisada para las neoplasias epiteliales colorrectales(16) junto con la de la OMS y la estadificación TNM (tabla 1).

1. No neoplasia – Categoría Viena 1 (negativo para neoplasia)
2. Neoplasia mucosa de bajo grado
- Categoría Viena 3 (neoplasia no invasiva de bajo grado, adenoma de bajo grado, displasia de bajo grado)
- Otras terminologías comunes: displasia leve o moderada
- OMS: neoplasia intraepitelial de bajo grado
3. Neoplasia mucosa de alto grado
- Categoría Viena 4.1-4.4 (neoplasia no invasiva de alto grado, adenoma de alto grado, displasia de alto grado)
- Carcinoma no invasivo (carcinoma in situ)
- Carcinoma intramucoso
- Otra terminología común: displasia de alto grado, displasia severa, neoplasia intraepitelial de alto grado
- OMS: neoplasia intraepitelial de alto grado
- TNM: pTis
4. Carcinoma que invade la submucosa o más en profundidad
- 4a. Carcinoma limitado a la submucosa
- Categoría Viena 5 (invasión submucosa del carcinoma)
- TNM: pT1
- 4b. Carcinoma más allá de la submucosa
- TNM: pT2-T4

Tabla 1: Clasificación de Viena revisada, OMS y TNM. No se recomienda emplear la categoría 2 de la Clasificación de Viena (indefinido para neoplasia).

Esta definición de CCR excluye lesiones con displasia de alto grado, en las que los cambios displásicos se limitan únicamente al epitelio, la lámina propia o *muscularis mucosae*. Los patólogos usan también el término carcinoma *in situ*, carcinoma intramucoso o neoplasia epitelial para referirse a este estadio tumoral, que no debe confundirse con el carcinoma invasivo. Estas lesiones son clasificadas como pTis según el sistema de estadificación TNM (tabla 1). En la siguiente imagen (figura 1) pueden verse representados los diferentes estadios tumorales según la clasificación TNM y la profundidad del cáncer.

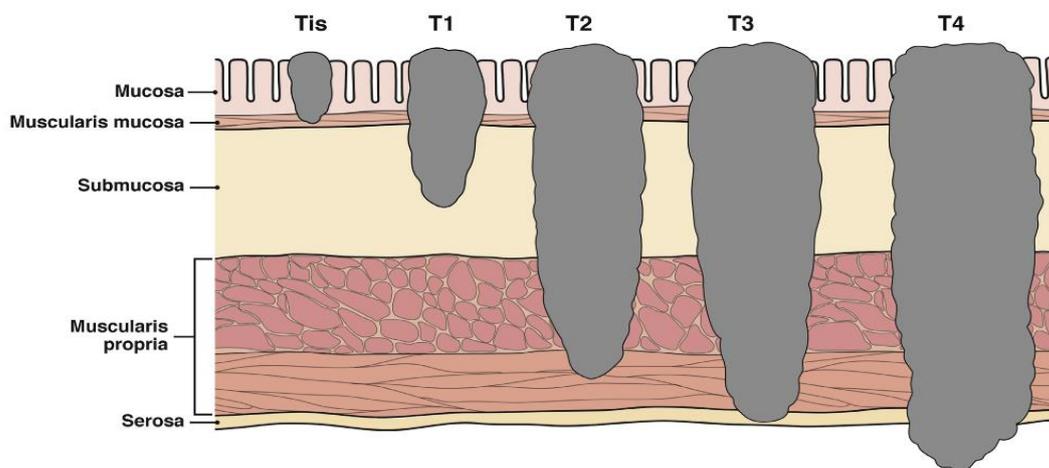


Figura 1: Profundidad del cáncer y estadificación TNM del AJCC (extraída de la revisión de Shaukat et al de la *US Multi-Society Task Force on Colorectal Cancer* (17))

2. PÓLIPOS COLORRECTALES

2.1. CLASIFICACIÓN HISTOLÓGICA

Los pólipos se clasifican en neoplásicos y no neoplásicos según tengan o no riesgo de degeneración. Dentro de los primeros se encuentran los adenomas y las lesiones serradas (LS) -concretamente las lesiones serradas sésiles (LSS) y los adenomas serrados tradicionales (AST)-. A pesar de que la evidencia científica disponible en relación a las lesiones neoplásicas colorectales está basada en los adenomas como lesión precursora del CCR, se estima que un 15-30% de todos los CCR se originan sobre LS(18,19). Los pólipos neoplásicos, a su vez, pueden presentar diferentes grados de displasia (bajo y alto).

Los **adenomas** constituyen una proliferación no invasiva de células epiteliales. El 85% son tubulares (< 20% de componente velloso), 10% tubulovellosos y 5% vellosos (> 80% de componente velloso). Los tubulares se caracterizan por una red compleja de glándulas ramificadas. En los vellosos, las glándulas se extienden directamente desde la superficie hasta el centro del pólipo, creando de esta manera proyecciones digitiformes largas. Los tubulovellosos se presentan como una combinación de estos dos tipos histológicos (figura 2).

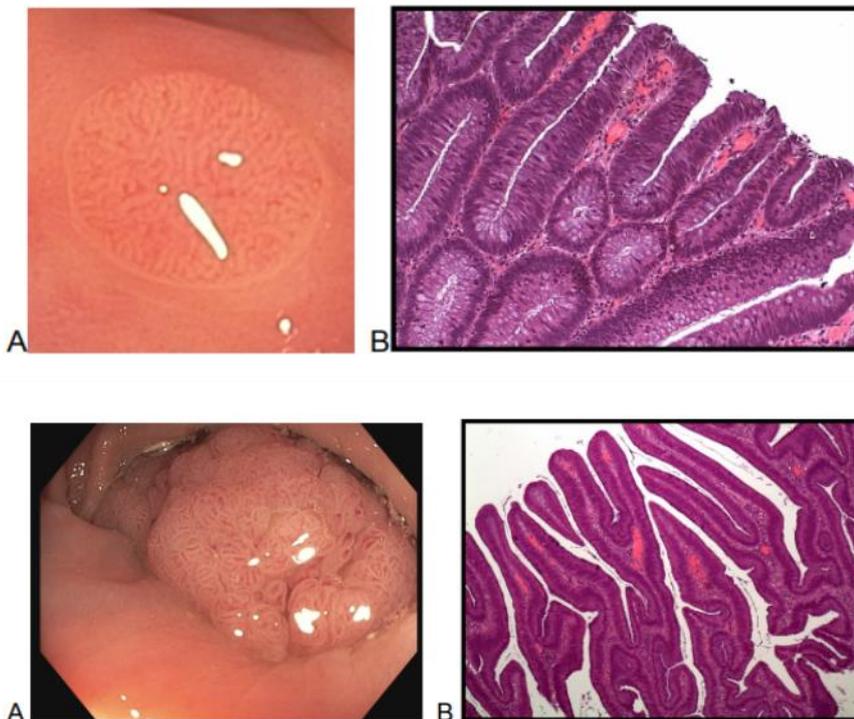


Figura 2: Representación de 2 ejemplos de adenomas. Primera fila. (A) Imagen endoscópica de un adenoma tubular, rico en vasos que rodean a criptas alargadas y ocasionalmente también curvadas. (B) Imagen histológica correspondiente con tinción de Hematoxilina-Eosina (HE) a 200 aumentos (HE 200x). Segunda fila. (A) Imagen endoscópica de un adenoma tubulovellosos, con patrón de criptas cerebroide, frecuentemente ramificadas y largas. (B) Imagen histológica correspondiente (HE 100x).

Las **LS** son un grupo heterogéneo de lesiones con características endoscópicas e histológicas que permiten diferenciarlas de los adenomas (figura 3). Comparten como característica común la presencia de una arquitectura en “dientes de sierra” en el epitelio de la cripta. La OMS propone asignar el término LS a los pólipos hiperplásicos (PH), las LSS y los AST. Los PH ocupan aproximadamente el 75% de todas las LS mientras

que las LSS (anteriormente llamados adenomas/pólipos serrados sésiles) representan aproximadamente el 25%. En general, las LSS se caracterizan por un tamaño mayor, ubicación en el colon proximal y una apariencia endoscópica distinta en comparación con los PH (ver figura 3). La progresión de LSS a LSS con displasia es infrecuente - aproximadamente entre el 4-8% de las LSS contienen displasia(20)-. Los AST son el tipo menos común de LS; son típicamente polipoides y se encuentran en el colon distal y recto. De acuerdo con los criterios actualizados de la OMS, la presencia de una única cripta inequívocamente distorsionada se considera diagnóstica de LSS(21).

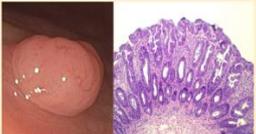
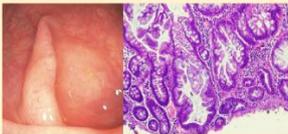
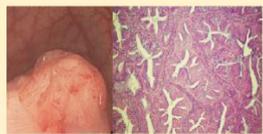
	PÓLIPOS HIPERPLÁSICOS	LESIONES SERRADAS SESILES	ADENOMAS SERRADOS TRADICIONALES
Características clínicas	Prevalencia: 20%-30% Tamaño: Pequeño o diminuto (≤5mm) - Morfología: Plana o sésil	Prevalencia: 5%-15% Tamaño: Mayor que los HP, 5-7mm Morfología: Plana (45%) o sésil	Prevalencia: <1% Tamaño: Mayor que las LSS Morfología: Pediculada, similar a un adenoma
Localización	 70-80% distal	 70-90% proximal	 Fundamentalmente distal
Características endoscópicas	Luz blanca: Pálido o del mismo color que la mucosa circundante. Forma redonda u ovalada. Se aplanan con la insuflación. Vasos finos o ausentes, lacys vessels. NBI: NICE tipo 1. Manchas uniformes oscuras o blancas	Luz blanca: Capa de moco. Superficie similar a una nube. Forma irregular. Anillo de detritus. NBI: NICE tipo 1. Manchas oscuras en las criptas.	Luz blanca: Eritematoso. Polilobulado. Apariencia 'pine cone'. Patrón de criptas tipo IV-S NBI: no está bien definido
Características histológicas	Glándulas con morfología en "diente de sierra" limitada a la superficie de la cripta, sin displasia ni potencial de transformación maligna. Tres subtipos: microvesicular, rico en células caliciformes y con depleción mucinosa	Toda la cripta está dilatada, la morfología en "diente de sierra" se extiende a su base y puede tener displasia citológica de bajo y alto grado	Dientes de sierra prominentes, citoplasma eosinófilo, displasia citológica similar al adenoma y posible morfología vellosa.
Imagen endoscópica con luz blanca e histológica			

Figura 3: Clasificación de las lesiones serradas. Características clínicas, endoscópicas e histológicas y localización con imagen endoscópica e histológica de ejemplo.

2.2. CLASIFICACIÓN MORFOLÓGICA

Las lesiones neoplásicas del tracto digestivo se dividen en precoces o superficiales y avanzadas, siendo el límite la submucosa. Las "superficiales" se denominan también grado 0, al igual que la clasificación para lesiones precoces gástricas de los japoneses publicada el siglo pasado(22), para poder diferenciarlas de las avanzadas o grados 1-5 según la terminología propuesta por Borrmann en 1926.

Con el objetivo de aunar una terminología común entre occidente y Japón para este tipo de lesiones surge la **clasificación de París** (2002) y París modificada (2005)(23,24). En base a esta, los pólipos de colon se clasifican según su forma en lesiones protruyentes [tipo 0-I, que a su vez pueden ser **pediculadas** (0-Ip) o **sésiles** (0-Is)], **planas** [tipo 0-II, que se subdividen en plano-elevadas (0-IIa), planas (0-IIb) y plano-deprimidas (0-IIc)] y **deprimidas** (tipo 0-III), figura 4.

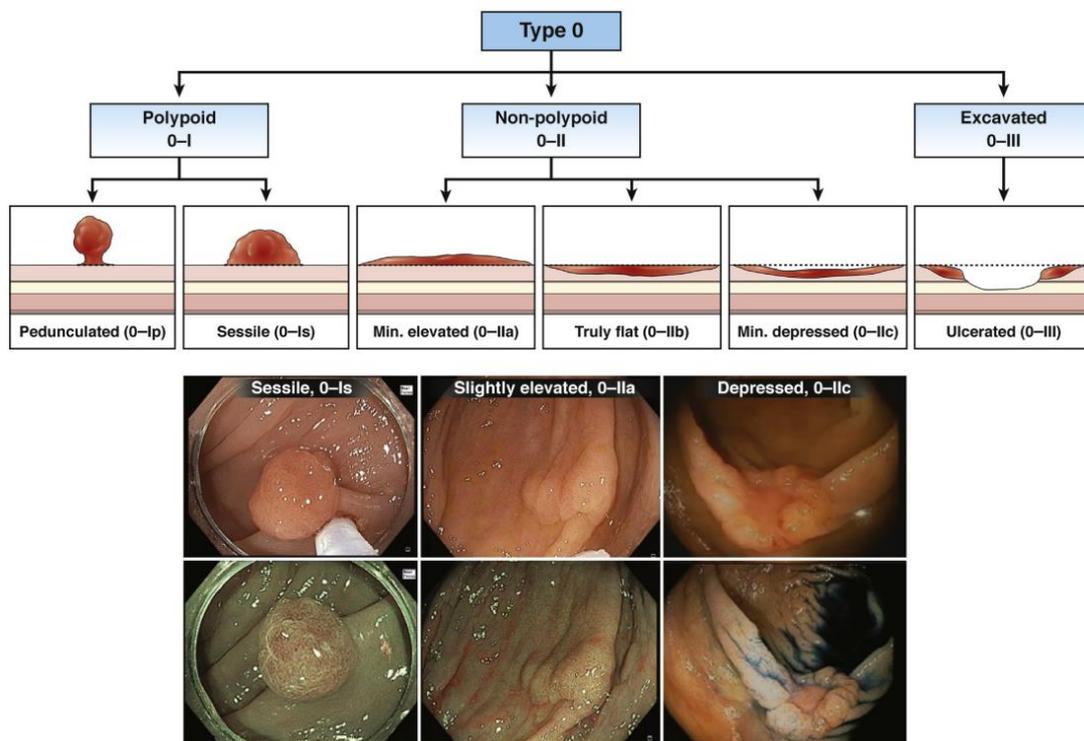


Figura 4: Clasificación morfológica de París (adaptación de la original de Axon)

Los pólipos planos de más de 10 mm son conocidos como **lesiones de extensión lateral** (LST o laterally spreading tumor) y pueden ser **granulares** (LST-G) o **no granulares** (LST-NG)(25). Posteriormente, Kudo et al subdividieron aún más las LST: el tipo granular en los tipos homogéneas (0-IIa) y nodulares mixtas (0-IIa+Is/0-Is+0-IIa) y el tipo no granular en los tipos plano-elevadas (0-IIa) y pseudodeprimidas (0-IIc, combinado o no con 0-IIa o 0-IIb)(26), como puede verse representado en la figura 5:

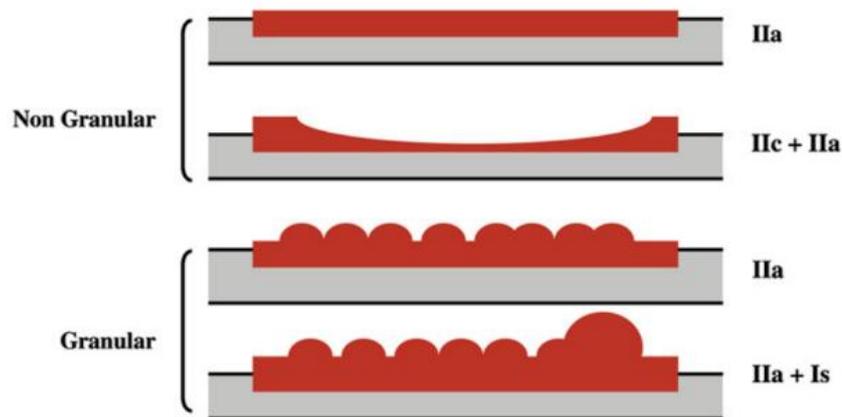


Figura 5: Modelos de LST granular y no granular, con la clasificación correspondiente según las categorías de la clasificación de París

Aunque la concordancia interobservador de la clasificación de París entre endoscopistas expertos es modesta(27), se recomienda su empleo y la de las LST para la caracterización morfológica.

3. SISTEMAS DE IMAGEN ENDOSCÓPICA Y CLASIFICACIONES UTILIZADAS PARA EL DIAGNÓSTICO ÓPTICO

Con el tiempo se han perfeccionado los sistemas de imagen endoscópica y desarrollado técnicas auxiliares de realce que pretenden mejorar, por un lado, la detección de pólipos y por el otro, la caracterización, es decir, la capacidad de predecir su histología sin necesidad de esperar al procesamiento de la muestra. Este último concepto se denomina “biopsia óptica”.

La cromoendoscopia (CE) clásica consiste en la aplicación tópica de colorantes (generalmente índigo carmín, azul de metileno o violeta cristal -en Oriente comercializado como violeta de genciana-) para delimitar y mejorar la caracterización de lesiones, al resaltar la arquitectura de la superficie mucosa. La **clasificación de Kudo** (figura 6) permite la evaluación de pólipos mediante el estudio del patrón de criptas glandulares. Utiliza colorantes y magnificación para la identificación de los diferentes tipos de lesiones (I: normal, II: PH o LSS -éstas muestran un patrón en estrella u ovalado, denominado II-O, de “open” en inglés, que pueden alternar o progresar a un patrón de

superficie adenomatoso tipo IV o invasivo tipo V-), IIIs y IIIIL: adenomatoso, IV: adenomatoso tipo tubulovelloso, V: invasivo). El tipo V se subclasifica en Vn (no estructural) y Vi (irregular)(28). La clasificación de Kudo con violeta cristal es el estándar de oro para la evaluación del patrón invasivo(29).

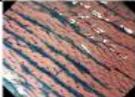
Patrón de criptas	Esquema	Imagen endoscópica	Descripción de las criptas	Histología más probable
I			Redondas, regulares	Mucosa normal
II			Más grandes, regulares, de morfología asteroide	Pólipo hiperplásico
II-O			Similar a II, más anchas y redondas	Lesión serrada sesil
IIIIL			Más alargadas	Adenoma
IV			Forman circunvoluciones	Adenoma
IIIs			Más pequeñas, tubulares o redondeadas	Cáncer precoz
Vi			Irregulares con tipos IIIs, IIIIL, IV	Cáncer intramucoso o SM1
Vn			No reconocibles	Cáncer con invasión submucosa >SM1

Figura 6: Patrón de criptas de Kudo y tipo de lesión

La CE clásica o convencional ha sido sustituida en los últimos años por el uso de técnicas “virtuales”, en las que el efecto de “coloración” se logra mediante el uso de filtros o fuentes de luz en el propio endoscopio (p. ej., el sistema de banda estrecha o Narrow Band Imaging (NBI) de la casa Olympus® y el Blue Light Imaging (BLI) de Fujifilm®) o mediante técnicas de procesamiento posterior de imagen (p. ej., FICE de la casa Fujifilm®

y el sistema i-Scan de la casa Pentax®). La CE virtual resalta los vasos (que se muestran en marrón) y permite la observación indirecta del patrón de la superficie (blanco) y su clasificación como regular, irregular o ausente.

Los siguientes sistemas de clasificación asocian las características de la superficie visualizadas con estas técnicas de realce de imagen con una histología específica y facilitan el diagnóstico óptico por parte del endoscopista:

La **clasificación de NICE** (NBI International Colorectal Endoscopic classification) surge en 2009 y está diseñada para utilizarse con endoscopios de alta definición que dispongan **NBI**, con o sin magnificación. La clasificación NICE unifica diferentes criterios y clasificaciones previamente utilizadas y simplifica las categorías diagnósticas en las tres clínicamente más relevantes: NICE **tipo 1** (PH o LSS), **tipo 2** (adenoma o adenocarcinoma superficial) y **tipo 3** (adenocarcinoma con invasión submucosa profunda)(30), figura 7. Estudios de validación internacional de la clasificación de NICE también han demostrado su utilidad en el diagnóstico óptico incluso sin magnificación(31).

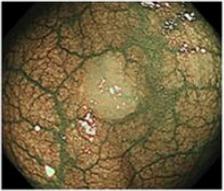
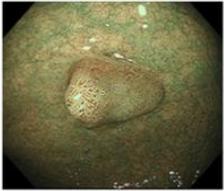
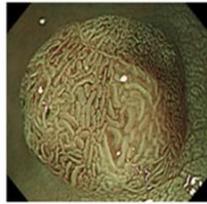
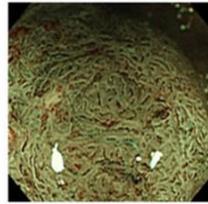
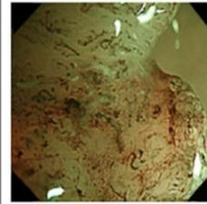
	Type 1	Type 2	Type 3
Color	Same or lighter than background	Browner relative to background (verify color arises from vessels)	Brown to dark brown relative to background; sometimes patchy whiter areas
Vessels	None, or isolated lacy vessels may be present coursing across the lesion	Brown vessels surrounding white structures	Has area(s) of disrupted or missing vessels
Surface pattern	Dark or white spots of uniform size, or homogeneous absence of pattern	Oval, tubular or branched white structures surrounded by brown vessels	Amorphous or absent surface pattern
Most likely pathology	Hyperplastic & sessile serrated polyp (SSP)	Adenoma	Deep submucosal invasive cancer
Endoscopic image			

Figura 7: Clasificación de NICE basada en el color, la vascularización y el patrón de superficie. Se muestra la histología más probable y una imagen endoscópica de ejemplo.

Una limitación de la clasificación de NICE es que no permite diferenciar entre displasia de bajo y alto grado e invasión submucosa superficial en lesiones de tipo 2. Para abordar

esta limitación, el grupo japonés de expertos en NBI (Japanese NBI Expert Team or JNET) publicó en 2014 una nueva clasificación para lesiones colorrectales (**clasificación JNET**), que sí que utiliza **magnificación** y proporciona una mayor precisión a la hora de diagnosticar la profundidad de invasión. JNET consta de 4 categorías: mantiene los tipos 1 y 3 de NICE, pero divide el tipo 2 de NICE en **JNET 2A**, con características asociadas a displasia de bajo grado y **2B** con características asociadas a displasia de alto grado e invasión submucosa superficial(32), figura 8.

	Type 1	Type 2A	Type 2B	Type 3
Vessel pattern	· Invisible*1	· Regular caliber · Regular distribution (meshed/spiral pattern) *2	· Variable caliber · Irregular distribution	· Loose vessel areas · Interruption of thick vessels
Surface pattern	· Regular dark or white spots · Similar to surrounding normal mucosa	· Regular (tubular/branched/papillary)	· Irregular or obscure	· Amorphous areas
Most likely histology	Hyperplastic polyp/ Sessile serrated polyp	Low grade intramucosal neoplasia	High grade intramucosal neoplasia/ Shallow submucosal invasive cancer *3	Deep submucosal invasive cancer
Endoscopic image				

*1. If visible, the caliber in the lesion is similar to surrounding normal mucosa.

*2. Micro-vessels are often distributed in a punctate pattern and well-ordered reticular or spiral vessels may not be observed in depressed lesions.

*3. Deep submucosal invasive cancer may be included.

Figura 8: Clasificación JNET basada en el patrón vascular de superficie. Se muestra la histología más probable y una imagen endoscópica de ejemplo.

Otros sistemas de clasificación endoscópica con magnificación y NBI son la clasificaciones de Hiroshima(33), Sano(34), Showa(35) y Jikei(36) (básicamente limitadas a países asiáticos).

Las LS no están incorporadas en la clasificación de NICE ni JNET. La **clasificación WASP** (Workgroup serrated polyps and Polyposis) combina la clasificación NICE y cuatro características morfológicas de las LS: superficie en forma de nube, borde mal definido, forma irregular y puntos negros dentro de las criptas(37) (figura 9). Al igual que los PH,

las LSS se consideran dentro del tipo 1 de la clasificación de NICE. Sin embargo, la presencia de al menos 2 de las características mencionadas se considera suficiente para diagnosticar una LSS y así poder diferenciarlas de los PH.

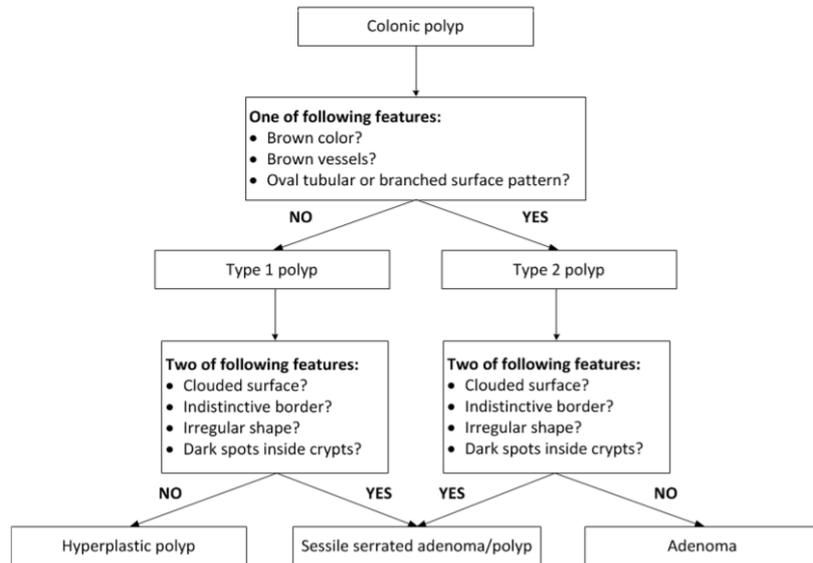


Figura 9: Clasificación WASP: combina la clasificación NICE y 4 características de las LSS: superficie en forma de nube, borde mal definido, forma irregular y puntos negros dentro de las criptas. La presencia de al menos 2 características es suficiente para diagnosticar una LSS.

La **clasificación BASIC** (BLI Adenoma Serrated International Classification)(38), creada para la caracterización de pólipos subcentimétricos con **BLI**, incluye características de superficie, patrón de criptas y vascularización y mejora la concordancia interobservador si se utiliza con magnificación.

Existen otros sistemas endoscópicos más complejos como la autofluorescencia (AFI) y técnicas microscópicas para la evaluación histológica in vivo como la endomicroscopía láser confocal (CLE por sus siglas en inglés) y la endocitoscopia, pero están menos extendidas y su uso depende del fabricante específico de dispositivos endoscópicos, además de estar pensadas para endoscopistas altamente especializados, mayoritariamente de Oriente.

4. MANEJO DE LOS PÓLIPOS COLORRECTALES

Como hemos visto en el apartado anterior, la apariencia endoscópica de un pólipo puede servir para predecir el tipo histológico y el grado de invasión submucosa, que a su vez se asocia con el riesgo de metástasis ganglionar(39–41), uno de los principales factores pronósticos asociados con recaída local y/o a distancia en el CCR.

Las lesiones colorrectales sin sospecha de componente invasivo, por su nula posibilidad de presentar metástasis ganglionar, pueden ser tratadas mediante técnicas de resección local (endoscópica o quirúrgica). Por consiguiente, al no ser necesario satisfacer los criterios de resección oncológicos, no es imprescindible realizar una resección en bloque de la lesión. En cambio, cuando hay sospecha de invasión submucosa, es imprescindible la extirpación en bloque para poder asegurar márgenes libres y una adecuada evaluación histológica de factores predictivos de metástasis ganglionar, que podrían determinar la necesidad de una cirugía ulterior. En estos casos, debido a su capacidad para proporcionar muestras de alta calidad, suelen requerirse técnicas de resección más complejas como la disección submucosa endoscópica (DSE) -figura 10-, la resección endoscópica transmural (endoscopic full-thickness resection o eFTR) y la cirugía mínimamente invasiva transanal.

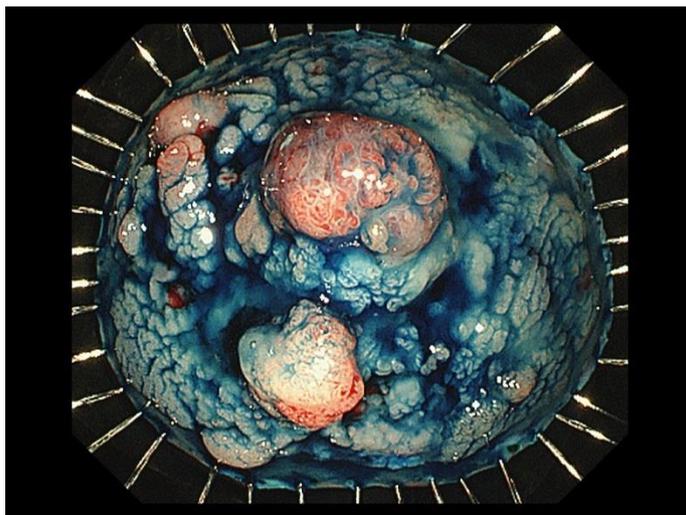


Figura 10: Pieza de DSE. Siempre que sea posible, el espécimen resecado debe enviarse al patólogo estirado y fijado a una superficie firme de material poroso mediante alfileres, estando toda la pieza sumergida en formalina.

Así pues, la indicación y posterior selección de la técnica de resección vienen condicionadas por la presencia o ausencia de signos de invasión submucosa y su profundidad. Para ello, es crucial tener en cuenta el tamaño, la localización y la morfología de la lesión, así como realizar una adecuada caracterización histológica de la misma.

Existe una relación lineal entre el **tamaño** y el riesgo de contener cáncer(42) (se ha descrito que puede llegar al 16.5% en pólipos de ≥ 30 mm(43)). La **localización** distal, principalmente el recto, también se ha asociado a carcinoma invasivo en diversos estudios(42,44,45).

En cuanto a la **morfología**, más del 80% de las LST colorrectales son de tipo LST-G homogénea y mixta. Solo un 15-20% se correspondería a lesiones LST-NG. Por orden de probabilidad, las LST-G nodulares mixtas y las LST-NG, sobre todo si asocian componente pseudodeprimido, tienen mayor riesgo de componente invasivo(46,47). Las LST-G nodulares mixtas de ≥ 4 cm (o también denominadas tipo 'bulky') y de localización rectal se consideran lesiones de alto riesgo ya que pueden ocultar cáncer invasivo con una probabilidad muy elevada(42,45,48).

Por último, el análisis del **patrón mucoso (criptas) y vascular** mediante el uso de CE (con o sin magnificación), puede ayudar a identificar la existencia y profundidad de invasión de la submucosa. Las recientes guías de la Sociedad Británica de Gastroenterología, la ESGE y la AGA sugieren utilizar la clasificación NICE y/o Kudo para evaluar la invasión de la submucosa, ya que han mostrado una buena concordancia interobservador y se adaptan fácilmente a la práctica clínica(17,49,50).

Otros signos más groseros que también pueden orientar a la existencia de componente invasivo son: la presencia de zonas induradas, convergencia de pliegues, úlceras, friabilidad, neovascularización, aspecto en columna, zonas de retracción, "piel de pollo" alrededor de la lesión y áreas deprimidas(51). El signo de no elevación (*non-lifting sign*) también puede deberse a invasión submucosa profunda en lesiones sin manipulación endoscópica previa(52).

Con el objetivo de poder estratificar el riesgo de invasión “encubierta” y predecir la profundidad, el grupo español liderado por Ignasi Puig et al ha creado un árbol de decisión mediante la evaluación de tres características (patrón NICE, presencia de depresión/úlceras y morfología)(46), que además se adapta al entorno occidental, donde pocos endoscopistas hacen uso de la magnificación, figura 11:

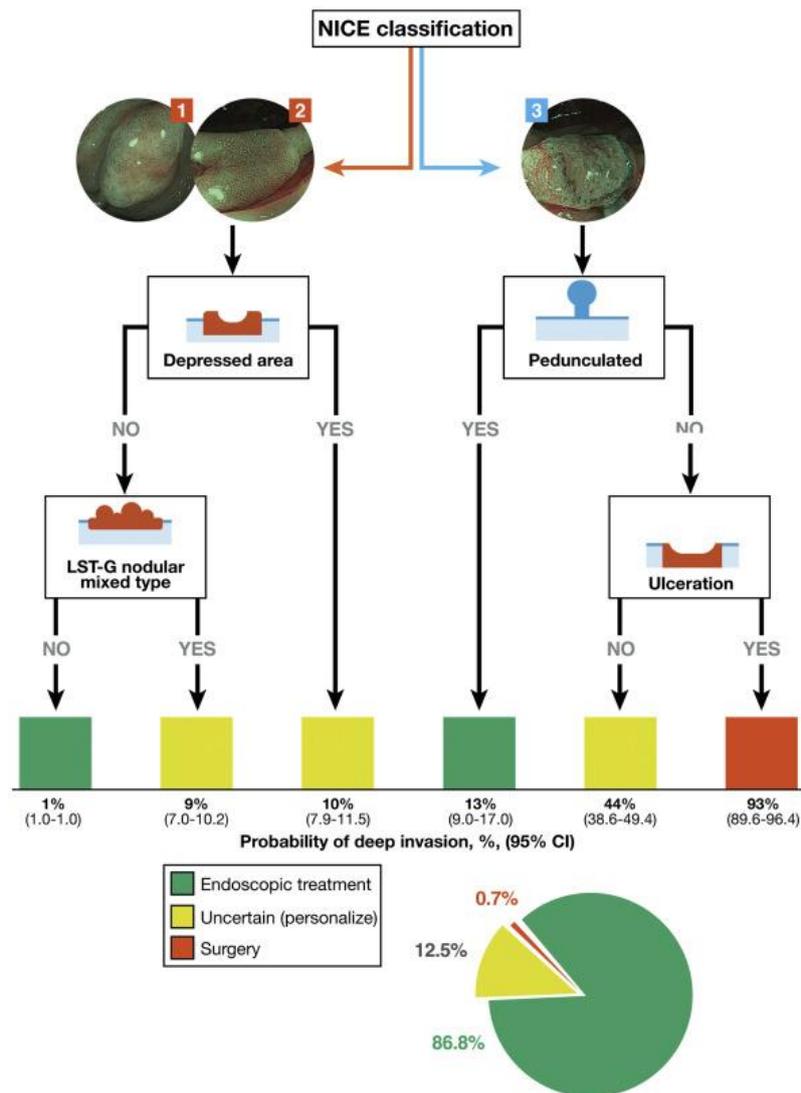


Figura 11: Algoritmo de tratamiento en función de la probabilidad de invasión profunda

Según este esquema, la CE sin magnificación sería útil para predecir la invasión submucosa profunda cuando una lesión NICE tipo 3 no pediculada está ulcerada y para descartarla cuando una lesión NICE tipo 1 o 2 no tiene áreas deprimidas ni nódulos(53).

La magnificación es especialmente valiosa en el escenario opuesto. En el caso de hacer uso de esta última, en lesiones JNET 2B, si existe un área de demarcación debería evaluarse el patrón de criptas de Kudo con violeta cristal(29). En lesiones JNET 2B sin Kudo invasivo o en evaluaciones de baja confianza, el uso de una técnica endoscópica que permita una escisión local completa, podría ser la primera opción de tratamiento.

En las lesiones pediculadas, puesto que un nivel más profundo de invasión submucosa podría correlacionarse con características histológicas favorables(54) y que la precisión del diagnóstico óptico es baja, parece razonable recurrir al tratamiento endoscópico, siempre que la resección sea en bloque y con un buen margen y recomendar una cirugía basada en la histología.

En la siguiente tabla se muestran las recomendaciones de tratamiento establecidas por consenso por el Grupo de Resección Endoscópica Español(55) para el manejo de las neoplasias precoces colorrectales en función del subtipo morfológico y tamaño (tabla 2).

Subtipo morfológico	Estrategia	
	Tamaño ≤ 20-30 mm	Tamaño > 20-30 mm
LST-G homogéneo	RME	RME (DSE en los pocos casos de gran extensión y sospecha invasión SM profunda, preferiblemente en centros con experiencia)
LST-G nodular mixto	RME	Individualizar casos (RME/DSE: si RME, resecar/analizar nódulos grandes por separado)
LST-NG plano elevado	RME	Individualizar casos (RME/DSE)
LST-NG pseudodeprimido	RME	Lesión sugestiva de serrada: la RME-F es en general la técnica de elección
0Is	Polipectomía/RME	DSE
<i>Non-lifting sign</i>	Individualizar casos (RME, RME híbrida, RME bajo agua, RME + avulsión o APC, DSE, cirugía)	
0Ip	Polipectomía	Polipectomía/RME (pedículos grandes o lesiones semipediculadas)
Carcinoides rectales	< 10 mm: RME (asistida por capuchón, híbrida) 10-20 mm: DSE	> 20 mm: cirugía
Pólipos menores de 20 mm:	< 3 mm: pinza de biopsia, asa fría, asa caliente 4-10 mm: asa fría, asa caliente, RME > 10 mm: asa caliente/RME	

Tabla 2: Recomendaciones de tratamiento endoscópico del Grupo de Resección Endoscópica Español

A pesar de todo lo anteriormente comentado con respecto al manejo de pólipos, la gran mayoría de lesiones detectadas en una colonoscopia son diminutas (tamaño ≤5 mm) y

no precisan técnicas de resección endoscópica avanzada. Puesto que la probabilidad de detectar un cáncer en este tipo de lesiones es insignificante(56,57), se ha hecho un enorme esfuerzo por intentar racionalizar los recursos y se han propuesto una serie de estrategias basadas en el diagnóstico óptico para evitar por un lado, resecciones innecesarias (**'leave in situ'** de lesiones no adenomatosas en recto-sigma) y por el otro, su análisis histológico (**'resect and discard'**)(58). En este sentido, se reducirían los riesgos asociados a la resección endoscópica y los costes asociados al diagnóstico anatomopatológico(59).

La Sociedad Americana de Endoscopia Gastrointestinal (ASGE), a través de un comité creado para la evaluación de nuevas tecnologías (PIVI), estableció los requisitos mínimos para poder implementar estas estrategias(58), los cuales han sido posteriormente respaldados por otras sociedades(60–62). Sin embargo, a pesar del buen rendimiento inicial en manos expertas, la precisión del diagnóstico óptico se ha visto reducida sensiblemente en ámbitos no académicos y con endoscopistas poco experimentados(63,64). El desarrollo de sistemas inteligentes para la caracterización de pólipos y la reciente publicación de umbrales alternativos y más fáciles de alcanzar(65) se espera que consigan superar las limitaciones para su aplicación en la práctica clínica diaria.

5. INTELIGENCIA ARTIFICIAL

5.1. CONCEPTOS GENERALES

No existe una definición consensuada de inteligencia artificial (IA). A grandes rasgos, es la inteligencia llevada a cabo por máquinas, es decir, sistemas que intentan hacer y pensar de manera similar a los humanos. John McCarty, Marvin Misky y Claude Shannon, los padres de la IA moderna, acuñaron por primera vez el término durante la conferencia de Darmouth en 1956, como: “la ciencia e ingenio de hacer máquinas inteligentes, especialmente programas de cálculo inteligente”.

Existen **subcampos** dentro de la IA que se centran en ciertas partes del sistema cognitivo humano y sus comportamientos. P. ej., la **robótica** imita los movimientos de los humanos, la **visión artificial** se encarga de la percepción del entorno, los **sistemas**

expertos se ocupan de la toma de decisiones y el reconocimiento del habla y el **procesamiento del lenguaje natural** analiza y trata la información recibida a través del lenguaje humano para posibilitar la comunicación. Sin embargo, el subcampo más revolucionario en las últimas décadas es el **aprendizaje computacional o Machine Learning (ML)**, encargado del reconocimiento de patrones y del aprendizaje (figura 12).

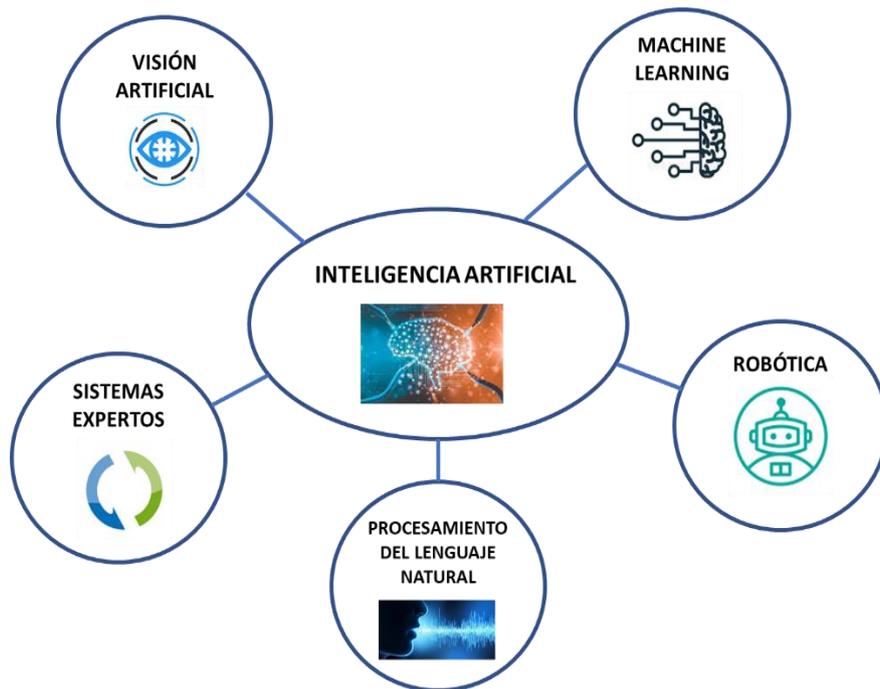


Figura 12: Ramas de la IA

5.2. MACHINE LEARNING

El ML es un área capital dentro de la IA, ya que se relaciona y conecta con el resto de las disciplinas. El ML engloba un conjunto de técnicas que permiten que las máquinas aprendan de manera automatizada, sin instrucciones explícitas de un humano, basándose en patrones e inferencias. Arthur Samuel, pionero en IA, definió el ML como *“el campo de estudio que da a las computadoras la habilidad de aprender sobre algo para lo que no han sido explícitamente programadas”*(66). En las últimas décadas se ha convertido en una herramienta muy común en prácticamente todas las tareas que requieren extraer información a partir de grandes cantidades de datos(67).

5.2.1. Tipos de Machine Learning

Existen diferentes tipos de ML, en función de si se dispone o no de la salida esperada o **Ground Truth (GT)** de cada uno de los datos de entrada: aprendizaje supervisado, no supervisado, semi-supervisado y por refuerzo. Cada uno de ellos tiene unas características y unos algoritmos determinados, por lo que se utilizarán unos u otros dependiendo del problema.

En el aprendizaje **supervisado** se entrena al algoritmo con datos etiquetados, es decir, aprende a partir de esos datos cómo predecir el resultado para otros en los que se desconoce su etiqueta. Este tipo de algoritmos necesitan conocer el GT de los datos de entrada. En contraposición con el anterior, en el aprendizaje **no supervisado** el algoritmo entrena con datos no etiquetados, por lo que busca patrones y estructuras similares y forma grupos en función de estas. En un punto intermedio entre ambos está el aprendizaje **semi-supervisado**, que se caracteriza por combinar datos de entrenamiento etiquetados y no etiquetados. Habitualmente se suele disponer de una mayor cantidad de datos sin etiquetar al ser estos más sencillos de obtener. Finalmente, el aprendizaje **reforzado (*reinforcement learning*)** es una forma de aprendizaje automático basado en un sistema de recompensas y castigos en el que un agente busca las decisiones óptimas para obtener la máxima recompensa tanto a corto como a largo plazo.

5.2.2. Etapas de un experimento con Machine Learning

Cualquier proyecto de ML no se centra únicamente en elegir un algoritmo y entrenarlo, sino que cuenta con una serie de pasos a seguir para aumentar sus probabilidades de éxito.

La primera etapa consiste en entender el problema que se quiere resolver. Después, se define la cantidad y el tipo de datos necesarios. Una vez recolectados y debidamente anotados se divide el conjunto total en tres subconjuntos: **entrenamiento, validación y test**. Los dos primeros conjuntos suelen suponer un 80% del total, dejando el 20% restante para el test. Es muy importante que no haya casos que se repitan entre los tres conjuntos para garantizar que el algoritmo de ML generalice y no se aprenda los ejemplos de memoria ('overfitting').

Dependiendo de la complejidad del problema y el tamaño de la base de datos, se elige el tipo de algoritmo de ML que mejor se vaya a ajustar a la tarea prevista: CNN, Bag of Words (BoW), regresión lineal, árboles de decisión, máquinas de vector soporte ('SVM'), etc.) así como la forma de entrenamiento (aprendizaje supervisado, no supervisado, etc.), como se puede ver en la figura 13.

La siguiente etapa consiste en entrenar el algoritmo seleccionado a partir de los datos de entrenamiento. Por tanto, el algoritmo debe crear una serie de "reglas" a partir de la observación de los datos de entrada para poder predecir su salida. A medida que se entrena el modelo, se ajustan sus parámetros (*'parameter tuning'*)(68) para mejorar su rendimiento.

Una vez entrenado el modelo se verifica su funcionamiento con los datos de test, los cuales el sistema no debería haber visto nunca. El rendimiento del sistema de ML se mide usando diferentes métricas como son la sensibilidad (S), especificidad (E), el valor predictivo positivo (VPP), el valor predictivo negativo (VPN) y la exactitud diagnóstica. La salida proporcionada por el modelo se compara siempre con el GT.

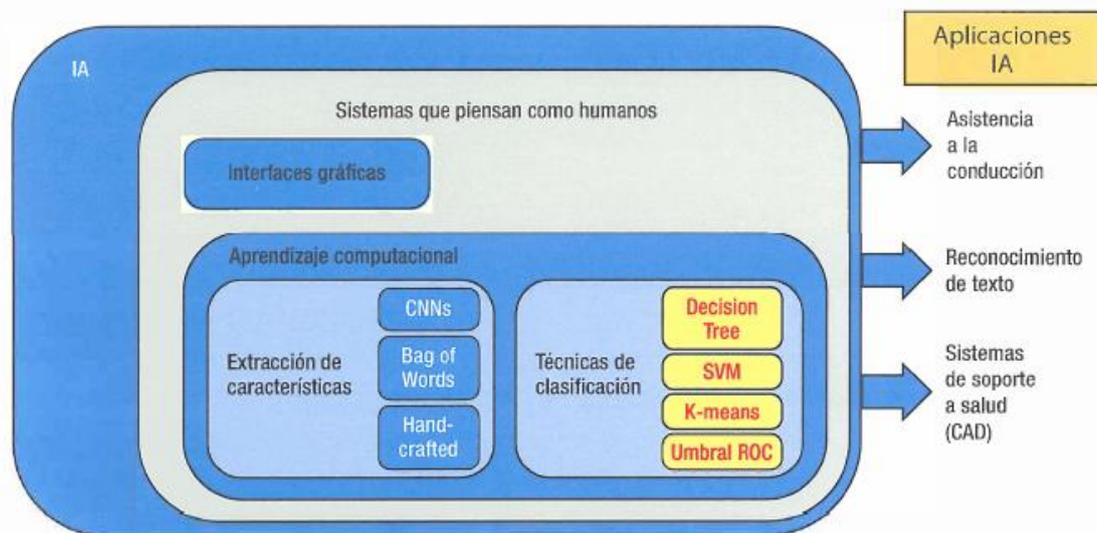


Figura 13: Gráfico explicativo de los componentes y las posibles aplicaciones de la IA (los sistemas de soporte a la salud entre ellas).

5.3. SISTEMAS AUTOMÁTICOS DE CLASIFICACIÓN DE IMÁGENES

La clasificación de imágenes es un proceso mediante el cual una máquina es capaz de decidir a qué clase pertenece cada uno de los objetos de una imagen. Los modelos de ML diseñados para clasificar imágenes constan de 2 etapas: 1. **Extracción de características** y 2. **Clasificación**. En este apartado vamos a explicar 2 técnicas de extracción de características que pueden ser acopladas a cualquier clasificador y que son las que se han utilizado para los trabajos de esta tesis.

5.3.1. Bag of Words

Los sistemas de búsqueda de información multimedia han evolucionado desde la “recuperación basada en **texto**” o **TBIR** (‘Text-based Image Retrieval’), donde las imágenes son descritas a través de un conjunto de palabras, a la “recuperación basada en **contenido**” o **CBIR** (‘Content-based Image Retrieval’), donde las imágenes son descritas a través de características visuales. Los sistemas CBIR y los sistemas híbridos, que combinan texto y contenido, surgen para superar las limitaciones de tiempo y coste que implican los sistemas TBIR. Los sistemas CBIR incluyen principalmente la elaboración de descriptores a partir del contenido de la imagen mediante los llamados rasgos de bajo nivel como son el color, la textura y la forma o una combinación de algunos de estos.

BoW es un modelo CBIR de clasificación de imágenes que representa una imagen en función de la frecuencia de una serie de elementos visuales (o “**palabras visuales**”). Se esperaría que, p. ej., una imagen de un edificio contenga más ventanas que una imagen de un árbol. La creación de un diccionario de “palabras” o “partes visuales” para cada imagen y su posterior agrupación y representación en un histograma de frecuencias, permite clasificarla y poder encontrar otras imágenes similares(69). A continuación (figura 14) se muestra un ejemplo de cómo funciona el BoW.

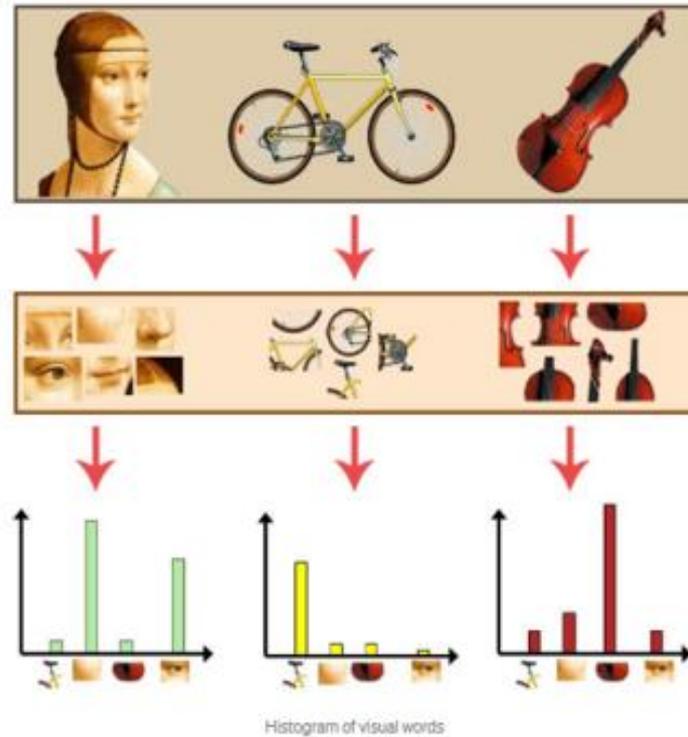


Figura 14: Ejemplo de cómo funciona el BoW. Cada imagen del conjunto de datos se representa mediante un histograma que contiene con qué frecuencia aparece en esa imagen cada una de las palabras visuales del diccionario. Los histogramas de imágenes de la misma categoría tendrán una distribución similar de palabras visuales mientras que los histogramas de imágenes de categorías distintas tendrán distribuciones diferentes.

5.3.2. Deep Learning

El aprendizaje profundo o DL ('Deep Learning') es un área específica dentro del ML que ha emergido en los últimos años y se basa en algoritmos de aprendizaje en múltiples niveles de representación y de abstracción con el fin de modelar relaciones más complejas entre los datos. En el DL estas representaciones por capas forman una **Red Neuronal** (RN) que, al fin y al cabo, es un conjunto estructurado de neuronas que permiten la extracción de las características de las imágenes a estudio.

Las capas de las RN están formadas por un número determinado de neuronas donde cada neurona recibe información de varias neuronas de la capa anterior. Las neuronas realizan cálculos internamente y devuelven un valor de salida que se transmite a las neuronas de la capa siguiente. En la última capa la información de las neuronas se combina para devolver la salida deseada(70). Una de las ventajas que ofrece la

estructura de una RN es la reducción del tiempo computacional al estar trabajando varias neuronas en paralelo, es decir, procesando información al mismo tiempo.

Las RN tienen tres tipos de capas: una **capa de entrada**, una o más **capas ocultas** y una **capa de salida** (figura 15). La **capa inicial de entrada** recibe información que proviene de sensores o de bases de datos con un peso y los transmite a la primera capa oculta. Esta capa tendrá tantas neuronas como entradas tenga la red. Las **capas ocultas** realizan todo el procesamiento de la información. La complejidad de cada RN vendrá determinada por el número de capas ocultas y la función de activación de las neuronas. La **capa final de salida** contiene el resultado realizado anteriormente en la capa oculta. Su número depende de la cantidad de clasificaciones que se requiera.

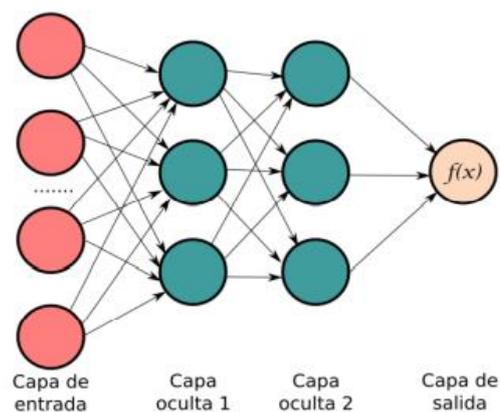


Figura 15: Esquema básico de una RN

Una red neuronal convolucional (**Convolutional Neural Network, CNN**), es un tipo especial de RN que se diferencia de las demás por el hecho de que cada una de las neuronas de las capas que la componen no recibe conexiones entrantes de todas las neuronas de la capa anterior, sino solo de algunas de ellas. Las CNN son hoy en día el estado del arte en el reconocimiento y clasificación de imágenes. Algunas implementaciones específicas que podemos encontrar sobre este tipo de redes son: VGG16, VGG19, ResNet50, Inception V3 y Xception.

Antes de la era DL, la decisión sobre las características a extraer de cada imagen se llevaba a cabo “manualmente”. En este caso, tras estudiar la cohorte de datos, los investigadores decidían qué características describían mejor cada imagen (y cada objeto dentro de una imagen) y construían el algoritmo necesario para poder extraerlas y así ofrecer al usuario la aplicación demandada. Normalmente las soluciones para cada problema tendían a ser específicas y hechas a medida para el mismo, de ahí su nombre ‘hand-crafted’. Este enfoque tradicional obligaba a pensar bien qué características eran importantes en cada imagen, pero a medida que el número y variabilidad de ejemplos aumentaba, la extracción de características se volvía cada vez más engorrosa.

El DL introduce el concepto de *end-to-end learning*, donde la máquina recibe un conjunto de datos de imágenes con anotaciones sobre qué clases de objetos están presentes en cada una de ellas y es capaz de extraer automáticamente las características más descriptivas con respecto a cada clase específica de objeto. Por tanto, las ‘**hand-crafted features**’ son características seleccionadas y programadas manualmente por el investigador mientras que las ‘**learned features**’ son las que se obtienen automáticamente de un algoritmo de DL (figura 16).

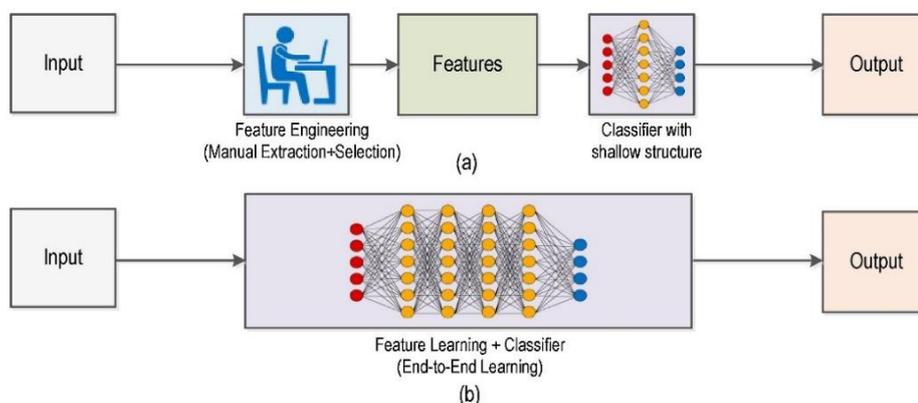


Figura 16: Flujo de trabajo del enfoque tradicional (a) y en la era DL (b)

Debido a que las características aprendidas se extraen automáticamente para resolver una tarea específica (en este caso la de clasificar), son extremadamente efectivas para ello, aunque requieren de un mayor esfuerzo computacional y de una gran cantidad de datos.

5.4. INTELIGENCIA ARTIFICIAL PARA ASISTIR A LA COLONOSCOPIA

La colonoscopia es el principal procedimiento diagnóstico para detectar y reconocer las lesiones precursoras del CCR, los pólipos. La detección precisa y la clasificación correcta dependen de las habilidades y la experiencia del endoscopista, sin embargo, incluso para los endoscopistas experimentados, trabajar durante muchas horas provoca fatiga mental y física. Los avances tecnológicos dirigidos a la mejora de la caracterización de lesiones, como son los sistemas de alta definición o la adición de filtros ópticos o electrónicos que optimizan la visualización de los mucosos y vasculares superficiales, siguen siendo insuficientes.

Los sistemas inteligentes tienen unas ventajas evidentes sobre los humanos: no se cansan, no se ven afectados por factores externos y son independientes del operador. La IA tiene al menos 2 aplicaciones principales para asistir en la colonoscopia: la **detección (CADe**, por sus siglas en inglés Computed-Aided Detection, incluida la localización) y la **clasificación (CADx**, Computed-Aided Diagnosis) de pólipos.

Los **sistemas CADe** tienen como objetivo decidir si hay un pólipo o no en la imagen y habitualmente están diseñados para alertar a los endoscopistas con un sonido o marcador cuando se sospecha la presencia de un pólipo durante la realización de una colonoscopia. Los **sistemas de localización** pueden entenderse como una extensión de los métodos de detección. En este caso, no solamente importa que el método automático indique correctamente la presencia o ausencia del pólipo en la imagen, sino también que en caso de detectar que hay pólipo, sea capaz de resaltar correctamente el área de la imagen donde está localizado dicho pólipo.

Los **sistemas CADx** se dedican fundamentalmente a clasificar los pólipos en función de su clase histológica (generalmente binaria, pero podría ser también multiclase).

5.5. ESTUDIOS DE SISTEMAS CADX YA PUBLICADOS:

El desarrollo de sistemas CADx, a diferencia de los CADe que han demostrado su rendimiento en diferentes ensayos aleatorizados y han sido autorizados ya por organismos reguladores, ha evolucionado más lentamente y todavía hoy no se han implementado en la práctica clínica. Se han publicado numerosos estudios a lo largo de

los últimos años, cada vez mejor diseñados y con resultados prometedores, pero la mayoría requieren tecnología óptica compleja y han sido probados en entornos reducidos y ex vivo, lo que sigue dificultando su aplicabilidad en vida real. Los algoritmos de prácticamente todos los CADx cuentan con un aprendizaje supervisado, es decir, imágenes o vídeos etiquetados (en función del GT) retrospectivamente por uno o varios expertos para poder hacer sus predicciones. A continuación, se describen los dispositivos publicados según la metodología óptica utilizada:

5.5.1. CADx con NBI (con y sin magnificación)

La primera publicación de un sistema CADx mediante NBI con magnificación fue descrita por primera vez por Tischendorf et al(71) en 2010. Su propuesta computacional consistía en: 1) preprocesado de imagen, 2) segmentación de estructuras vasculares de la superficie del pólipo, 3) extracción de características de dichas estructuras vasculares (longitud, brillo, perímetro y otras) y 4) clasificación final de la imagen (neoplásico y no neoplásico). Estos autores analizaron 209 imágenes (76,6% pólipos neoplásicos) obtenidas prospectivamente de 128 pacientes, consiguiendo una S, E y exactitud diagnóstica del 90, 70,2 y 85,3%, respectivamente.

Después de su trabajo, varios investigadores desarrollaron sistemas CADx con esta modalidad óptica (NBI + magnificación) para diferenciar pólipos neoplásicos de no neoplásicos mediante algoritmos de ML (en concreto, SVM). Gross et al(72) en 2011, en un estudio prospectivo con 214 pacientes, alcanzaron una exactitud diagnóstica del 93,1%. Su modelo se centró en el análisis del mapa vascular en 434 imágenes (59,4% pólipos neoplásicos), con una S, E y valores predictivos que superaron el 90% (y resultados comparables para el subgrupo de pólipos diminutos, n = 255).

En 2012, Takemura et al(73), analizaron retrospectivamente su sistema basado en el color, la forma y el tamaño de las criptas en 371 imágenes (87,3% pólipos neoplásicos) y obtuvieron todavía mejores resultados: exactitud diagnóstica 97,8%, con similar S y E. En la misma línea, Tamai et al(74) analizaron computacionalmente 121 imágenes (82,6% pólipos neoplásicos) y lograron los siguientes resultados: S: 83,9%, E: 82,6%, valor predictivo positivo (VPP): 53,1%, valor predictivo negativo (VPN): 95,6% y exactitud diagnóstica: 82,8%.

Kominami et al(75) describieron un algoritmo más robusto y adecuado para su uso clínico en tiempo real, que ofrecía clasificación in vivo de pólipos durante la endoscopia, mediante el análisis de características texturales de las criptas. El estudio incluyó 118 imágenes (62% de pólipos neoplásicos) y obtuvo una exactitud diagnóstica del 93,2%, con S, E y valores predictivos que también superaron el 90%.

Posteriormente, Chen(76) y Byrne et al(77) publicaron 2 estudios con CADx y NBI (+ magnificación/near focus respectivamente) basados en algoritmos de DL (en vez de ML), que alcanzaron los umbrales PIVI, concretamente >90 % VPN para el diagnóstico de adenomas diminutos (≤ 5 mm).

El estudio de Chen et al(76) con 284 imágenes recolectadas prospectivamente (66,2% pólipos neoplásicos) obtuvo las siguientes métricas: S: 96,3%, E: 78,1%, VPP: 89,6%, VPN: 91,5% y exactitud diagnóstica del 90,1%. Por su parte, Byrne et al(77) validaron su sistema en tiempo real con 125 vídeos (y no imágenes estáticas) con un 94% de eficacia diagnóstica, S del 98%, E del 83%, VPP del 90% y VPN del 97%.

5.5.2. CADx con Blue Laser/Light Imaging

Recientemente, Van der Zander et al(78) han presentado un estudio sobre un novedoso sistema CADx basado en CNN con 60 imágenes (45 (pre) malignos, 33 diminutos, 54 pacientes) en el que demuestran que al combinar imágenes de luz blanca y BLI (las llamadas imágenes multimodales), la exactitud diagnóstica general alcanza el 95% y es superior a la de los endoscopistas expertos (81,7%) y novatos (66,7%) para diferenciar pólipos benignos (PH) de (pre)malignos (LSS, adenomas, ADK). Curiosamente, también demostraron que la caracterización inicial de pólipos sin IA (definida como “intuición”) tenía una baja precisión diagnóstica, que no aumentaba al utilizar la clasificación BASIC ni para expertos (intuición 79,5% vs BASIC 81,7%; $p=0,14$) ni para principiantes (intuición 66,7% vs BASIC 66,5%; $p=0,95$). A la vista de los resultados, los autores enfatizan el valor de la IA en el proceso de caracterización.

5.5.3. CADx con endocitoscopia

La endocitoscopia también ha sido utilizada para la clasificación automática de pólipos. Los primeros modelos publicados por el grupo japonés liderado por Mori(79) se basaban en la extracción automática de características magnificadas (380 y 450x) del núcleo y de

los elementos con vascularización. Estas características se integraban en un clasificador de ML y permitían alcanzar una exactitud diagnóstica en torno al 90% para la identificación de adenomas/lesiones neoplásicas, con una demora de resultado desde la captura de la imagen de 0,2-0,3 segundos. Al tipo de pólipo se añadía una probabilidad que era de “alta confianza” si era > 90%(80).

Posteriormente, Takeda et al(81) utilizaron también endocitoscopia con una magnificación de aproximadamente 400x. Entrenaron el sistema con 5.543 imágenes de 238 lesiones y lo aplicaron a 200 imágenes (100 adenomas y 100 cánceres invasivos) de 76 lesiones. La S, E, VPP, VPN y exactitud diagnóstica obtenidas en esta serie fueron: 89,4, 98,9, 98,8, 90,1 y 94,1%. En 2018, Mori et al(82) publicaron otra serie retrospectiva con 144 imágenes de pólipos diminutos (36,8% pólipos neoplásicos) y obtuvieron los siguientes resultados: S 98%, E 71%, VPP 67%, VPN 98% y exactitud diagnóstica 81%. Ese mismo año, Mori et al(83) mejoraron sus resultados en un estudio prospectivo a gran escala utilizando endocitoscopia con magnificación 520x y añadiendo NBI en 466 imágenes (61,6% pólipos neoplásicos) en tiempo real e in vivo, analizando características texturales y obteniendo una exactitud diagnóstica en la valoración de pólipos diminutos neoplásicos/no neoplásicos del 98,1% (457/466); la S, E y valores predictivos fueron en torno al 90%.

El mismo grupo (Misawa et al) evaluó la siguiente generación de su CADx basado en endocitoscopia con NBI (EndoBRAIN Cybernet Systems®) primero con 100 imágenes magnificadas 380x (50% pólipos neoplásicos)(84) y posteriormente con 173 (124 pólipos neoplásicos)(85) obteniendo una exactitud diagnóstica del 90% y 87,8% respectivamente. Más tarde, Kudo et al(86) validaron retrospectivamente y de forma multicéntrica dicho modelo y lo compararon con 30 endoscopistas (20 en formación y 10 expertos) de cinco centros de alto volumen, obteniéndose unas métricas por encima del 95%, con mayor exactitud diagnóstica que los endoscopistas en formación (similar a la de los expertos). Finalmente, Barua et al(87), también del mismo equipo, han publicado el primer estudio multicéntrico (que incluye 3 centros entre oriente y occidente) in vivo que evalúa y compara el rendimiento de EndoBRAIN; Cybernet Systems® con el de los endoscopistas para diferenciar pólipos neoplásicos diminutos de

rectosigma (excluyeron a los endoscopistas nóveles). La S de los endoscopistas para los pólipos neoplásicos fue del 88,4% (IC del 95%: 84,3 a 91,5) y la del CADx del 90,4% (IC 95%: 86,8 a 93,1), $p=0,33$. Los autores apuntan que su CADx funciona especialmente bien para poder decidir qué lesiones dejar in situ.

5.5.4. CADx con otras modalidades ópticas complejas

Existen publicados otros estudios que también han empleado tecnologías ópticas complejas (CLE(88,89), espectroscopia(90) y AFI(91)) y han mostrado buenos resultados.

5.5.5. CADx con luz blanca

En lo que a predicción histológica automática con luz blanca se refiere la literatura es escasa.

Una CNN desarrollada por Komeda et al(92) mostraron una exactitud diagnóstica del 75,1% para la diferenciación de adenomas en 1.800 imágenes (66,6% adenomas) extraídas de vídeos de colonoscopia con luz blanca, NBI y cromoendoscopia. Sánchez-Montes et al(93) desarrollaron un modelo predictivo '*hand-crafted*' basado en 3 métricas (contraste, tubularidad y ramificación) de la superficie del pólipo (figura 17). Este sistema fue validado off-line en 225 imágenes de alta calidad (63,1% displásicos), obteniendo una exactitud diagnóstica del 91,1%, S del 92.3, E del 89.2, VPP del 93.6 y VPN del 87.1. Para el subgrupo de pólipos diminutos del rectosigma ($n= 54$) el VPN fue del 96,7%. Sin embargo, el sistema no era completamente automático al requerir de una segmentación previa del pólipo y no funcionaba en tiempo real.

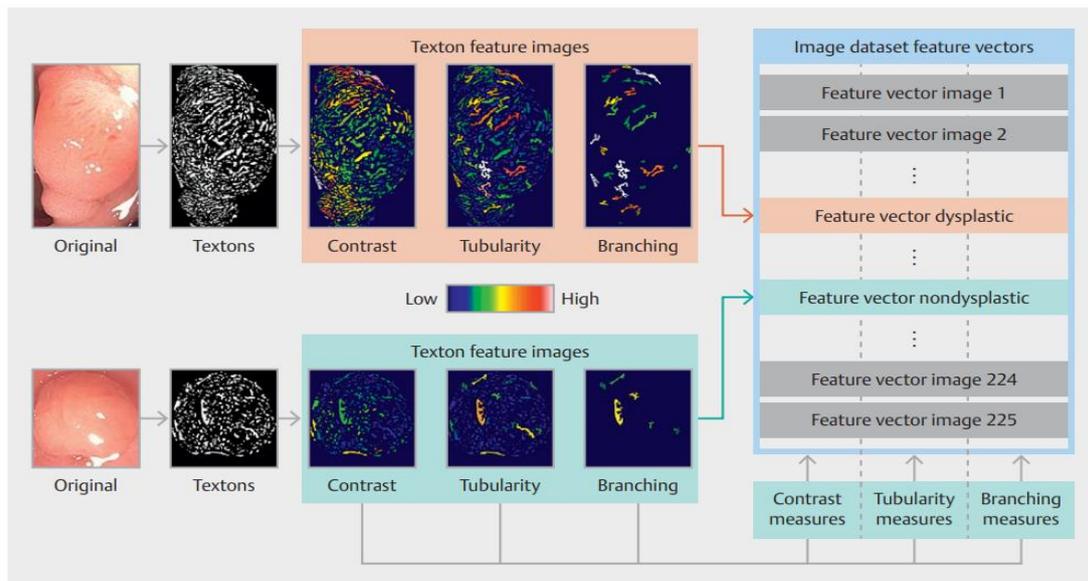


Figura 17: Representación gráfica del modelo CADx de Sánchez-Montes et al. Extracción de vectores de características a partir de la obtención de los textones de la imagen de entrada. De ellos se extraen tres características globales: contraste, tubularidad y ramificación. La presencia de textones de aspecto tubular y con ramificaciones se asocia a la categoría de pólipo con displasia.

Más adelante, Yang et al(94) propusieron un modelo CNN cuya exactitud diagnóstica fue similar a la de los endoscopistas expertos (71,5 % frente a 67,5 %) en 240 imágenes de luz blanca que incluían 3 imágenes de CCR avanzado, 8 imágenes de adenomas con displasia de alto grado, 116 imágenes de adenomas tubulares (con displasia de bajo grado) y 113 imágenes de lesiones no neoplásicas.

Dada la gran cantidad de estudios publicados en los últimos años, se han publicado varias revisiones y metaanálisis, cuyas conclusiones, dada la heterogeneidad de los sistemas CADx incluidos, deberían tomarse con mucha cautela. Por destacar uno de ellos, en el meta-análisis de Bang et al(95) que incluye 13 estudios de caracterización automática de pólipos diminutos (todos adoptaron la definición de ≤ 5 mm a excepción de Shahidi et al(96) que adoptaron una definición más estricta con un tamaño < 3 mm), la gran mayoría alcanza unos umbrales de S y E superiores al 80% y de VPN del 90% en rectosigmo, lo que facilitaría la implementación de la estrategia 'leave in situ'.

En la tabla 3 se resumen todos los estudios mencionados:

Autores	Año publicación	País	Reclutamiento	Algoritmo CADx	Modalidad óptica	N pacientes	N lesiones	Exactitud diagnóstica	Clasificación histológica del CADx	Tipo de BD
Tischendor et al(71)	2010	Alemania	Piloto, prospectivo	SVM	NBI + magnificación	128	209	85.3%	Neoplásico vs no neoplásico	Privada
Gross et al(72)	2011	Alemania	Prospectivo	Hand-crafted+ SVM	NBI + magnificación	214	434	93.1%	Neoplásico vs no neoplásico	Privada
Takemura et al(73)	2012	Japón	Retrospectivo	SVM	NBI + magnificación	ND	371	97.8%	Neoplásico vs no neoplásico	Privada
Tamai et al(74)	2017	Japón	Retrospectivo	Hand-crafted	NBI + magnificación	103	121	82.8%	Neoplásico (adenoma/adenocarcinoma) vs no neoplásico	Privada
Kominami et al(75)	2016	Japón	Retrospectivo	SIFT+ SVM	NBI + magnificación	41	118	93.2%	Neoplásico vs no neoplásico	Privada
Chen et al(76)	2018	Taiwán	Prospectivo	DL	NBI + magnificación	193	284	90.1%	Neoplásico vs no neoplásico	Privada
Byrne et al(77)	2017	Canadá	Retrospectivo	DL	NBI +/- near focus	ND	125	94%	Adenoma vs PH	Privada
Komeda et al(92)	2017	Japón	Retrospectivo	DL	LB, NBI y CE	ND	ND	75.1%	Adenoma vs no adenoma	Privada
Sánchez-Montes et al(93)	2019	España	Retrospectivo	Hand-crafted+ SVM	LB	ND	225	91,10%	Displásico vs no displásico	Privada
Mori et al(79)	2015	Japón	Retrospectivo	Hand-crafted + Regresión lineal	EC	152	176	89.2%	Neoplásico vs no neoplásico	Privada
Mori et al(80)	2016	Japón	Retrospectivo	Hand-crafted + SVM	EC	123	205	ND	Neoplásico vs no neoplásico	Privada
Mori et al(82)	2018	Japón	Retrospectivo	ND	EC	ND	144	81%	Neoplásico vs no neoplásico	Privada

Mori et al(83)	2018	Japón	Prospectivo	Hand-crafted + SVM	EC con NBI	325	466	98,10%	Neoplásico vs no neoplásico (solo diminutos)	Privada
Takeda et al	2017	Japón	Retrospectivo	Hand-crafted + SVM	EC	76	200	94.1%	Adenoma vs CCR invasivo	Privada
Misawa et al(84)	2016	Japón	Retrospectivo	EndoBRAIN	EC con NBI	ND	100	90%	Neoplásico vs no neoplásico	Privada
Misawa et al(85)	2017	Japón	Retrospectivo	EndoBRAIN / ECV-CAD	EC con NBI	ND	173	87.8%	Neoplásico vs no neoplásico	Privada
Yang et al(94)	2020	Corea	Retrospectivo	DL	LB	ND	240	ND	CCR invasivo y precoz/DAG, DGB y no adenomas	Privada
Van der Zander et al(78)	2020	Países Bajos	Prospectivo	DL	LB y BLI (multimodal)	54	60	95.0%	PH vs (pre)malignos (adenomas, LSS y CCR)	Privada
Barua et al(87)	2022	Noruega, Reino Unido, Japón	Prospectivo	EndoBRAIN system	EC con LB y NBI	518	892	90,40%	Neoplásico vs no neoplásico	Privada
Kudo S et al(86)	2019	Japón (5 centros)	Retrospectivo	EndoBRAIN system	EC con LB y NBI	89	100	98%	Neoplásico vs no neoplásico	Privada

Tabla 3: Resumen estudios CADx publicados. EC: Endocitoscopia. LB: Luz blanca. NF: Near Focus. DL: Deep Learning. SVM: Support Vector Machine. BLI: Blue Laser/Light Imaging. NBI: Narrow Band Imaging. DAG: Displasia de alto grado. DBG: Displasia de bajo grado. PH: Pólipo hiperplásico. LSS: Lesión serrada sénil. CCR: Cáncer colorrectal. ND: No disponible. BD: Base de datos.

5.6. IMPLEMENTACIÓN DE LA INTELIGENCIA ARTIFICIAL EN LA VIDA REAL

Los siguientes requisitos técnicos son imprescindibles para la integración de los sistemas CADx en el flujo de trabajo clínico:

1. Capacidad de trabajar en **tiempo real**, esto es, el tiempo que transcurre desde la captura de la imagen hasta la salida de los resultados por pantalla debe ser menor o igual al inverso de la tasa de captura de imágenes. P. ej. si el sistema adquiere 25 imágenes por segundo, el sistema IA deberá tardar un máximo de 40 milisegundos para procesar cada imagen. Para acortar este tiempo, una estrategia es contar con mejor equipamiento informático u optimizar los algoritmos. No debe confundirse “tiempo real” con la interpretación que en muchos artículos científicos se realiza de ‘real-time colonoscopy’, refiriéndose a la ejecución del sistema en el mismo momento que se realiza la colonoscopia (*‘on-line’*, *‘on-site’* y *‘during the colonoscopy’* son términos que reflejan mejor a esta definición).

2. **Viabilidad de implementación física del sistema en la sala de exploración.** Las tecnologías de DL emergentes requieren hardware específico, voluminoso y de alto consumo energético, lo que impone restricciones en cuanto a su uso en la sala de exploración. Una alternativa es usar la computación en la nube (cloud-computing) aunque en este caso aparecen restricciones asociadas al uso de redes inalámbricas.

3. **Interfaz de usuario no disruptiva.** Idealmente la salida de los dispositivos de IA debería mostrarse en la misma pantalla en la que se está viendo la exploración, dado que el uso de otra pantalla podría distraer al endoscopista y/o requerir de personal adicional no siempre disponible. En este caso se recomienda el uso de alarmas visuales cuando aparezca un objeto de interés y la combinación de alertas visuales y sonoras. En todo caso, en el diseño de la interfaz de usuario deben colaborar tanto los ingenieros como los usuarios finales de cara a tener un sistema que sea fácil de usar e interpretar.

5.7. BASES DE DATOS

Las bases de datos de imágenes son necesarias para validar los métodos computacionales, así como para facilitar la comparativa entre ellos y garantizar que éstos representen adecuadamente el problema.

Los casos incluidos en una base de datos deben representar toda la variabilidad clínica existente en la práctica diaria. Además, cuanto más diferentes sean los ejemplos, más robusto será el sistema desarrollado. El tamaño final de la base de datos debería permitir la extracción de conclusiones estadísticamente significativas (el número mínimo para alcanzar la meseta de aprendizaje aún se desconoce).

A diferencia de las grandes bases de datos como ImageNet(97), con más de 14 millones de imágenes anotadas a mano en 20.000 categorías o COCO(98) de Microsoft, en la medicina es difícil crear tales anotaciones y de tan alta calidad debido a los altos costes que ello supone (económicos y de experiencia requerida). En el caso de la colonoscopia, cada vez hay disponibles más bases de datos públicas para la detección y clasificación de pólipos.

En la siguiente tabla 4 resumimos las más importantes:

Nombre	Número de imágenes	Número de pólipos distintos	Resolución imagen	Aplicación
CVC-ClinicDB	612	31	388x284 (subSD)	Detección, localización, segmentación
CVC-ColonDB	300	15	574x500 (SD)	Detección, localización, segmentación
ETIS-Larib	196	44	1920x1080 (HD)	Detección, localización, segmentación
CVC-ClinicVideoDB	9830 (entrenamiento) 2124 (validación) 18733 (test)	15 (entrenamiento) 3 (validación) 18 (test)	388x284 (subSD)	Detección, localización
Kvasir-SEG	1000	1000	Resolución variable	Detección, localización, segmentación
CVC-HDClassif	1153		1920x1080 (HD)	Segmentación, clasificación
ASU Mayo	15432 (entrenamiento) 18321 (test)	20 (entrenamiento) 20 (test)	Resolución variable	Detección, localización

Tabla 4: Bases de datos públicas disponibles hasta la fecha. La mayoría generadas por el equipo del Centro de Visión por Computador. Solo hay una base de datos específica para la clasificación de pólipos; la mayoría están dedicadas a las tareas de detección y localización.

Puesto que todavía no son lo suficientemente voluminosas como para desarrollar modelos de ML precisos, la mayoría de estudios han recopilado sus propias bases de datos (privadas) para llevar a cabo su investigación.

Para mitigar los problemas asociados al reducido tamaño de las bases de datos de imágenes médicas, los investigadores a menudo usan 2 estrategias: (1) *transfer learning*, and (2) *data augmentation*. En *transfer learning*, el modelo se actualiza a partir de un modelo de red preentrenado que se obtiene de otros campos, como la mencionada ImageNET(97), para entrenar la RN. En *data augmentation*, los nuevos datos son generados artificialmente mediante el uso de los datos disponibles con ciertas manipulaciones realistas como rotar, traducir, agregar ruido, voltear, etc.

Como ya se ha comentado, el exitoso desarrollo de dispositivos inteligentes para asistir a la colonoscopia requiere la disponibilidad de grandes conjuntos de datos (imágenes o videos de pólipos) junto con anotaciones manuales de alta calidad proporcionadas por expertos. Estas anotaciones deben proporcionar el **GT**. Si p. ej. se desarrolla un sistema para detección de pólipos el GT consistirá en un archivo Excel que indica si hay o no un pólipo en cada imagen y únicamente se requerirá el etiquetado de la imagen como TRUE/FALSE. Un sistema de localización implicará definir el área de la imagen que ocupa el pólipo. Finalmente, un sistema de clasificación de pólipos requerirá asignar una etiqueta a la imagen según la clase histológica.

5.8. LIMITACIONES DE LOS ESTUDIOS CADX EXISTENTES

Se han identificado varias limitaciones en los estudios ya publicados sobre sistemas CADx, algunas ya mencionadas en los apartados anteriores, que se desarrollarán seguidamente.

La primera hace referencia a los datos que se usan para entrenar los modelos. Los sistemas CADx usan como GT para su entrenamiento la histología, que a pesar de ser considerada universalmente como el estándar oro, también tiene una variación considerable entre observadores, especialmente en la diferenciación de LSS con PH(99).

El reducido tamaño de las bases de datos es otra gran limitación ya que impide desarrollar modelos que sean generalizables, sobre todo teniendo en cuenta el desbalance entre las distintas clases, habiendo muchos más ejemplos de adenomas. En este aspecto alcanzar los umbrales establecidos por la ASGE-PIVI (que es el objetivo que más estudios sobre CADx se proponen) es complicado dado que, además, se suelen incluir imágenes de pólipos de diferentes tamaños, y no sólo diminutos.

Por otro lado, el sesgo de selección inherente a los estudios retrospectivos y ex vivo, donde a menudo se seleccionan las “mejores” imágenes como imágenes para el entrenamiento y el test respectivamente, lo que hace que los resultados no sean verdaderamente representativos ni extrapolables a la práctica clínica habitual. Hay todavía pocos estudios multicéntricos prospectivos que analicen imágenes en tiempo real (es poco común que se especifique cuánto tiempo tarda el método en calcular un resultado lo cual imposibilita saber si puede usarse o no en la sala de exploración) e in vivo. Asimismo, los estudios retrospectivos no contemplan otros aspectos asociados al uso en vida real de la IA como tiempo adicional para el examen y sobrecarga del endoscopista, nivel de experiencia y confianza(100). Además, en la práctica clínica, la caracterización de los pólipos de la IA se ve afectada por los movimientos de la cámara, las ondas peristálticas, las burbujas/residuos y las alteraciones ópticas relacionadas con el enfoque y el reflejo de la luz.

También hay que decir que casi todos los sistemas CADx publicados emplean tecnologías ópticas complejas (NBI con y sin magnificación, CLE y endocitoscopia, espectroscopia y AFI), lo que impide la universalización de estos. Finalmente, y dado que la mayoría de los estudios son unicéntricos y se evalúan en bases de datos privadas, sus resultados pueden no ser extrapolables ni generalizables. Además, de cara a establecer el beneficio clínico asociado al uso de los sistemas de IA y la autorización para su comercialización se requieren ensayos clínicos multicéntricos con endoscopistas con diferente nivel de experiencia en colonoscopia(101).

5.9. DISPOSITIVOS INTELIGENTES COMERCIALIZADOS Y REGULACIÓN

Las agencias reguladoras tienen la tarea de proteger la salud pública al garantizar la seguridad de los dispositivos médicos. Existe cierto temor a que el uso de la IA pueda traer alguna consecuencia negativa. Progresivamente, las preguntas sobre la regulación, qué forma toma y quién es responsable de la gobernanza comienzan a ser respondidas. La Organización para la Cooperación y el Desarrollo Económico y los países socios adoptaron en 2019 el primer conjunto de directrices políticas intergubernamentales sobre IA(102). La Comisión Europea (CE) publicó posteriormente un Libro Blanco titulado “La inteligencia artificial: un enfoque europeo para la excelencia y la confianza” (Informe

de la CE, 2020). En el Libro Blanco se definen las opciones existentes para alcanzar el doble objetivo de promover la adopción de la IA y abordar los riesgos vinculados a determinados usos de esta nueva tecnología(103). El uso de cualquier dispositivo en Europa requiere una marca CE. Desde mayo de 2021, se aplica una nueva Regulación de Dispositivos Médicos (MDR por sus siglas en inglés) en Europa, que define las normativas y obligaciones que deben cumplir todos los fabricantes y distribuidores para comercializar un dispositivo médico en el mercado europeo. Así pues, la aprobación de futuros sistemas CAD requerirá un nivel de evidencia significativamente más alto y una vigilancia posterior a la comercialización más exhaustiva que incluya la preparación de un informe periódico de actualización de seguridad ("PSUR")(104).

Como se ha podido ver en la sección 5.5, la calidad de los sistemas desarrollados hasta la fecha y la de los estudios que exploran su desempeño ha sido muy variable y escasamente evaluada en la práctica clínica real, lo que ha ralentizado su incorporación al mercado. En la tabla 5 se resumen los sistemas CAD comercializados hasta el momento, incluyendo los CADx(104).

Nombre	EndoBRAIN	GI Genius	Discovery	ReiLi CAD EYE	ai4gi	EndoScreener
Empresa	Cybernet Systems	Medtronic	Pentax	Fujifilm	Imagia, Satis	Shanghai Wision AI Co
Tipo de colonoscopio	CF-H290 ECI	Admite varias marcas	Pentax	Fujifilm serie 700	Admite varias marcas	Admite varias marcas
Doble monitor	No	No	Si	No	No	Yes
Estudio	Mori y Misawa et al(79,85)	Repici et al(105)	ND	ND	Byrne et al(77)	Wang et al(106)
Modo	CADe y CADx	CADe	CADe	CADe y CADx	CADe y CADx	CADe
Regulación	Aprobado por PMDA 12/2018	Aprobado por EMA 07/2019, FDA 2021	Aprobado por EMA 01/2020	Aprobado por EMA 02/2020	Aprobado por EMA 10/2021	Aprobado por EMA 11/2021

Tabla 5: Descripción general de los sistemas CAD comercializados. EMA: European Medicines Agency. PMDA: Japanese Pharmaceuticals and Medical Devices Agency. ND: No disponible.

Al hilo de lo anterior, y hasta que haya nuevos datos, la ESGE sugiere la posibilidad de incorporar sistemas inteligentes para asistir a la colonoscopia siempre que se haya demostrado su rendimiento, aunque con una calidad de evidencia baja(60). La interacción de la IA y el endoscopista puede ser **activa**, p. ej. si se encuentra con un pólipo y le pide al sistema que confirme su diagnóstico como un "segundo lector", o **pasiva**, es decir, si se ejecuta continuamente en un segundo plano, de manera que proporcione una "lectura concurrente", p. ej. en la detección. Puede haber también situaciones en las que la IA actúe de forma completamente autónoma para tomar una decisión sin ninguna entrada del endoscopista(60) (figura 18).

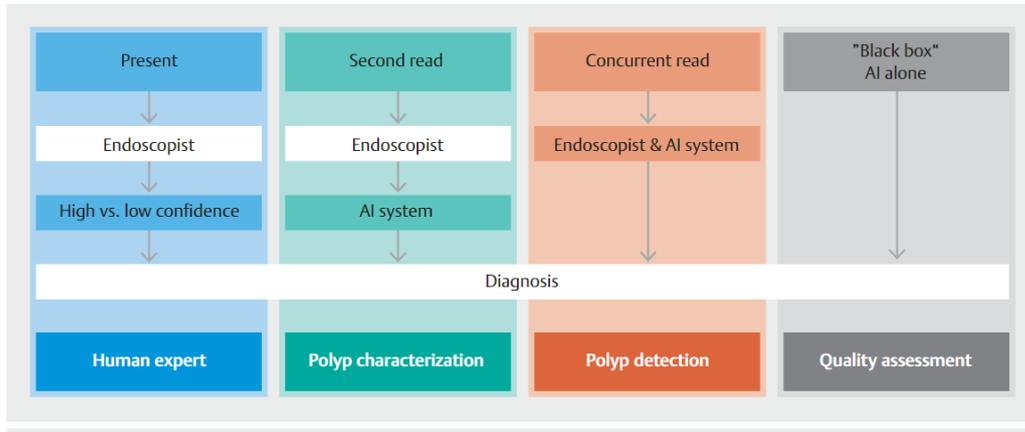


Figura 18: Posibilidades de interacción entre una IA y endoscopistas

Hay cierta preocupación por los efectos de la implementación de una IA en la curva de aprendizaje del endoscopista con la posibilidad de descualificación y sobre la influencia del dispositivo en la atención del endoscopista durante el procedimiento y posible aumento del tiempo procesal. Las sociedades científicas apuestan por una **IA híbrida** con el endoscopista y no una IA como tomadora de decisiones independiente.

En conclusión, la IA en la endoscopia digestiva debe basarse en la colaboración de los clínicos, la informática básica, las agencias reguladoras y las sociedades científicas(107).

HIPÓTESIS

La inteligencia artificial posee diferentes aplicaciones en el área de la colonoscopia y, más concretamente, sobre los pólipos de colon, considerados de gran relevancia clínica por su frecuencia e implicación en la prevención del cáncer colorrectal, que a su vez conlleva la realización de numerosas colonoscopias con el consiguiente impacto en la calidad de vida de los pacientes y coste económico para el sistema de salud. El desarrollo y la utilización de sistemas de soporte asistidos por inteligencia artificial en la colonoscopia podría: 1) mejorar el diagnóstico óptico de los pólipos de colon posibilitando la distinción entre clases histológicas y 2) reconocer una serie de características que los hacen singulares y fácilmente diferenciables, de manera que no se produzcan errores en su identificación. Para poder integrar la inteligencia artificial en esta área, la disponibilidad de una herramienta que permita crear el patrón oro de un conjunto de imágenes dado, de forma fácil y segura, también sería de gran utilidad.

OBJETIVOS

1. Desarrollar una aplicación que permita, por una parte, la visualización y manipulación de imágenes endoscópicas y por otra, la anotación semántica de las mismas, en un entorno colaborativo donde diferentes expertos (Medicina e Informática) puedan participar, constituyéndose como herramienta de apoyo a la tarea de clasificación.
2. Desarrollar y evaluar un sistema de clasificación automática de pólipos en tiempo real e in vivo utilizando solo imágenes de luz blanca de alta definición y comparar su rendimiento con el de los endoscopistas.
3. Desarrollar y evaluar un sistema de recuperación de imágenes mediante la extracción de características de pólipos en imágenes de luz blanca que permitan su identificación inequívoca (huella digital) y potencial agrupación en categorías histológicas por analogía.

MATERIAL Y MÉTODOS Y RESULTADOS

Cada uno de los objetivos de esta tesis ha dado lugar a una publicación.

Los artículos de esta tesis doctoral son:

- Jorge Bernal, Aymeric Histace, Marc Masana, Quentin Angermann, Cristina Sánchez-Montes, Cristina Rodríguez de Miguel, Maroua Hammami, **Ana García-Rodríguez**, Henry Córdova, Olivier Romain, Gloria Fernández-Esparrach, Xavier Dray, F. Javier Sánchez
« GTCreator: a flexible annotation tool for image-based datasets » Int J Comput Assist Radiol Surg. 2019 Feb 1;14(2):191–201.
- **Ana García-Rodríguez**, Yael Tudela, Henry Córdova, Sabela Carballal, Ingrid Ordás, Leticia Moreira, Eva Vaquero, Oswaldo Ortiz, Liseth Rivero, F Javier Sánchez, Miriam Cuatrecasas, Maria Pellisé, Jorge Bernal, Glòria Fernández-Esparrach
« In vivo computer-aided diagnosis of colorectal polyps using white light endoscopy » Endosc Int open. 2022
- **Ana García-Rodríguez**, Jorge Bernal, F. Javier Sánchez, Henry Córdova, Rodrigo Garcés Durán, Cristina Rodríguez de Miguel, Gloria Fernández-Esparrach
« Polyp fingerprint : automatic recognition of colorectal polyps' unique features » Surg Endosc. 2020;34(4):1887–9

ARTÍCULO 1:

TÍTULO

GTCreator: una herramienta de anotación flexible para conjuntos de datos de imágenes.

OBJETIVO

Desarrollar una aplicación que permita, por una parte, la visualización y manipulación de imágenes endoscópicas y por otra, la anotación semántica de las mismas, en un entorno colaborativo donde diferentes expertos (Medicina e Informática) puedan participar, constituyéndose como herramienta de apoyo a la tarea de clasificación.

RESUMEN

Propósito: La evaluación del rendimiento de los sistemas de apoyo a la toma de decisiones para la salud es una tarea que requiere mucho tiempo. Los expertos clínicos tienen que lidiar con la anotación de miles de imágenes. Actualmente, las herramientas existentes podrían mejorarse en términos de flexibilidad y facilidad de uso.

Métodos: GTCreator es una nueva herramienta que proporciona anotaciones de imagen y texto a grandes conjuntos de datos. Mantiene las funcionalidades básicas de otras herramientas similares al mismo tiempo que amplía otras capacidades como permitir múltiples anotadores simultáneamente y acelerar la navegación y transferencia de anotaciones.

Resultados: La comparación con otras herramientas similares mostró que GTCreator permitía obtener una rápida y precisa anotación de imagen, siendo el único que ofrecía capacidades completas de edición y navegación de anotaciones.

Conclusión: GTCreator es eficiente para la anotación de conjuntos de datos de imágenes de gran tamaño y muestra potencial de uso en otras etapas de evaluación de sistemas, como la configuración experimental o el análisis de resultados.



GTCreator: a flexible annotation tool for image-based datasets

Jorge Bernal¹ · Aymeric Histace² · Marc Masana¹ · Quentin Angermann² · Cristina Sánchez-Montes³ ·
Cristina Rodríguez de Miguel³ · Maroua Hammami² · Ana García-Rodríguez³ · Henry Córdova³ ·
Olivier Romain² · Gloria Fernández-Esparrach³ · Xavier Dray^{2,4} · F. Javier Sánchez¹

Received: 27 January 2018 / Accepted: 12 September 2018
© CARS 2018

Abstract

Purpose: Methodology evaluation for decision support systems for health is a time-consuming task. To assess performance of polyp detection methods in colonoscopy videos, clinicians have to deal with the annotation of thousands of images. Current existing tools could be improved in terms of flexibility and ease of use.

Methods: We introduce GTCreator, a flexible annotation tool for providing image and text annotations to image-based datasets. It keeps the main basic functionalities of other similar tools while extending other capabilities such as allowing multiple annotators to work simultaneously on the same task or enhanced dataset browsing and easy annotation transfer aiming to speed up annotation processes in large datasets.

Results: The comparison with other similar tools shows that GTCreator allows to obtain fast and precise annotation of image datasets, being the only one which offers full annotation editing and browsing capabilities.

Conclusion: Our proposed annotation tool has been proven to be efficient for large image dataset annotation, as well as showing potential of use in other stages of method evaluation such as experimental setup or results analysis.

Keywords Annotation tool · Validation framework · Benchmark · Colonoscopy · Evaluation

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11548-018-1864-x>) contains supplementary material, which is available to authorized users.

✉ Jorge Bernal
jorge.bernal@cvc.uab.cat

Aymeric Histace
aymeric.histace@ensea.fr

Marc Masana
marc.masana@cvc.uab.cat

Quentin Angermann
quentin.angermann@ensea.fr

Cristina Sánchez-Montes
crsanchez@clinic.cat

Cristina Rodríguez de Miguel
crodrigm@clinic.cat

Maroua Hammami
maroua.hammami@ensea.fr

Ana García-Rodríguez
anagrod4@gmail.com

Henry Córdova
hcordova@clinic.cat

Introduction

The development and validation of decision support systems (DSS) for health has seen an increase in interest in recent years due to the improvement in computing capabilities and the evolution of imaging technology. These systems provide

Olivier Romain
olivier.romain@ensea.fr

Gloria Fernández-Esparrach
mgfernan@clinic.cat

Xavier Dray
xavier.dray@aphp.fr

F. Javier Sánchez
javier.sanchez@cvc.uab.cat

- 1 Computer Vision Center and Universitat Autònoma de Barcelona, Barcelona, Spain
- 2 ETIS lab, ENSEA, CNRS, University of Cergy-Pontoise, Cergy, France
- 3 Endoscopy Unit, ICMDiM, Hospital Clinic, IDIBAPS, CIBEREHD, University of Barcelona, Barcelona, Spain
- 4 St. Antoine Hospital, APHP, Paris, France

objective information to clinicians to help them in the different stages of a clinical procedure, from intervention planning (automatic location of cancer nodules in a CT scan [1]), to assistance during in vivo explorations (renal lesion detection in ultrasound images [2]), or in decision-making processes (resection of colon polyps with automatic histology prediction [3]).

The scope of our research is the development and validation of intelligent systems for colonoscopy. These systems couple the expertise of clinicians and technicians to develop efficient solutions that can help to tackle some of the most demanding clinical needs. For the case of colonoscopy, polyp detection and in vivo histology prediction [4] are being identified as the key areas in which a computational system could be of great help.

Several computational methods have been proposed to tackle both polyp detection (the interested reader can find a comparison between several detection methodologies in [5]) and in vivo histology prediction [6]. Despite the large number of approaches, none of them is used in clinical facilities. The two main reasons for it are low practical feasibility due to not meeting real-time constraints, and the requirement of excessive computational resources. Furthermore, the lack of uniform and public frameworks hinders the performance assessment of these systems before their use by clinicians.

With respect to the latter, several attempts have been made to develop public validation frameworks, especially in the context of challenges in medical imaging conferences such as MICCAI or ISBI. However, there is still a lack in terms of quality and quantity with respect to clinically generated medical image annotations. For the case of colonoscopy video analysis, ground truth definition is a highly time-consuming task, requiring clinicians to precisely annotate thousands of images.

Several tools have been proposed to assist in ground truth generation which can be divided into two groups: general and domain specific. General tools include LabelMe [7] or VGG image annotator (VIA) [8], while tools such as Rat-snake [9] or Arthemis [10] were specifically designed for medical imaging domains. None of the tools from the latter group were designed along their final users, which results in powerful applications that, unfortunately, are not practically usable due to being complex to use.

In conclusion, the contributions of this article are twofold:

- GTCreator, a novel flexible tool for providing image and text-based annotations of image datasets,
- A comparison study between image annotation tools.

GTCreator has been designed to ease the task of ground truth (GT) annotation and management. It allows a great flexibility with respect to the number and type of annotations to

be provided for each image, making it suitable to use for any image dataset annotation task (see Sect. 2.1.2).

To show the benefits of our tool, we present a study in which we compare qualitatively and quantitatively six different annotation tools. In the proposed use case of annotation of colonoscopy videos and still images by experts, our tool keeps the main functionalities provided by other tools as well as incorporating new ones such as annotation transfer or full annotation editing capabilities (see Sect. 3).

The rest of the paper is structured as follows: Sect. 2 details the features and usage of GTCreator. We present the results of the comparison study against other similar tools in Sect. 3. Conclusions are presented in Sect. 4.

GTCreator: A flexible image annotation tool

Overview of GTCreator

GTCreator is a tool for creating and managing annotations of image databases. The result of an image annotation is a set of metadata (image and text-based) which will be associated with that image. The main innovation of GTCreator is its flexibility in comparison with other existing tools, as it allows to freely determine the number and type of metadata to be associated with each image during the annotation task. This makes our tool suitable for its use in any image annotation task.

We show in Fig. 1 our proposed pipeline for image dataset annotation. GTCreator consists of 3 stages after dataset collection: (1) definition file configuration for ground truth definition, (2) image annotation using graphical user interface (GUI) and (3) ground truth data exportation.

Definition file configuration

The definition file is the key element of GTCreator and consists of three different parts: the header containing global information of the annotation task, the list of metadata to be associated with each image, and the list of dataset images to be annotated.

The header contains: (a) name of the annotation set to be generated, (b) description of the annotation task and (c) the relative paths in which both dataset images and annotations will be stored. This allows to univocally identify and locate the annotation task and the data. To cover the range of possible annotation tasks, GTCreator admits three types of metadata:

- Images/masks, to represent the objects in the image. This annotation type admits two different options: annotation through freehand drawing or by using predetermined shapes.



Fig. 1 Proposed pipeline for image dataset annotation using GTCreator

- Text, to provide formatted information associated with the image. These category subtypes¹ include bool (for binary metadata), string or list of values.
- Semantic labels, to classify each of the predefined regions in the image.

To make annotations by multiple users easier and more coherent, each metadata can have an associated description (visual and/or text) to provide an example of how the annotation process is expected to be done. Finally, the definition file includes the list of dataset images to be annotated, and image names can be added manually or automatically through the GTCreator GUI.

Image annotation through Graphical User Interface

The GUI of our tool has been designed to allow an easy access to all functionalities. The GUI is divided into six different areas, as shown in Fig. 2:

1. Definition file and database management operations, which include the possibility of adding images to the annotation task.
2. Metadata editing area, in which clinicians can modify the value of the different metadata associated with each image.
3. Annotation assistance area, which might include a representative image and additional text to guide the annotation task.
4. Image annotation tools, including the possibility to change image scale change, mask transparency, brush size as well as additional capabilities to improve freehand contour definition and to change local image contrast to enhance image visibility.
5. Image annotation area.
6. Database browsing tools, including metadata-based filtering and shortcuts to ease navigation through non-annotated images.

An annotation session starts by loading the definition file associated with the task. This also allows to resume an already started annotation session. To ease the annotation of video datasets, we allow annotation transfer between consecutive images. Finally, considering that a large annotation task might be divided among different annotators, annotation merging from different observers is also supported. All those

functionalities are accessible through the described areas of the GUI.¹

Ground truth data exportation

Once the desired dataset has been annotated, it is important that it can easily be used widely by both clinical and technical users. As it will be discussed in Sect. 3, the format used to export the annotation data should be simple and easy to use. Therefore, all information related to the annotation task is organized as a table and stored in a CSV file, which is compatible with most commonly used software environments (MATLAB, Python or SPSS). This allows the use of the annotations for method validation purposes in many frameworks without further interaction with the data. In the CSV file, rows correspond to the different images in the dataset and the columns to the different metadata included in the definition file. Image-based annotations are stored as binary masks in TIFF files to ensure efficiency in data storage.

The way our tool is designed allows modification of the definition file during the annotation process while keeping intact the information already stored. In this case, the user should only complete the new metadata information for the already labeled images.

Use-case scenarios

GTCreator has been mainly designed to support annotation tasks, but it can also be used in other stages of the development and validation of an image processing method, as shown in Fig. 3. More precisely, we foresee the following scenarios in which our tool could play a key role:

- Provide image and text-based annotations of image datasets (**GT Creation & Annotation**). The interface of our tool allows to resume an already started annotation process as well as incorporating new metadata on the fly without having to start the annotation process from scratch. For the case of large datasets, our tool incorporates annotation merging capabilities in case annotation task was split into different annotators.
- Annotation review (**GT Revision**). Our tool incorporates metadata types that fosters annotation review by other

¹ A demo version of GTCreator is available at <http://www.cvc.uab.es/CVCColon/>.

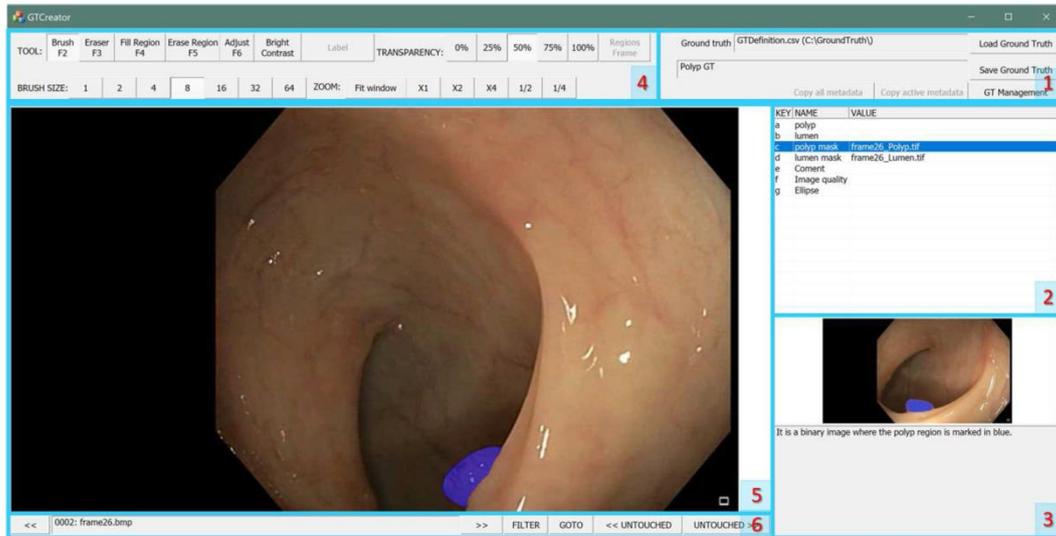


Fig. 2 Distribution of the different areas of GTCreator graphical user interface as described in Sect. 2.1.2

users. In an example case, a novice clinician marks mark an image for later inspection by an expert by using a Boolean metadata type. The expert can easily navigate through the images marked for inspection by using the filtering capabilities of our tool.

- The CSV file produced as a result of dataset annotation can be used to perform experimental setups for methodology validation. For instance, specific text-based metadata (set by the user or imported from an external source) can be used to split a dataset into training, validation and test sets (**GT Experiment Preparation**).
- Our tool can also be used to assist in method validation (**Result Revision and Annotation**) by using the semantic label metadata type. In this case, the user can assign any of the predefined labels (set in the definition file) to each of the regions of interest provided by a given image processing method. The user can also browse through numerical or textual output values using the filtering functions once these values are imported into the corresponding CSV file.

Comparison study of annotation tools

The development of manual image annotation tools has already been covered in the literature. We can divide existing tools into two groups: domain specific (such as Artemis [10] or iPad [11]) or general (such as LabelMe [7] or VGG image

annotator (VIA) [8]). We present in this section a comparative study between GTCreator and other similar tools.

The inclusion criteria for the comparison have been the free availability of the tool, ease of installation and use by means of the potential final users. Five tools were finally selected: RatSnake [9], LabelMe [7], VGG image annotator (VIA) [8], video image annotation tool (VIAT) [12] and ImageJ [13].

RatSnake is based on the use of a snake model to enable semiautomatic segmentation, being model customization capabilities key to succeeding in providing accurate object masks. LabelMe is a web-based tool allowing individual region annotation (which needs no installation), being its usability compromised by requiring a network connection. VIA is another web-based tool which stands out for its simplicity in use, including the possibility of adding formatted text annotations. VIAT offers great flexibility in the type of image annotations to be created, ranging from polygons to freehand annotations, supporting semiautomatic segmentation capabilities. Finally, ImageJ offers a great variability in output formats and incorporates a full image processing suite to support the annotation process.

Qualitative comparison

Table 1 presents a qualitative comparison of the mentioned annotation tools. We describe next some of the most relevant differences among those tools.

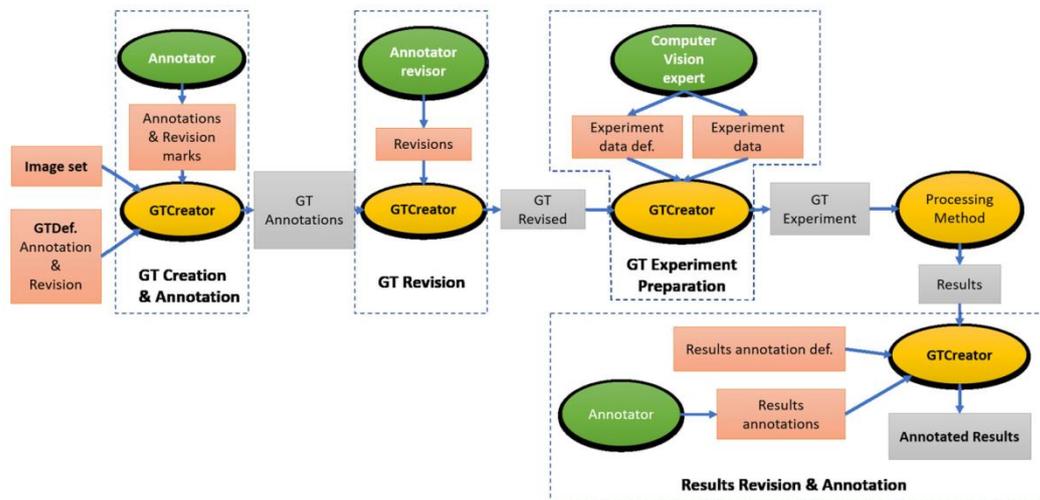


Fig. 3 Graphical scheme showing all the potential interactions of GTCreator through a complete validation of an image processing method

A desirable feature of an annotation tool is to allow users a fast and easy navigation experience through the images to be annotated. With respect to this, we can clearly divide the tools into two groups: those which allow the creation of image collections (LabelMe, VIA and GTCreator) and those which require the user to manually open images and saving results. It is clear that, for the case of large dataset annotation, the use of tools from the second group is less practical since a large amount of the annotation time would be devoted to tasks outside actual image annotation.

Regarding the type of annotations that can be generated, VIA and GTCreator are the only tools allowing the inclusion of formatted text metadata. LabelMe allows to incorporate semantic labels to the objects marked by the annotator and add unformatted comments for each annotation. Finally, VIAT allows to include a general unformatted comment for each image to be annotated. Despite other tools offering the possibility of including text annotations, GTCreator is the only one allowing browsing the dataset using filters defined according to text metadata values.

All tools allow image annotation using predetermined shapes, but only three of them (VIAT, ImageJ and GTCreator) allow the user to draw freely the contour of the object to be annotated. The use of predetermined shapes allows a faster dataset annotation at the cost of losing annotation precision, which we consider is key when validating segmentation methods. Our tool is the only one supporting complete pixel-wise editing of image annotations, whereas other approaches simply allow slight polygon points displacement.

Some tools (RatSnake, VIA and GTCreator) allow annotation transfer among images, which is specially useful when

providing image annotations in video datasets in which variability between consecutive frames tends to be low. GTCreator is the only tool allowing text metadata transfer between images.

The majority of tools accept all common image input formats with the exception of LabelMe and VIAT with respect to TIFF images, which needs of image format conversion. On ground truth data exportation, only GTCreator and RatSnake directly provide as a result the binary ground truth masks. The rest of tools require further file editing operation, which are not always immediate for the final user. Both GTCreator and VIA allow users to also obtain CSV files with the full annotation results (text metadata values and mask annotation image names).

Finally, some of the tools include additional capabilities aiming to provide an extra value to the annotation process. For instance, RatSnake allows the creation of a semantic ontology for the annotations and VIAT provides directly the output of MPEG-7 descriptors for the target image. With respect to image processing capabilities, LabelMe incorporates semiautomatic segmentation capabilities, whereas ImageJ includes a full image processing suite. Furthermore, the use of some of these processing capabilities requires excessive user training which might prevent less experienced users from actually taking profit from them. GTCreator incorporates some basic image processing capabilities such as the possibility of adjusting free handmade contours to actual image contours using watershed segmentation with markers or the possibility to change image contrast locally.

Table 1 Comparison of mean and SD of the annotation precision with respect to the annotation tool

Feature	RatSnake	LabelMe	VIA	VIAT	ImageJ	GTCreator
Annotation types	Image masks	Image masks, semantic labels	Image masks, formatted text	Image masks, unformatted text	Image masks	Image mask, formatted text
Mask annotation	Polygon	Polygon	Polygon, predetermined shapes	Polygon, predetermined shapes, freehand	Polygon, predetermined shapes, freehand	Polygon, predetermined shapes, freehand
Mask editing	Annotation transfer	None	Annotation transfer	None	Annotation transfer	Annotation transfer, pixel-wise editing
Dataset browsing	Single image	Collection	Collection	Single image	Single image	Collection
Input format	BMP, JPEG, PNG, TIFF	BMP, JPEG, PNG	BMP, JPEG, PNG, TIFF, CSV file	BMP, JPEG, PNG	BMP, JPEG, PNG, TIFF	BMP, JPEG, PNG, TIFF
Output format	Binary masks	XML file	CSV file	XML file	Text file	Binary masks, CSV file
Extra features	Semantic ontology	Semiautomatic segmentation	None	MPEG-7 descriptors	Image processing suite	Filtering-based browsing, annotation merging and reviewing

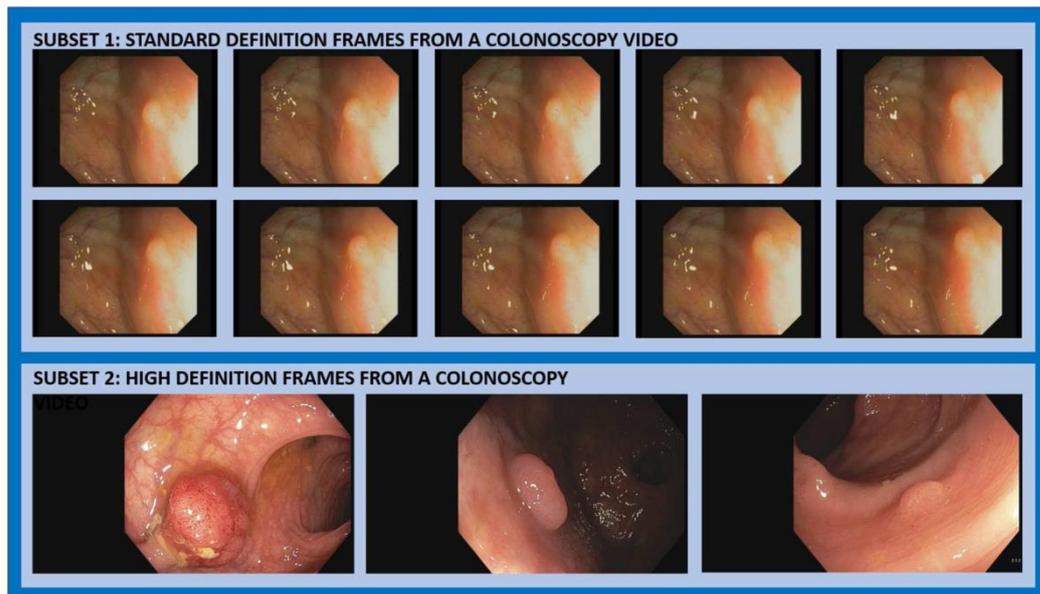


Fig. 4 Content example of a set of images used in the comparison study

Quantitative comparison

Experimental Setup

To explore practical differences in image annotation, we performed an experiment in which 6 different experts were asked to annotate 6 different sets of images, extracted from the datasets of GIANA challenge.² Each set consists of 13 different images divided into two subsets: ten consecutive standard definition frames from a colonoscopy video, and three high-definition images. All images contain a polyp, and each set was chosen to present a similar difficulty level. The annotation task consists of generating a mask covering the polyp. Figure 4 shows a set of images used in the experiment.

The annotation experiment was performed as follows: Each annotator would use a different set of images for each of the tools included in the comparison. Those sets were shared across all annotators, and they were randomly assigned to each tool so that no set and tool would be repeated for two different annotators. Each of the tools was presented to the annotators before the start of the experiment, and during the experiment, support was given if necessary to explain where the boundaries of the polyps were.

Quantitative comparison was based on the measurement of annotation time and precision. For annotation time, a sep-

arate measurement was done to account for the time spent in the interaction with the system, like dataset browsing. All annotators performed the experiment under the same equipment. An expert clinician, not participating in the annotation experiment, was asked to provide pixel-wise annotations of the different images using common image editing software. The different annotations generated in the experiment were compared to this clinical ground truth using common segmentation metrics (DICE coefficient, Jaccard index).

Experimental results

Annotation time results with respect to the annotation tool can be found in Table 2. We can observe a big difference in mean total annotation time between the group of tools that include the use of image collections (LabelMe, VIA and GTCreator) with respect to the ones that do not. This latter group of tools (RatSnake, VIAT and ImageJ) all presents a mean interaction time per image higher than 28 s. As results show, some of these tools would benefit greatly if this interaction time was removed. As an example, mean total annotation time for ImageJ would be reduced to its half.

GTCreator appears as the tool with a lowest mean total annotation time, followed by VIA. It is worth to mention that this difference is higher if we only consider video sequence subsets. Both GTCreator and VIA benefit from including annotation transfer capabilities between consecutive frames,

² Available at <https://giana.grand-challenge.org/>.

Table 2 Annotation time with respect to the annotation tool

	RatSnake	LabelMe	VIA	VIAT	ImageJ	GTCreator
Total annotation time (mean \pm SD)						
All	13'18" \pm 125"	6'3" \pm 88"	4'54" \pm 54"	12'9" \pm 96"	10'11" \pm 81"	4'52" \pm 56"
VidFr	9'38" \pm 120"	4'12" \pm 75"	3'10" \pm 51"	8'30" \pm 88"	7'35" \pm 68"	2'59" \pm 54"
HDFr	3'38" \pm 19"	1'52" \pm 27"	1'44" \pm 33"	3'48" \pm 30"	2'40" \pm 22"	1'43" \pm 14"
Interaction time per image (mean \pm SD)						
	27" \pm 2.4	0" \pm 0"	0" \pm 0 s	29" \pm 4.1"	28" \pm 3.4"	0" \pm 0 s
Actual image annotation time (minutes and seconds)						
All	7'31" \pm 105"	6'3" \pm 88"	4'54" \pm 54"	5'51" \pm 75"	4'13" \pm 76"	4'52" \pm 56"
VidFr	5'12" \pm 107"	4'12" \pm 75"	3'10" \pm 51"	3'40" \pm 70"	3'01" \pm 66"	2'59" \pm 54"
HDFr	2'18" \pm 13"	1'52" \pm 27"	1'44" \pm 33"	2'13" \pm 21"	1'17" \pm 23"	1'43" \pm 14"

VidFr stands for video sequence frames, HDFr for still HD images. Bold values refer to the best performance for each category. In this case, the lowest annotation time

Table 3 Mean and SD of the annotation time with respect to the set to be annotated

	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6
All	8'14" \pm 199"	10'16" \pm 280"	7'50" \pm 218"	7'36" \pm 223"	8'41" \pm 267"	8'50" \pm 217"
VidFr	5'27" \pm 158"	7'54" \pm 216"	5'32" \pm 173"	5'10" \pm 174"	5'49" \pm 203"	6'10" \pm 165"
HDFr	2'37" \pm 56"	2'27" \pm 68"	2'18" \pm 50"	2'25" \pm 53"	2'51" \pm 67"	2'39" \pm 56"

VidFr stands for video sequence frames, HDFr for still HD images. Bold results point to highest values of annotation time which we associate with higher difficulty

which points into this feature to be key for any given annotation tool.

In order to better understand the reason behind differences between tools, we present results with respect to sets instead of annotators in Table 3. We observe that set 2 of the video sequences appears to be the most difficult one as it has the largest mean annotation time. The small differences between mean annotation time among still frame subsets indicate a similar difficulty level.

Finally, we show in Fig. 5 a breakdown of the mean annotation time per tool and video sequence subset. The graph shows how GTCreator is the tool with smaller mean annotation time in 3 out of 6 subsets, followed by VIA.

Annotation precision results with respect to the annotation tool can be found in Table 4. We observe that GTCreator is the tool which achieves the highest score in both DICE and Jaccard metrics, showing the robustness in annotation quality when generated by our tool. Differences between GTCreator and the rest of the tools increase if we only consider video sequence subsets. We associate these differences to the intensive use of the annotators of pixel-wise mask editing capabilities offered by GTCreator, especially for the case of small polyps. In this case, a precise delimitation of the polyp contour can be more difficult to achieve by polygon approximation offered by the other tools. Differences among tools are smaller when only still frames are considered.

As for annotation time, we study in Table 5 the dependence of annotation precision with respect to the specific subset to be annotated. We observe how video sequence of subset 6

appears to be the most difficult, whereas difficulty appears to be balanced between still frame subsets.

Finally, we show in Fig. 6 a breakdown of the mean DICE score per tool and video sequence subset. The graph shows how GTCreator is the tool with highest mean DICE score in all subsets, being this difference remarkable for the case of the mentioned subset 6.

We present in Fig. 7 a graphical comparison of the annotations generated with the different proposed tools. We have chosen three different representative examples: First row shows annotation results of an image belonging to a subset in which all annotation tools lead to achieve good annotations. Second rows shows an example of an image from subset 6 in which all annotators found difficulties when delimiting the area occupied by the polyp. Finally, the last row shows an example of HD image annotation.

From these examples, we can observe that indeed there are no big differences in annotation quality when the polyp appears clearly defined in the image with respect to the example image from subset 6. In this case, we can observe clear discrepancies between annotators and the external expert which provided the ground truth, which suggest that both having several annotators for a same image and annotation review and editing capabilities are key to generate high-quality annotations of image datasets. Finally, HD image annotation task allows us to observe better the limitation of some tools when providing high-quality pixel-wise annotations. We can observe steep contour direction changes in those tools which allow less flexibility in mask generation

Fig. 5 Mean annotation time with respect to image subset and annotation tool

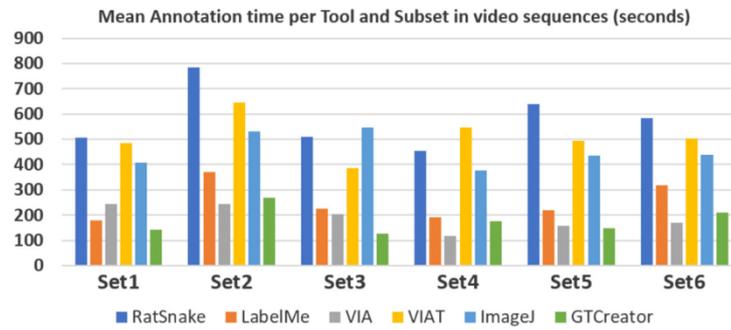


Table 4 Mean and SD of the annotation precision with respect to the annotation tool

	RatSnake	LabelMe	VIA	VIAT	ImageJ	GTCreator
DICE score (mean ± SD)						
All	0.886 ± 0.040	0.860 ± 0.080	0.876 ± 0.059	0.848 ± 0.068	0.852 ± 0.082	0.908 ± 0.032
VidFr	0.870 ± 0.058	0.851 ± 0.103	0.862 ± 0.076	0.824 ± 0.089	0.832 ± 0.099	0.899 ± 0.037
HDFr	0.935 ± 0.026	0.917 ± 0.022	0.924 ± 0.015	0.927 ± 0.038	0.920 ± 0.031	0.937 ± 0.029
Jaccard index (mean ± SD)						
All	0.938 ± 0.025	0.925 ± 0.051	0.932 ± 0.035	0.915 ± 0.042	0.917 ± 0.052	0.951 ± 0.017
VidFr	0.929 ± 0.035	0.916 ± 0.064	0.924 ± 0.046	0.901 ± 0.056	0.905 ± 0.063	0.946 ± 0.002
HDFr	0.965 ± 0.009	0.956 ± 0.013	0.960 ± 0.008	0.960 ± 0.024	0.958 ± 0.017	0.967 ± 0.010

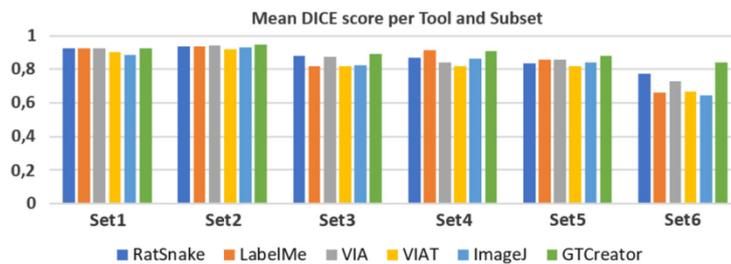
VidFr stands for video sequence frames, HDFr for still HD images. Bold values refer to the highest scores in annotation precision for each of the metrics and annotation sets

Table 5 Mean and SD of the annotation precision with respect to the set to be annotated

	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6
DICE score (mean ± SD)						
All	0.921 ± 0.011	0.928 ± 0.012	0.871 ± 0.027	0.887 ± 0.026	0.869 ± 0.014	0.761 ± 0.058
VidFr	0.915 ± 0.014	0.936 ± 0.008	0.812 ± 0.032	0.869 ± 0.033	0.847 ± 0.019	0.720 ± 0.071
HDFr	0.939 ± 0.014	0.900 ± 0.023	0.937 ± 0.007	0.942 ± 0.011	0.940 ± 0.007	0.907 ± 0.026
Jaccard index (mean ± SD)						
All	0.959 ± 0.006	0.962 ± 0.007	0.934 ± 0.014	0.939 ± 0.015	0.929 ± 0.008	0.861 ± 0.038
VidFr	0.955 ± 0.007	0.967 ± 0.005	0.919 ± 0.019	0.929 ± 0.018	0.917 ± 0.011	0.839 ± 0.004
HDFr	0.969 ± 0.006	0.945 ± 0.015	0.967 ± 0.004	0.970 ± 0.006	0.969 ± 0.004	0.948 ± 0.013

VidFr stands for video sequence frames, HDFr for still HD images. Bold results point to lowest values for each of the metrics which we associate with higher difficulty

Fig. 6 Mean DICE score with respect to image subset and annotation tool



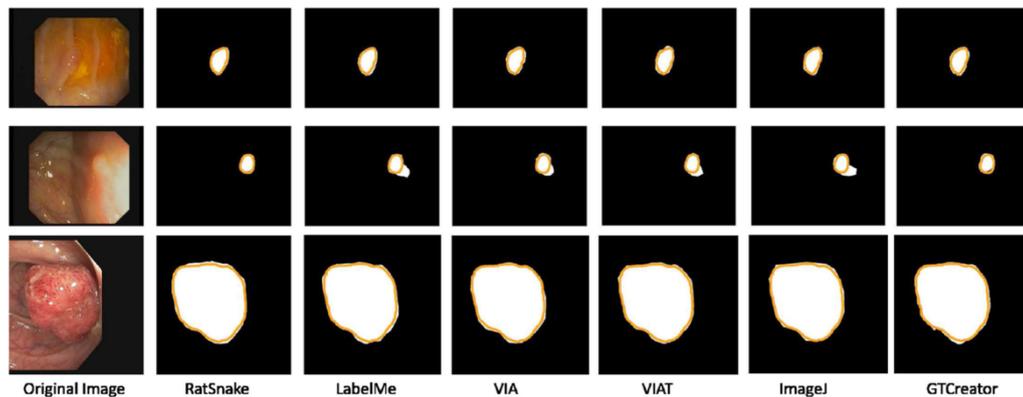


Fig. 7 Comparison between annotations performed with the different tools and the ground truth provided by an external expert. In annotation images, pixels in white represent those marked by the annotators, whereas the orange contour represents the ground truth for the particular image

using predetermined shapes such as VIAT or ImageJ; in these cases, the final mask is obtained by joining the different points marked by the user requiring a large amount of user interaction by marking several consecutive points to obtain a smoother final contour.

Conclusion and perspectives

Evaluating decision support systems for health is a challenging task, mainly due to the lack of public annotated datasets. Video annotation by clinicians is a high time-consuming task as precise masks have to be provided for thousand of images. We have proposed an annotation tool designed to ease annotation by clinicians while keeping the functionalities of other existing tools. Flexibility, efficient data management and browsing capabilities allow our tool to be used during the main stages of method evaluation for any image domain.

We have performed a comparison study to rank our tool among similar ones with the conclusion of GTCreator being the tool which offers the best compromise between annotation time and precision. Easy image browsing and the inclusion of image editing capabilities play a key role in the generation of fast and precise annotations.

As future work, we plan to extend GTCreator by incorporating image processing tools to support annotation tasks as well as video annotation capabilities. We plan to use GTCreator to propose new benchmarks in colonoscopy image analysis such as polyp segmentation in high-definition images or to extend the current one on automatic histology prediction.

Funding This work has been funded by Spanish Government through iVENDIS (DPI2015-65286-R), DeepMTL (TIN2016-79717-R) and

HISINVIA(PI17/00894) projects, Catalan government through SGR-2017-1669, SGR-2017-653 and CERCA programme, Région Île de France through SATT funding “iPolyp” (Project 184). A. Histace and J. Bernal acknowledge the Institute of Advanced Studies from UCP (Invited Prof. Position grant) as well as Initiative Paris Seine through which the position was obtained in the context of “iPolyp”. M. Masana acknowledges 2017FIB-00218 grant of Generalitat de Catalunya. We also acknowledge the generous GPU support from NVIDIA.

Compliances with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. El-Regaily SA, Salem MA, Abdel Aziz MH, Roushdy MI (2018) Survey of computer aided detection systems for lung cancer in computed tomography. *Curr Med Imaging Rev* 14(1):3–18
2. Gui L, Yang X (2018) Automatic renal lesion segmentation in ultrasound images based on saliency features, improved lbp, and an edge indicator under level set framework. *Med Phys* 45(1):223–235
3. Sánchez FJ, Bernal J, Sánchez-Montes C, de Miguel CR, Fernández-Esparrach G (2017) Bright spot regions segmentation and classification for specular highlights detection in colonoscopy videos. *Mach Vis Appl* 28(8):917–936
4. Mori Y, Kudo Se, Berzin TM, Misawa M, Takeda K (2017) Computer-aided diagnosis for colonoscopy. *Endoscopy* 49(08):813–819

5. Bernal J, Tajkbaksh N, Snchez FJ, Matuszewski BJ, Chen H, Yu L, Angermann Q, Romain O, Rustad B, Balasingham I, Pogorelov K, Choi S, Debard Q, Maier-Hein L, Speidel S, Stoyanov D, Brandao P, Crdova H, Snchez-Montes C, Gurudu SR, Fernndez-Esparrach G, Dray X, Liang J, Histace A (2017) Comparative validation of polyp detection methods in video colonoscopy: results from the miccai 2015 endoscopic vision challenge. *IEEE Trans Med Imaging* 36(6):1231–1249
6. Chen PJ, Lin MC, Lai MJ, Lin JC, Lu HHS, Tseng VS (2018) Accurate classification of diminutive colorectal polyps using computer-aided analysis. *Gastroenterology* 154(3):568–575
7. Russell BC, Torralba A, Murphy KP, Freeman WT (2008) Labelme: a database and web-based tool for image annotation. *Int J Comput Vis* 77(1):157–173
8. Dutta A, Gupta A, Zissermann A (2016) VGG image annotator (VIA). <http://www.robots.ox.ac.uk/~vgg/software/via/>. Accessed 21 Jan 2018
9. Iakovidis D, Goudas T, Smailis C, Maglogiannis I (2014) Ratsnake: a versatile image annotation tool with application to computer-aided diagnosis. *Sci World J* 2014:1–12
10. Liu D, Cao Y, Kim KH, Stanek S, Dounggratanaex-Chai B, Lin K, Tavanapong W, Wong J, Oh J, De Groen PC (2007) Arthemis: annotation software in an integrated capturing and analysis system for colonoscopy. *Comput Methods Programs Biomed* 88(2):152–163
11. Rubin DL, Rodriguez C, Shah P, Beaulieu C (2008) ipad: semantic annotation and markup of radiological images. In: AMIA annual symposium proceedings, vol 2008. American Medical Informatics Association, p 626
12. Several authors (2013) Video image annotation tool. <https://sourceforge.net/projects/via-tool/>. Accessed 22 Jan 2018
13. Abràmoff MD, Magalhães PJ, Ram SJ (2004) Image processing with imagej. *Biophoton Int* 11(7):36–42

ARTÍCULO 2:

TÍTULO

Diagnóstico in vivo de pólipos colorrectales asistido por computador mediante endoscopia de luz blanca

OBJETIVO

Desarrollar y evaluar un sistema de clasificación automática de pólipos en tiempo real e in vivo utilizando solo imágenes de luz blanca de alta definición y comparar su rendimiento con el de los endoscopistas.

RESUMEN

Antecedentes y objetivos: La inteligencia artificial es capaz de predecir la histología de los pólipos de colon. Sin embargo, los sistemas desarrollados hasta la fecha utilizan tecnología óptica compleja y se han probado ex vivo.

Objetivo: Evaluar la eficacia de un nuevo sistema de clasificación automática basado en aprendizaje profundo, ATENEA, en un entorno clínico real utilizando solo luz blanca de alta definición y comparar su desempeño con el de los endoscopistas.

Métodos: Estudio de validación de ATENEA in vivo con pólipos consecutivos detectados en colonoscopias de cribado en el Hospital Clínic. La predicción histológica de ATENEA (adenoma vs no adenoma) se comparó con la predicción de cuatro endoscopistas del staff, ciegos al resultado de ATENEA y sin formación específica en diagnóstico óptico para este estudio. La histología fue el patrón oro.

Resultados: Se analizaron 90 pólipos de 31 pacientes [tamaño mediano: 5 mm, 69 adenomas (76,7 %)]. ATENEA predijo correctamente la histología de 63/69 adenomas (91.3%, IC del 95%: 82–97%) y 12/21 no adenomas (57.1%, IC del 95%: 34–78%) mientras los endoscopistas acertaron en 52/69 (75.4%, IC del 95%: 60–85%) y 20/21 (95,2%, IC del 95%: 76–100%), respectivamente. La exactitud diagnóstica global fue del 83,3% (IC del 95%: 74–90%) para ATENEA y del 80% (IC del 95%: 70–88%) para los endoscopistas.

Conclusión: ATENEA obtiene una predicción precisa de la histología de los pólipos in vivo utilizando solo endoscopia de luz blanca y ofrece un rendimiento similar a los endoscopistas a pesar de clasificar peor los no adenomas.

In vivo computer-aided diagnosis of colorectal polyps using white light endoscopy



Authors

Ana García-Rodríguez¹, Yael Tudela², Henry Córdova^{1,3,4}, Sabela Carballal^{1,3,4}, Ingrid Ordás^{1,3,4}, Leticia Moreira^{1,3,4}, Eva Vaquero^{1,3,4}, Oswaldo Ortiz¹, Liseth Rivero^{1,3,4}, F. Javier Sánchez², Miriam Cuatrecasas^{3,4,5}, Maria Pellisé^{1,3,4}, Jorge Bernal², Glòria Fernández-Esparrach^{1,3,4}

Institutions

- 1 Endoscopy Unit, Gastroenterology Department, ICMDiM, Hospital Clínic of Barcelona, University of Barcelona, Barcelona, Catalonia, Spain
- 2 Computer Science Department, Autonomous University of Barcelona and Computer Vision Center, Barcelona, Catalonia, Spain
- 3 IDIBAPS, Barcelona, Catalonia, Spain
- 4 CIBEREHD, Spain
- 5 Pathology Department, Hospital Clínic of Barcelona, University of Barcelona, Barcelona, Catalonia, Spain

submitted 1.11.2021

accepted after revision 10.6.2022

Bibliography

Endosc Int Open 2022; 10: E1201–E1207

DOI 10.1055/a-1881-3178

ISSN 2364-3722

© 2022. The Author(s).

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

Corresponding author

Glòria Fernández-Esparrach, MD, PhD, Endoscopy Unit, Institut de Malalties Digestives, Hospital Clínic, Villarroel 170, 08036 Barcelona, Spain

Fax: +34 93 227 93 87

mgfernan@clinic.cat

ABSTRACT

Background and study aims Artificial intelligence is currently able to accurately predict the histology of colorectal polyps. However, systems developed to date use complex optical technologies and have not been tested in vivo. The objective of this study was to evaluate the efficacy of a new deep learning-based optical diagnosis system, ATENEA, in a real clinical setting using only high-definition white light endoscopy (WLE) and to compare its performance with endoscopists.

Methods ATENEA was prospectively tested in real life on consecutive polyps detected in colorectal cancer screening colonoscopies at Hospital Clínic. No images were discarded, and only WLE was used. The in vivo ATENEA's prediction (adenoma vs non-adenoma) was compared with the prediction of four staff endoscopists without specific training in optical diagnosis for the study purposes. Endoscopists were blind to the ATENEA output. Histology was the gold standard.

Results Ninety polyps (median size: 5 mm, range: 2–25) from 31 patients were included of which 69 (76.7%) were adenomas. ATENEA correctly predicted the histology in 63 of 69 (91.3%, 95% CI: 82%–97%) adenomas and 12 of 21 (57.1%, 95% CI: 34%–78%) non-adenomas while endoscopists made correct predictions in 52 of 69 (75.4%, 95% CI: 60%–85%) and 20 of 21 (95.2%, 95% CI: 76%–100%), respectively. The global accuracy was 83.3% (95% CI: 74%–90%) and 80% (95% CI: 70%–88%) for ATENEA and endoscopists, respectively.

Conclusion ATENEA can accurately be used for in vivo characterization of colorectal polyps, enabling the endoscopist to make direct decisions. ATENEA showed a global accuracy similar to that of endoscopists despite an unsatisfactory performance for non-adenomatous lesions.

Introduction

Colorectal cancer (CRC) is the third most common cancer in both sexes and the second leading cause of death in the world [1]. Screening for CRC and the removal of neoplastic polyps by

colonoscopy has led to a substantial improvement in survival [2].

Optical diagnosis aims to predict the histology of a polyp based on its endoscopic features. This practice could avoid pathological analysis in several cases and reduce the derived

costs. The American Society for Gastrointestinal Endoscopy's PIVI working group established diagnostic thresholds for real-time implementation of optical diagnosis for diminutive polyps (≤ 5 mm) [3]. However, PIVI criteria have not yet been met in community-based practices or in non-expert hands [4,5]. In this regard, the European Society of Gastrointestinal Endoscopy (ESGE) emphasizes the importance of being able to ensure and maintain competence in optical diagnosis as well as considering only the proportion of high-confidence diagnosis as a benchmark [6].

During the past few decades, considerable technological advances have been made in the application of artificial intelligence (AI) to medicine. Computer-aided diagnosis (CADx) is a promising solution to overcome human variation in characterization of polyps by providing decision support. In this specific field, CADx approaches based on deep learning (DL) represent an advantage over previous machine learning by combining both the automatic extraction and classification of image characteristics using a multilayered system called convolutional neural networks (CNNs) [7].

The quality and design of published CADx systems has varied over time. Initial studies were carried out retrospectively and were tested *ex vivo* using selected stored images [8–12]. More recent studies, most of them DL approaches, have been prospectively conducted reporting higher accuracy with faster processing times, which allows diagnosis in real time [13–16]. However, to date, only CADx systems using complex optical technologies as endocytoscopy have been tested *in vivo* [13], that is, while the colonoscopy is being performed, which still hinders the adoption of this technology in daily practice. Furthermore, the newest CADx systems are currently using advanced imaging modalities, which clearly limits their implementation worldwide.

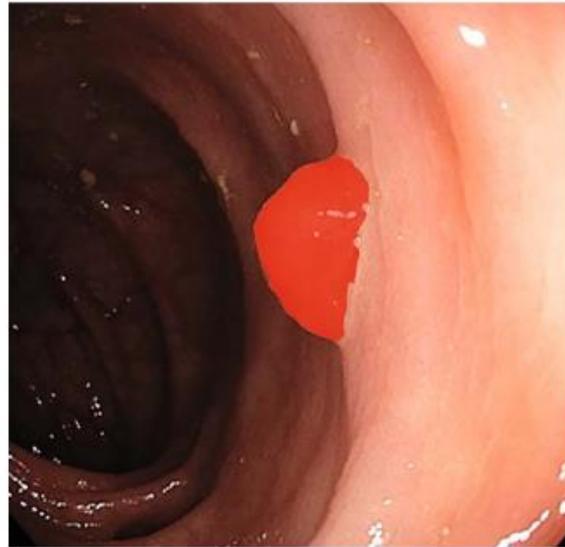
The aim of this study was to evaluate the efficacy of a new CADx system based on DL called ATENEA for *in vivo* optical diagnosis in consecutive patients using only white light endoscopy (WLE) and compare its performance with endoscopists.

Methods

Development of ATENEA

As a first step, images of any polyp from routine colonoscopies performed at Hospital Clínic of Barcelona from January 2016 to December 2020 were prospectively collected. The images were captured with high-definition colonoscopes (CF-HQ180, CF-HQ-185, CF-HQ-190 and EVIS EXERA III videoprocessor, Olympus Europe, Hamburg, Germany) using an external computer with a frame grabber to ensure image acquisition with the highest quality. Only white light images without magnification or chromoendoscopy were used. Data from morphology, location, and size of the polyp were collected and periodically transferred by an assistant into the database along with the histological category of the lesion obtained after pathological analysis.

The development of ATENEA consisted of two stages: feature extraction and image classification. Feature extraction was made by using a faster region-based CNN with ResNet50 as backbone. For system training purposes, a region of interest



► **Fig. 1** Example of a manually defined region of interest (ROI) delineating the polyp using GTCreator in the training phase.

(ROI) delineating the polyp was manually defined by clinicians using a program called GTCreator (► **Fig. 1**) [17]. The extracted features, ROI, and actual histologic class of the polyp were used to train the system. ATENEA learned to classify images into adenoma or non-adenoma categories using an 80% (high confidence) threshold value (predictions with a confidence value $< 80\%$ were not considered valid to represent the actual performance of the system as they could not guarantee a robust performance).

ATENEA was trained and validated with a total of 1049 high-definition white light images of 483 polyps from 354 patients, with a maximum of three images of the same polyp (but with a different view or perspective). Images had variable quality but all had a visible mucosal pattern and only blurred images and polyps covered by mucus were excluded. Images containing patient data or without histological analysis were also excluded. About two-thirds were adenomas and one-third non-adenomas, following a similar proportion to what is found in real life, with a median size of 4.5 mm (range 2 mm–20 mm). These images were randomly split into training ($n=837$) and validation sets ($n=212$), with the condition that all the images of the same polyp were in the same set.

In vivo experiment

This observational prospective cohort study was performed in Hospital Clínic of Barcelona between January and March 2021 and included individuals from the fecal immunochemical test-based (cut-off of $\geq 20 \mu\text{g}$ of hemoglobin/g of feces) organized population CRC screening program in which all individuals aged 50 to 69 years are invited to participate. Colonoscopies were performed by four staff endoscopists with more than 1000 colonoscopies and adenoma detection rates of 47%,

48%, 51%, and 55%, respectively (average in this program and time period was 49%). No specific training in optical diagnosis was performed for the study purpose. The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of Hospital Clinic of Barcelona (HCB2017/0506, 7/18/2017) and informed consent was obtained from all patients involved in the study. It was also registered in ClinicalTrials.gov (NCT03775811).

All polyps detected and resected with a final pathological report were prospectively included regardless of image quality (not centered polyps, blurred or covered by mucus). Only the image that was considered adequate for the prediction (either by the orientation of the polyp within the image or by the proximity for its correct assessment) was selected by the endoscopist during the exploration and was analyzed. The following variables were collected: estimated size (in mm), location (rectum-sigma, descending, transverse, ascending) and morphology according to Paris classification [18], which were properly noted in the colonoscopy report.

In real time, ATENEA classified each polyp as an adenoma or non-adenoma and provided the confidence value for each prediction. To avoid losing any prediction, we were more flexible than in train/validation phases and the minimum threshold was reduced to 50%. Values between 50% and 80% were considered low-confidence predictions. In this phase, the system was fully automatic (without delineation of ROI), providing for each lesion a bounding box and its corresponding histological class. Time needed to acquire and process each image to obtain the automatic prediction was of 40 milliseconds (real time).

For the optical diagnosis process, the endoscopists were asked to categorize each lesion into two categories (adenoma and non-adenoma) based on its surface pattern and to provide its diagnostic confidence (high vs low) without any time limitation. Endoscopists were blinded to the ATENEA output. The decision was intuitive and based on their previous experience but without using chromoendoscopy or any of the existing classifications. Serrated sessile lesions (SSLs) were included in the non-adenoma category. ► **Fig. 2** shows the set-up in the exploration room.

Histopathology

All polyps were removed using the usual techniques and sent separately for further evaluation to the Pathology Department, which was the gold standard. The diagnosis of dysplasia in neoplastic polyps was made based on the Modified Vienna Classification [19]. The pathologist was blinded to the predictions of both endoscopist and ATENEA.

Statistical analysis

All the polyps were globally analyzed, independently of the endoscopist who performed the exploration. The numbers of polyps that were true positive (adenomatous polyps predicted to be adenomatous), true negative (non-adenomatous polyps predicted to be non-adenomatous), false positive (non-adenomatous polyps predicted to be adenomatous) or false negative (adenomatous polyps predicted to be non-adenomatous) were calculated. Sensitivity, specificity, positive predictive value,



► **Fig. 2** Setting in the endoscopy room showing the position of the assistant sitting in front the computer: the endoscopist is blind to the image displayed in the computer and ATENEA's output.

negative predictive value and accuracy with 95% confidence intervals were calculated. Comparisons of these metrics between ATENEA and endoscopists were performed with Chi squared test.

Receiver operator characteristics (ROC) curves for use of different prediction confidence values to determine ATENEA in vivo performance for diagnosis of adenoma were constructed with Matlab. The area under the curve (AUC) and optimal operating point with its sensitivity and specificity and 95% confidence intervals were calculated.

Two-sided $P < 0.05$ was considered statistically significant. All calculations were performed with R version 4.1.2 and Matlab for Windows version 2020.

Results

During the study period, 90 polyps from 31 consecutive patients were included. Sixty-nine (76.7%) were adenomas with a median size of 5.0 mm (range: 2–25 mm). Characteristics of the polyps are described in ► **Table 1**.

ATENEA provided an output for all 90 polyps and was able to correctly predict the histology in 63 of 69 adenomas (sensitivity: 91.3%, 95% CI: 82%–97%) and 12 of 21 non-adenomas (specificity: 57.1%, 95% CI: 34%–78%). Endoscopists were able to provide a correct optical diagnosis in 52 of 69 adenomas (sensitivity: 75.4%, 95% CI: 64%–85%) and 20 of 21 non-adenomas (specificity: 95.2%, 95% CI: 76%–100%), with an accuracy of 83.3% (95% CI: 74%–90%) and 80% (95% CI: 70%–88%) for CADx and endoscopists, respectively (► **Table 2**).

With respect to diminutive polyps (≤ 5 mm), ATENEA was able to correctly predict the histology in 31 of 35 adenomas (sensitivity: 88.2%, 95% CI: 73%–97%) and 11 of 18 non-adenomas (specificity: 61.1%, 95% CI: 36%–83%) and endoscopists in 20 of 35 adenomas (sensitivity: 58.8%, 95% CI: 41%–75%) and 17 of 18 non-adenomas (specificity: 94.4%, 95% CI: 73%–100%), respectively, with an accuracy of 78.8% (95% CI: 65%–89%) and 71.1% (95% CI: 57%–83%) for ATENEA and

► **Table 1** Characteristics of the polyps included in the study.

In vivo test (n=90)	
Histological type	
• Adenoma	69 (76.7%)
• Non-adenoma	21 (23.3%)
• Hyperplastic	16
• SSL	5
Size (in mm)	
• ≤5 mm	52 (57.8%)
• 6 mm to <10 mm	21 (23.3%)
• ≥10 mm	17 (18.8%)
Location	
• Rectum-sigma	22 (24.4%)
• Others	68 (75.6%)
Morphology	
• 0-Ip	5 (5.5%)
• 0-Is	30 (33.3%)
• 0-IIa/IIb	55 (61.1%)

endoscopists, respectively (► **Table 2**). Results of prediction for small polyps (<10 mm) are also shown in ► **Table 2**.

ATENEA and the endoscopists disagreed in their prediction in 31 of 90 cases (34.5%). Endoscopists made their prediction with high confidence in 79 cases (87%) and with low confidence in 11. In all these 11 cases, ATENEA made a good prediction (► **Fig. 3**).

The ROC curve for ATENEA in vivo performance (► **Fig. 4**) showed an AUC of 0.782. The optimal operating point was achieved by using 74.2% as the threshold value with a sensitivity of 86.9% (95% CI: 79.9%–93.9%) and a specificity of 66.7% (95% CI: 56.9%–76.4%).

Discussion

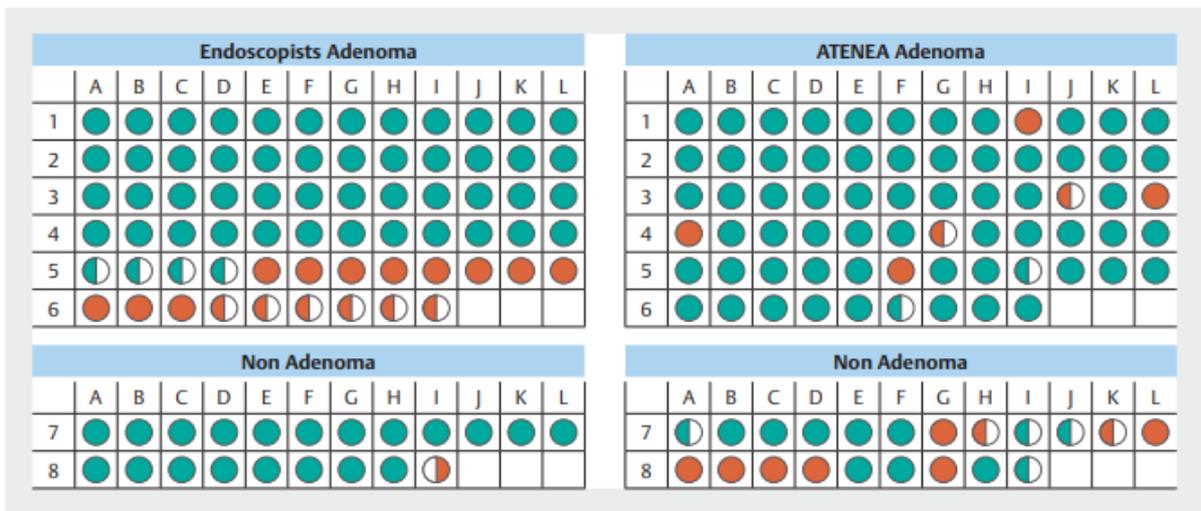
This study demonstrates that ATENEA, a fully automatic optical diagnostic system, can accurately predict polyp histology in an in vivo setting using only white light, which answers one of the 10 key research questions identified by international experts related to AI implementation in colonoscopy [20]. ATENEA is particularly useful for identifying adenomatous lesions, showing its readiness to be used in a clinical environment.

Until now, the majority of CAD systems developed for characterizing colorectal polyps that use AI have required advanced optical diagnostic equipment, such as narrow band imaging (NBI) [9–12, 21] and endocytoscopy [8, 22, 23]. However, WLE, which is the most common endoscopic modality, has not been

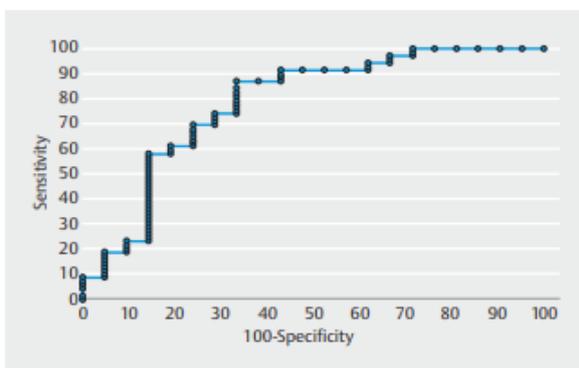
► **Table 2** Performance characteristics of ATENEA and the endoscopists for diagnosis of adenoma.

	TP	FP	TN	FN	PPV	SENS	NPV	Spec	Accuracy
ATENEA, n=90	63	9	12	6	87.5% (95% CI: 78%–94%)	91.3% (95% CI: 82%–97%)	66.7% (95% CI: 41%–87%)	57.1% (95% CI: 34%–78%)	83.3% (95% CI: 74%–90%)
Endoscopists, n=90	52	1	20	17	98.1% (95% CI: 90%–100%)	75.4% (95% CI: 64%–85%)	54.0% (95% CI: 37%–71%)	95.2% (95% CI: 76%–100%)	80.0% (95% CI: 70%–88%)
P value					0.07	0.02	0.55	0.01	0.7
ATENEA diminutive polyps ≤5 mm, n=52	30	7	11	4	81.1% (95% CI: 65%–92%)	88.2% (95% CI: 73%–97%)	73.3% (95% CI: 45%–92%)	61.1% (95% CI: 36%–83%)	78.8% (95% CI: 65%–89%)
Endoscopists diminutive polyps ≤5 mm, n=52	20	1	17	14	95.2% (95% CI: 76%–100%)	58.8% (95% CI: 41%–75%)	54.8% (95% CI: 36%–73%)	94.4% (95% CI: 73%–100%)	71.1% (95% CI: 57%–83%)
P value					0.27	0.01	0.38	0.04	0.5
ATENEA small polyps <10 mm, n=73	47	9	12	5	83.9% (95% CI: 72%–92%)	90.4% (95% CI: 79%–97%)	70.6% (95% CI: 44%–90%)	57.2% (95% CI: 34%–78%)	80.8% (95% CI: 70%–89%)
Endoscopists small polyps <10 mm, n=73	36	1	20	16	97.3% (95% CI: 86%–100%)	69.2% (95% CI: 55%–81%)	55.5% (95% CI: 38%–72%)	95.2% (95% CI: 76%–100%)	76.7% (95% CI: 65%–86%)
P value					0.09	0.02	0.46	0.01	0.69

TP, true positive; FP, false positive; TN, true negative; FN, false negative; PPV, positive predictive value; SENS, sensitivity; NPV, negative predictive value; SPEC, specificity.



► Fig. 3 ATENEA and endoscopists' predictions for all polyps. Each circle represents a polyp and colors correspond to a correct prediction (green) and incorrect prediction (red) with high confidence (full circle) or low confidence (half circle).



► Fig. 4 Receiver operating characteristic (ROC) curve for different prediction confidence values for ATENEA. The optimal operating point (defined as the point in the curve with better balance of specificity and sensitivity) was achieved by using 74.2% as the threshold value.

extensively studied in this context yet. For this reason, we used only white light images to allow for widespread use of ATENEA regardless of the manufacturer and model of endoscope used.

Our group previously developed a hand-crafted predictive model based on extraction of surface patterns (textons) over white light images. This system was validated in a small dataset containing only 225 high-quality images with a diagnostic accuracy of over 90% [24]. Unfortunately, the system was not fully automatic and could not perform in real time and, consequently, it could not be used in an *in vivo* experiment. In contrast, ATENEA can perform *in vivo* automatic polyp classification due to its capability of calculating the result in less than 40 ms (real time). It is important to make a difference between what real-time and *in vivo* means: the first refers to the time the

system takes to process an image, whereas the second is related to where and when the system is applied (*in vivo*: in the exploration room versus *ex vivo*: off-line experiments). These concepts are commonly confused in the literature and, in most of the cases, real-time is used instead of *in vivo* without mentioning the actual processing time [13, 14].

The results obtained in the present experiment showed that ATENEA had better global accuracy than endoscopists, particularly for adenomatous lesions. Conversely, clinicians performed better for the non-adenoma category. It has to be pointed out that they had more information than the system because they knew the location and size of the lesion (variables for which ATENEA was blind), so this could lead to a major pretest probability of better diagnosis of non-adenomatous lesions when faced with a small or diminutive polyp in the rectosigmoid colon. In contrast, ATENEA performance depended solely on the number and variability of examples from each of the classes in the training set.

The lower performance of ATENEA for non-adenomatous lesions could limit implementation of a "leave in situ" policy. However, it must be stated that the training of the AI system was greatly affected by the reduced number of examples of this class in the dataset. There are two reasons for this: first, the prevalence of non-adenomatous lesions is generally lower and second, they are not systematically removed when located in the rectosigmoid colon, making their collection difficult. Hence, it is necessary to enlarge the dataset both in quantity and percentage of examples of the minority class.

CADx systems are not intended to replace endoscopists, but rather, to help them in their tasks. As stated by the ESGE review on advanced endoscopic imaging, the most likely scenario is that the intelligent systems will be used as a "second reader" to support the endoscopist's final diagnosis [25]. In this sense, an important result from our study was that if endoscopists

with low confidence had followed the output of the system, all their predictions would have been correct. This shows the potential of ATENEA to assist clinicians with less expertise.

Similar to endoscopists, AI systems do also provide a confidence value in their predictions using a percentage instead of a binary assessment (high versus low). By varying the confidence value, we represented different performances of ATENEA using a ROC curve. Our results show that if the confidence value is low, the sensitivity of the system is lower as it provides more outputs, some of them erroneous. Conversely, if we increase the allowed confidence value, less but more confident outputs will be provided, with the risk of missing some of the lesions without a prediction. The so-called Simple Optical Diagnosis Accuracy or SODA criteria, which were recently published by the ESGE Curricula Working Group [26], are more flexible than the PIVI criteria and emphasize the importance of not leaving any diminutive lesion with advanced neoplasia in situ. In accordance with SODA criteria, we considered sensitivity to be the most important outcome of the intelligent system. The AUC value obtained in our study is similar to the 0.84 reported in a recent meta-analysis using only WLE [27] (which is usually lower than the AUC in studies using chromoendoscopy or magnification techniques).

Unlike other studies in which low-quality images or the known "difficult for AI cases" were not included, the main strength of this study is that it was performed under real clinical practice conditions and polyps were included regardless of image quality (not centered polyps, blurred or covered by mucus). If we had excluded these polyps from the analysis, the ATENEA performance would have been better. Nevertheless, the lack of publicly available annotated datasets does not allow for a fair comparison between ATENEA and other CADx systems. In this sense, comparison of metrics in meta-analysis are difficult to understand as the number, quality, and class of the images are different in each study and affect training and testing stages of the validation of computational systems.

The study has the following limitations. First, the training dataset was small in terms of number of different polyps. To mitigate this, we used pre-trained weights from ImageNet [28], which is a more general dataset and we applied data augmentation operations (color transformation, rotation, horizontal and vertical flip) to enlarge its size. The collaboration with other centers and the public availability of other datasets could also be of great use to both enlarge the dataset and perform multi-center validations, increasing the robustness of the results. Second, our study did not consider a separate class for SSL due to the low number of examples in the dataset. The problem of SSLs has not adequately been addressed in other previous studies. It is well known that SSLs are neoplastic lesions and there is not an ideal optical method for their characterization [29]. Due to the clinical relevance of SSLs, some studies propose a classification between neoplastic and non-neoplastic lesions instead of using adenoma vs. non-adenoma. Following this logic, if we had included all SSL in the same category as adenomas, four of five SSLs would have been correctly classified.

Conclusions

In conclusion, ATENEA, a CADx system based only on WLE data, is ready to be used for in vivo and real-time characterization of colorectal polyps, enabling the endoscopist to make direct decisions. ATENEA achieved a global accuracy similar to endoscopists, despite having lower performance for non-adenomatous lesions.

Acknowledgments

This work was supported by the Instituto de Salud Carlos III (PI17/00894 and PI19/01050), co-funded by the European Union; the Spanish Ministry for Science and Innovation (MCIN) (ALETHEIA project, PID2020-120311RB-I00 funded by MCIN/AEI/10.13039/501100011033); Fundació Marató de TV3 (201932–30); Secretaria d'Universitats i Recerca de la Generalitat de Catalunya (2014-SGR-1470, 2014-SGR-135, SGR-2017–1669 and SGR-2017–653), and by CERCA Programme/Generalitat de Catalunya. Ana García-Rodríguez received a personal grant from the Societat Catalana de Digestologia. The authors thank Dr. Andrés Cárdenas for his English medical writing assistance and Roger Borràs for performing the statistical analysis.

Competing interests

Glòria Fernández-Esparrach was employed on an advisory board of Medtronic and is shareholder of MiWendo Solutions S.L. Miriam Cuatrecasas is shareholder of MiWendo Solutions S.L. Maria Pellisé has served on clinical advisory boards for Fujifilm Europe; has served on the clinical advisory board and owns share options in MiWendo S.L.; reports speaker fees from Casen Recordati, Fujifilm, Medtronic and Olympus Europe; and has received research funding from Fujifilm Europe, Casen Recordati and Ziuz.

Funding

Fundació la Marató de TV3 201932-30
Iniciació a la recerca de la Societat Catalana de Digestologia
Generalitat de Catalunya CERCA Programme
Departament d'Innovació, Universitats i Empresa, Generalitat de Catalunya 2014-SGR-135, 2014-SGR-1470, SGR-2017-1669, SGR-2017-653
Instituto de Salud Carlos III PI17/00894, PI19/01050, PID2020-120611RB-I00
Spanish Ministry for Science and Innovation (MCIN) PID2020-120311RB-I00

Clinical trial

ClinicalTrials.gov
NCT03775811
TRIAL REGISTRATION: Prospective trial at <http://www.clinicaltrials.gov/>

References

- [1] Bray F, Ferlay J, Soerjomataram I et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018; 68: 394–424
- [2] Levin B, Lieberman DA, McFarland B 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. *Gastroenterology* 2008; 134: 1570–1595
- [3] Abu Dayyeh BK, Thosani N, Konda V et al. ASGE technology committee systematic review and meta-analysis assessing the ASGE PIVI thresholds for adopting real-time endoscopic assessment of the histology of diminutive colorectal polyps. *Gastrointest Endosc* 2015; 81: 502–516
- [4] Vu HT, Sayuk GS, Hollander TG et al. Resect and discard approach to colon polyps: real-world applicability among academic and community gastroenterologists. *Dig Dis Sci* 2015; 60: 502–508
- [5] Rees CJ, Rajasekhar PT, Wilson A et al. Narrow band imaging optical diagnosis of small colorectal polyps in routine clinical practice: The Detect Inspect Characterise Resect and Discard 2 (DISCARD 2) study. *Gut* 2016; 66: 887–895
- [6] Dekker E, Houwen BBSL, Puig I et al. Curriculum for optical diagnosis training in Europe: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. *Endoscopy* 2020; 52: C10
- [7] Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; 521: 436–444
- [8] Misawa M, Kudo SE, Mori Y et al. Characterization of colorectal lesions using a computer-aided diagnostic system for narrow-band imaging endocytoscopy. *Gastroenterology* 2016; 150: 1531–1532.e3
- [9] Byrne MF, Chapados N, Soudan F et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut* 2019; 68: 94–100
- [10] Gross S, Trautwein C, Behrens A et al. Computer-based classification of small colorectal polyps by using narrow-band imaging with optical magnification. *Gastrointest Endosc* 2011; 74: 1354–1359
- [11] Kominami Y, Yoshida S, Tanaka S et al. Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy. *Gastrointest Endosc* 2016; 83: 643–649
- [12] Chen PJ, Lin MC, Lai MJ et al. Accurate Classification of diminutive colorectal polyps using computer-aided analysis. *Gastroenterology* 2018; 154: 568–575
- [13] Mori Y, Kudo SE, Misawa M et al. Real-time use of artificial intelligence in identification of diminutive polyps during colonoscopy a prospective study. *Ann Intern Med* 2018; 169: 357–366
- [14] Rodriguez-Diaz E, Baffy G, Lo WK et al. Real-time artificial intelligence-based histologic classification of colorectal polyps with augmented visualization. *Gastrointest Endosc* 2021; 93: 662–670
- [15] Horiuchi H, Tamai N, Kamba S et al. Real-time computer-aided diagnosis of diminutive rectosigmoid polyps using an auto-fluorescence imaging system and novel color intensity analysis software. *Scand J Gastroenterol* 2019; 54: 800–805
- [16] Zachariah R, Samarasekera J, Luba D et al. Prediction of polyp pathology using convolutional neural networks achieves “resect and discard” thresholds. *Am J Gastroenterol* 2020; 115: 138–144
- [17] Bernal J, Histace A, Masana M et al. GTCreator: a flexible annotation tool for image-based datasets. *Int J Comput Assist Radiol Surg* 2019; 14: 191–201
- [18] Lambert R, Lightdale CJ. The Paris endoscopic classification of superficial neoplastic lesions: Esophagus, stomach, and colon – Paris, France November 30 to December 1, 2002. *Gastrointest Endosc* 2003; 58: S3–S43
- [19] Schlemper RJ, Riddell RH, Kato Y et al. The vienna classification of gastrointestinal epithelial neoplasia. *Gut* 2000; 47: 251–255
- [20] Ahmad OF, Mori Y, Misawa M et al. Establishing key research questions for the implementation of artificial intelligence in colonoscopy: A modified Delphi method. *Endoscopy* 2021; 53: 893–891
- [21] Takemura Y, Yoshida S, Tanaka S et al. Computer-aided system for predicting the histology of colorectal tumors by using narrow-band imaging magnifying colonoscopy (with video). *Gastrointest Endosc* 2012; 75: 179–185
- [22] Mori Y, Kudo SE, Chiu PWY et al. Impact of an automated system for endocytoscopic diagnosis of small colorectal lesions: an international web-based study. *Endoscopy* 2016; 48: 1110–1118
- [23] Mori Y, Kudo SE, Mori K. Potential of artificial intelligence-assisted colonoscopy using an endocytoscope (with video). *Dig Endosc* 2018; 30: S2–S3
- [24] Sánchez-Montes C, Sánchez FJ, Bernal J et al. Computer-aided prediction of polyp histology on white light colonoscopy using surface pattern analysis. *Endoscopy* 2019; 51: 261–265
- [25] Bisschops R, East JE, Hassan C et al. Advanced imaging for detection and differentiation of colorectal neoplasia: European Society of Gastrointestinal Endoscopy (ESGE) Guideline – Update 2019. *Endoscopy* 2019; 51: 1155–1179
- [26] Houwen BBSL, Hassan C, Coupé VMH et al. Definition of competence standards for optical diagnosis of diminutive colorectal polyps: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. *Endoscopy* 2022; 54: 88–99
- [27] Lui TKL, Guo CG, Leung WK. Accuracy of artificial intelligence on histology prediction and detection of colorectal polyps: a systematic review and meta-analysis. *Gastrointest Endosc* 2020; 92: 11–22.e6
- [28] Deng J, Dong W, Socher R et al. ImageNet: a Large-Scale Hierarchical Image Database in 2009 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops). Miami, FL: 2009: pp. 248–255 doi: 10.1109/CVPR.2009.5206848 Corpus ID: 57246310
- [29] Bustamante-Balén M, Satorres C, Ramos-Soler D et al. Evaluation of the optical criteria for sessile serrated lesions of the colon: A prospective study on a colorectal cancer screening population. *Endosc Int Open* 2021; 9: E14–E21

ARTÍCULO 3:

TÍTULO

Polyp fingerprint: reconocimiento automático de las características unívocas de los pólipos colorrectales

OBJETIVO

Desarrollar y evaluar un sistema de recuperación de imágenes mediante la extracción de características de pólipos en imágenes de luz blanca que permitan su identificación inequívoca y potencial agrupación en categorías histológicas por analogía

RESUMEN

Antecedentes: La recuperación de imágenes según contenido es una aplicación del aprendizaje computacional que se utiliza para recuperar imágenes por similitud. El objetivo era desarrollar un sistema que pudiera identificar diferentes instantáneas que contuvieran el mismo pólipo.

Métodos: Aplicación de técnica CBIR llamada Bag of Words para describir cada imagen endoscópica con un pólipo de una manera única (huella digital o Polyp fingerprint). El sistema se probó con 243 imágenes de luz blanca pertenecientes a 99 pólipos diferentes de colonoscopias de rutina en el Hospital Clínic y proporcionó para cada imagen la coincidencia más cercana dentro el conjunto de datos.

Resultados: El sistema hizo coincidir otra imagen del mismo pólipo en 221/243 casos (exactitud diagnóstica 91%). No se observaron diferencias significativas en el número de coincidencias correctas según la clasificación de París (protruidas: 90,7% vs no protruidas: 91,3%) ni el tamaño (<10 mm: 91,6% vs >10 mm: 90%).

Conclusiones: Un sistema de recuperación de imágenes puede hacer coincidir con precisión dos imágenes que contienen el mismo pólipo, lo que podría ser una ayuda útil para reconocimiento y clasificación de pólipos.



Polyp fingerprint: automatic recognition of colorectal polyps' unique features

Ana García-Rodríguez¹ · Jorge Bernal² · F. Javier Sánchez² · Henry Córdova¹ · Rodrigo Garcés Durán¹ · Cristina Rodríguez de Miguel¹ · Gloria Fernández-Esparrach¹

Received: 26 May 2019 / Accepted: 28 October 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Background Content-based image retrieval (CBIR) is an application of machine learning used to retrieve images by similarity on the basis of features. Our objective was to develop a CBIR system that could identify images containing the same polyp ('polyp fingerprint').

Methods A machine learning technique called Bag of Words was used to describe each endoscopic image containing a polyp in a unique way. The system was tested with 243 white light images belonging to 99 different polyps (for each polyp there were at least two images representing it in two different temporal moments). Images were acquired in routine colonoscopies at Hospital Clínic using high-definition Olympus endoscopes. The method provided for each image the closest match within the dataset.

Results The system matched another image of the same polyp in 221/243 cases (91%). No differences were observed in the number of correct matches according to Paris classification (protruded: 90.7% vs. non-protruded: 91.3%) and size (≤ 10 mm: 91.6% vs. > 10 mm: 90%).

Conclusions A CBIR system can match accurately two images containing the same polyp, which could be a helpful aid for polyp image recognition.

Keywords Content-based image retrieval · Artificial intelligence · Colorectal polyps

Colonoscopy allows the identification and extirpation of polyps for its later histological analysis [1]. However, depending on the type of polyp and the risk of submucosal invasion, endoscopists may opt for 'resect and discard' (and omit the pathological diagnosis) or 'leave in situ' strategies or if not send it to surgery [2]. That is why it is important that the same polyp could be recognized in two different time instants, in order to differentiate it from others and handle each case properly. The so-called content-based image retrieval (CBIR) techniques could help us to recognize polyps quickly.

CBIR are artificial intelligence (AI) systems that provide as output the most similar image to a given query image. These systems describe images by means of visual features such as color, texture and shape in a way that similar images should have a close representation in the same feature space. CBIR has already been used in medical imaging domain such as lung nodule recognition in CT scans or brain tumor identification in MR images [3, 4].

We propose to study the potential of these systems to distinguish those unique characteristics that could lead to the unambiguous recognition of each colorectal polyp ('polyp fingerprint').

✉ Gloria Fernández-Esparrach
mgfern@clinic.cat

¹ Endoscopy Unit, Gastroenterology Department, Hospital Clínic, IDIBAPS, CIBEREHD, University of Barcelona, Barcelona, Spain

² Computer Science Department, Universitat Autònoma de Barcelona and Computer Vision Center, Barcelona, Spain

Materials and methods

Our system is based on a machine learning technique called Bag of Words [5] and uses a color descriptor to characterize each endoscopic image containing a polyp to build the vocabulary that will describe each image in a unique way.

Once each image in the dataset is described (training phase), the system can identify as similar those images sharing the same prevalence of vocabulary words (CBIR). To test this methodology, we used 243 white light images belonging to 99 different polyps acquired in routine colonoscopies at the Hospital Clínic using high-definition Olympus endoscopes. For each polyp at least two images were included (up to a maximum of eight), showing different views of each polyp. The method provided for each image the closest match within the dataset. When the chosen image contained the same polyp, it was considered a correct match.

The study was approved by the Research Ethics Committee of our hospital. All patients signed the consent before performing the colonoscopy.

Results

The mean size of polyps was 11.54 ± 10.79 mm. Overall, the system matched another image of the same polyp in 221/243 cases (91%). No differences were observed in the number of correct matches according to Paris classification (protruded: 90.7% vs. non-protruded: 91.3%) and size (≤ 10 mm: 91.6% vs. > 10 mm: 90%) (Fig. 1).

Failures were due to the presence of vessels crossing the surface of small polyps ($n=11$, 4.5%), specular highlights ($n=5$, 2%) and folds ($n=2$, 1%). We also observed an incorrect functioning of the system when the size of the polyp greatly varied across images of the same polyp ($n=4$, 1.6%):

in this case, image description contained more mucosa information or an extensive luminal area which could lead to mismatches.

Discussion

It is evident that there is a growing interest in the investigation of computational support systems for decision making through endoscopic images. For the first time, we have used a CBIR technique to assist in the unambiguous recognition of polyps in white light colonoscopy images. Our proposal can accurately identify as the pair image of another image of the same polyp, despite differences in point of view, size and appearance. This kind of systems have already been used in other medical imaging domain as tools to assist in the retrieval of similar images to one problematic target image by means of other descriptors as brightness or texture.

We foresee different applications of our system. First, in case of referred endoscopies for complex polypectomy, it might ensure the identification of the target polyp. This would be especially useful in colonic polyposis and when the target polyp location is poorly described, or it is not marked with a tattoo. Second, it could assist clinicians to provide a prediction of polyp histology by associating the histological category of the most similar image in the database. Future iterations of the method should include its adaptation to in vivo explorations. This could be especially useful for

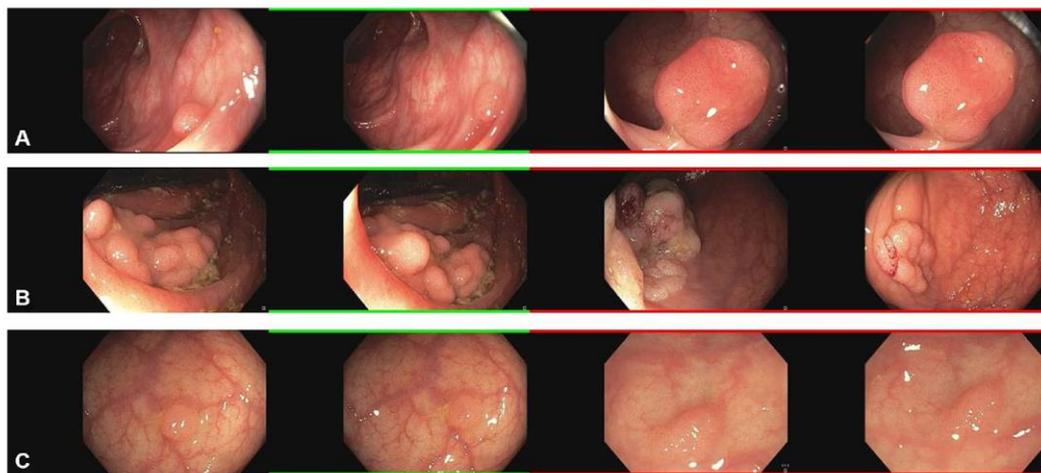


Fig. 1 Examples of the performance of the method in polyps with only two images in the database. For each row, the first image corresponds to the query image and the second one to the match provided by the system (outlined with a green rectangle because the match is

correct). Additional images in each row correspond to the second and third closest images that belong to different polyps (outlined in red) with similar appearance to the query one (Color figure online)

searching in the retrieval phase a polyp visualized during the intubation.

The main limitation of this system is the necessity of having a large dataset of images because of the machine learning methodology involved. Those images should be acquired in similar conditions, especially regarding endoscopic devices (which might affect image resolution) and endoluminal content. Some elements in the endoluminal content associated with failures (p.e. specular highlights) are potentially correctable, leading to a better performance of the method.

To conclude, we have shown that an AI system is able to extract meaningful color information from colonoscopy images with polyps, leading to the unambiguous recognition of each colorectal polyp ('polyp fingerprint'). This system could be a valuable tool for endoscopists in their daily decision making practice.

Funding This manuscript has not received any funding.

Compliance with ethical standards

Disclosures Ana García-Rodríguez, Jorge Bernal, F. Javier Sánchez, Henry Córdova, Rodrigo Garcés Durán, Cristina Rodríguez de Miguel and Gloria Fernández-Esparrache have no conflicts of interest or financial ties to disclose.

References

1. Von Renteln D, Bouin M, Barkun AN (2017) Current standards and new developments of colorectal polyp management and resection techniques. *Expert Rev Gastroenterol Hepatol* 11(9):835–842
2. Abu Dayyeh BK, Thosani N, Konda V, Wallace MB, Rex DK, Chauhan SS, Hwang JH, Komanduri S, Manfredi M, Maple JT, Murad FM, Siddiqui UD, Banerjee S (2015) ASGE technology committee systematic review and meta-analysis assessing the ASGE PIVI thresholds for adopting real-time endoscopic assessment of the histology of diminutive colorectal polyps. *Gastrointest Endosc* 81(3):502–516
3. Qin P, Li Q, Zeng J, Liu H, Cui Y (2019) Fully convolutional-based dense network for lung nodule image retrieval algorithm. *Int J Perform Eng* 15(1):326–336
4. Swati ZNK, Zhao Q, Kabir M, Ali F, Ali Z, Ahmed S, Lu J (2019) Content-based brain tumor retrieval for MR images using transfer learning. *IEEE Access* 7:17809–17822
5. Zhang Y, Jin R, Zhou Z-H (2010) Understanding bag-of-words model: a statistical framework. *Int J Mach Learn Cybern* 1(1–4):43–52

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

DISCUSIÓN

El trabajo de la presente tesis ha permitido:

Primero, diseñar una novedosa herramienta (GTCreator) para crear y facilitar anotaciones de bases de datos de imágenes (pensada específicamente para las imágenes de pólipos colorrectales) y compararla con otras herramientas de anotación de imágenes y segundo, evaluar dos sistemas inteligentes de soporte a la colonoscopia para mejorar el diagnóstico de pólipos colorrectales. Por un lado, se ha desarrollado un nuevo sistema CADx bautizado con el nombre de ATENEA, que ha sido probado in vivo y comparado su rendimiento con el de los endoscopistas y por el otro, se ha utilizado por primera vez un sistema CBIR para ayudar a la identificación inequívoca de pólipos (como si de una huella dactilar se tratase, de ahí el nombre de Polyp fingerprint), ambos con imágenes de luz blanca de alta definición. ATENEA ha mostrado un buen rendimiento para la predicción histológica en tiempo real e in vivo, similar al de los endoscopistas, lo que permitiría, por un lado, su integración en la práctica clínica diaria y por el otro, su difusión. Polyp fingerprint ha sido capaz de reconocer con precisión el mismo pólipo en 2 (o más) escenas endoluminales diferentes (tiempo, espacio y perspectiva).

El desarrollo de sistemas inteligentes depende de un gran volumen de imágenes como fuente de datos para poder obtenerse la información que se desea. Para poder validar el rendimiento de cualquier sistema de IA se debe haber definido previamente el GT. GTCreator es una herramienta de anotación que permite su creación (de ahí su nombre). GTCreator admite tres tipos de metadatos: 1. **Imágenes/máscaras**, para representar los objetos en la imagen (este tipo acepta anotación mediante dibujo a mano alzada o mediante el uso de formas), 2. **Texto**, para proporcionar información formateada asociada con la imagen (este tipo incluye a su vez variables booleanas (binarias), cadena o lista de valores) y 3. **Etiquetas semánticas**, para clasificar cada una de las regiones predefinidas en la imagen.

La principal innovación de GTCreator frente a otras herramientas similares ya existentes es su flexibilidad, ya que permite determinar libremente el número y tipo de metadatos a asociar a cada imagen durante la tarea de anotación. Esto hace que GTCreator sea apta

para su uso en cualquier tarea de anotación de imágenes. GTCreator también se puede utilizar en otras etapas del desarrollo y validación de un sistema de procesamiento de imágenes. Más específicamente aplicado al área de la colonoscopia y la IA, éstas son sus principales prestaciones:

- Crear anotaciones y ser gestionadas por varios usuarios. GTCreator permite retomar una anotación ya iniciada, así como incorporar nuevos metadatos sobre la marcha sin tener que iniciar el proceso de anotación desde cero. Para el caso de grandes conjuntos de datos en los que la tarea de anotación se dividiera en diferentes anotadores incorpora además la capacidad de combinación de anotaciones.
- Revisar anotaciones. GTCreator incluye tipos de metadatos que permiten la revisión de anotaciones por parte de otros usuarios. P. ej., un médico novato marca una imagen para que un experto la inspeccione más tarde; en este caso, el experto puede navegar fácilmente a través de las imágenes marcadas para su inspección utilizando las capacidades de filtrado de dicha herramienta.
- El archivo CSV producido como resultado de la anotación se puede utilizar para realizar configuraciones experimentales para la validación de un sistema inteligente. P. ej., metadatos específicos basados en texto pueden utilizarse para dividir un conjunto de datos en entrenamiento, validación y test.
- Ayudar en la validación de sistemas de IA mediante el tipo de metadatos “etiqueta”. En este caso, el usuario puede asignar cualquiera de las etiquetas predefinidas (establecidas en el archivo de definición) para cada una de las regiones de interés en cada imagen (p. ej., categoría histológica: “adenoma” vs “no adenoma”).

El estudio incluyó la comparación cualitativa y cuantitativa con otras herramientas similares (RatSnake(108), LabelMe(109), VGG Image Annotator (VIA)(110) [8], Video Image Annotation Tool (VIAT) [[https:// sourceforge.net/projects/via-tool/](https://sourceforge.net/projects/via-tool/)] and ImageJ(111)). La comparación cuantitativa se basó en la medición del tiempo de anotación y la exactitud alcanzada en las anotaciones. El experimento consistió en anotar 6 conjuntos diferentes de imágenes de colonoscopia con pólipos (13 imágenes divididas en dos subconjuntos: 10 de calidad estándar (extraídas de un vídeo) y 3 de alta definición). Participaron 6 personas del equipo de investigación Clínic-CVC (médicos, enfermeras y técnicos). La tarea de anotación solicitada fue la creación de una máscara

que cubriera el pólipo. Para medir la precisión, las diferentes anotaciones se compararon con el GT generado por un experto mediante un software de edición de imágenes utilizando métricas de segmentación como el coeficiente DICE y el índice Jaccard.

GTCreator fue la herramienta con un tiempo de anotación total medio más bajo, seguida de VIA y la que logró la puntuación más alta tanto en las métricas DICE y Jaccard, lo que añade robustez en la calidad de las anotaciones cuando son generadas por dicha herramienta.

En cuanto a las aplicaciones que se han desarrollado para mejorar el diagnóstico asistido de pólipos colorrectales mediante IA, ATENEA, surge como un nuevo CADx para superar las limitaciones del diagnóstico óptico en la práctica clínica y facilitar la implementación de las estrategias de *'resect and discard'* y *'leave in situ'*(58).

En los últimos años se han desarrollado y validado con éxito distintos modelos CADx. Sin embargo, aún no se ha podido determinar su aplicabilidad en el flujo de trabajo diario debido a limitaciones metodológicas y escasez de modelos probados en vida real. A esto se añade que la evidencia que apoya el uso de dicha tecnología se ha difundido en ausencia de unos estándares de calidad y requisitos específicos para cada IA(112).

Prácticamente todos los sistemas CADx publicados han sido probados ex vivo, con imágenes procedentes de un único centro y uso de tecnología óptica compleja, no disponible en todas las unidades de endoscopia. Recientemente se ha publicado el primer estudio multicéntrico e internacional (incluye a 3 hospitales, 2 en Occidente y 1 en Japón) de un sistema CADx in vivo(87), pero que utiliza imágenes de pólipos obtenidas con endocitoscopia y magnificación. A destacar como primer punto fuerte de ATENEA su capacidad para realizar una predicción precisa utilizando sólo imágenes de luz blanca de alta definición, que pueden obtenerse con cualquier tipo de endoscopio. Creemos que, si bien el operador humano depende de la aplicación de CE (con colorantes o virtual) para mejorar la visualización del patrón mucoso y vascular y, por ende, realizar una adecuada predicción histológica, los sistemas CADx deberían poder predecir la histología independientemente de la modalidad de visualización óptica

utilizada, marca y tipo de endoscopio. Este aspecto es crucial para facilitar la implementación de cualquier CADx de forma universal.

El modelo CADx previo al desarrollo de ATENEA que nuestro grupo había diseñado(93), había sido validado off-line en 225 imágenes de alta calidad consiguiendo una exactitud diagnóstica de más del 90%. Desafortunadamente, el sistema no era completamente automático y tampoco funcionaba en tiempo real, por lo que no podía usarse in vivo. Con el advenimiento de las redes neuronales y la adquisición de un mayor número de imágenes surge ATENEA, un dispositivo totalmente automático y que trabaja en tiempo real, debido a su capacidad de calcular el resultado en menos de 40 ms. Es importante no mezclar los conceptos “tiempo real” e “in vivo”, ya que muchas veces se utiliza el primero en lugar del segundo sin mencionar el tiempo de procesamiento real y viceversa(83,113).

Los resultados obtenidos en el experimento con ATENEA han mostrado que tenía un mejor rendimiento global que los endoscopistas, particularmente para lesiones adenomatosas. Por el contrario, los médicos clasificaron mejor los no adenomas. Las posibles causas de estas diferencias se desarrollan a continuación.

Por un lado, los endoscopistas tenían más información que ATENEA, primero, porque analizaban una secuencia de imágenes en movimiento (y ATENEA solo una imagen aislada) y segundo, porque conocían la localización y el tamaño de la lesión (variables que ATENEA no tenía en cuenta para su predicción) y ello podría haberles inducido a una mayor probabilidad pretest para la categoría “no adenoma” cuando se enfrentaban a un pólipo diminuto en el rectosigma. La inclusión de metadatos como el tamaño y la localización de los pólipos junto con la información clínica del paciente podría valorar incorporarse en futuras estrategias de DL. Sin embargo, en un estudio piloto previo realizado por nuestro grupo(114), la inclusión de estas características ayudó muy ligeramente a mejorar la predicción del CADx, por lo que nuestra hipótesis es que la adición de dichas variables es insignificante en comparación con la infinitud de datos que el algoritmo de DL es capaz de extraer de forma automática (y que por otro lado desconocemos).

Por otro lado, el entrenamiento de ATENEA se vio afectado por el menor número de ejemplos de la clase “no adenoma” en el conjunto de datos. Esto es debido a que la prevalencia de lesiones no adenomatosas es generalmente menor y no se resecan sistemáticamente cuando se localizan en el rectosigma, lo que dificulta su recolección. Este problema, que ocurre especialmente cuando el conjunto de datos de entrenamiento es pequeño y poco heterogéneo, se denomina ‘*overfitting*’ o “sobreajuste”, es decir, el modelo se ajusta demasiado a los datos de entrenamiento (en este caso a la clase mayoritaria) y no es capaz de aprender más allá ni poder extrapolar su aprendizaje a la exposición de nuevos y diferentes datos. En nuestra base de datos, se intentó evitar el solapamiento de imágenes del mismo pólipo entre los datos de entrenamiento y validación y así minimizar el sobreajuste. La solución al desequilibrio de clases, que por otro lado representa la realidad del día a día en la sala de endoscopia es un auténtico desafío.

Al hilo de lo anterior, el menor rendimiento de ATENEA para lesiones no adenomatosas podría limitar la implementación de la política ‘leave in situ’. El escaso número de pólipos diminutos en rectosigma no permite sacar conclusiones al respecto, pero el VPN no llegó al umbral establecido por el PIVI. Por tanto y como ya se ha expuesto, es necesario ampliar el conjunto de datos en cantidad global y en porcentaje de ejemplos de la clase minoritaria.

Además de los adenomas, los futuros sistemas CADx también deberían aprender a clasificar adecuadamente las LSS. Con ATENEA, por su similitud endoscópica con los pólipos hiperplásicos, se incluyeron en la categoría de “no adenomas”. Sin embargo, en algunos estudios, las LSS se han incluido de acuerdo con el riesgo de malignización en la categoría “neoplásicos” junto con los adenomas. Si hubiese sido así, ATENEA habría clasificado correctamente cuatro de las cinco LSS.

Los sistemas CADx no pretenden sustituir a los endoscopistas sino ayudarles en sus tareas. Según lo establecido en las guías de práctica clínica de la ESGE, el escenario más probable es que los sistemas inteligentes se utilicen como un "segundo observador" para respaldar el diagnóstico final del endoscopista(60). En nuestro estudio, si los endoscopistas que realizaron sus predicciones con baja confianza se hubieran fiado de la salida del sistema, todas sus predicciones habrían sido correctas. Esto demuestra el

potencial de ATENEA para ayudar a los médicos que realizan su predicción con menor grado de confianza (y presumiblemente menor experiencia). En nuestro caso, se asumió que todos los endoscopistas tenían un nivel de competencia en diagnóstico óptico semejante (tenían una tasa de detección de adenomas similar, trabajaban en el mismo hospital y tenían un número también análogo de endoscopias al año y años trabajados) y no se dio formación extra para la realización de este experimento.

Al igual que los endoscopistas, los sistemas inteligentes también brindan un valor de confianza en sus predicciones utilizando un porcentaje en lugar de una evaluación binaria (confianza alta versus baja). Al variar el valor de corte de la confianza puede observarse cómo clasifica ATENEA en una curva ROC (Receiver Operating Characteristic). Los resultados muestran que, si el valor de confianza es bajo, la sensibilidad del sistema es menor, ya que proporciona más salidas, siendo algunas de ellas erróneas. Por el contrario, si se aumenta el valor de confianza permitido, se proporcionan menos resultados, pero más seguros, con el riesgo de perder alguna lesión por no tener una salida. Así, p. ej., si ATENEA para una "imagen x" da una predicción con una probabilidad del 60% de que sea un "adenoma" (y del 40% de "no adenoma"), la predicción final será de "adenoma". Pero, si se predetermina que ATENEA sólo brinde una salida cuando la probabilidad de diagnóstico sea superior al 70%, para este ejemplo concreto, no dará ninguna predicción. En nuestro experimento decidimos utilizar como punto de corte un nivel de confianza del 50% para tener el máximo de predicciones posibles. La reciente publicación de los criterios SODA (Simple Optical Diagnosis Accuracy)(65), más flexibles que los criterios PIVI, enfatiza la importancia de no dejar ninguna lesión diminuta con neoplasia avanzada in situ. De acuerdo con estos criterios, consideramos que la sensibilidad es el resultado más importante del sistema CADx. El valor de área bajo la curva obtenido con ATENEA es similar al 0,84 reportado en un metaanálisis reciente utilizando solo endoscopia con luz blanca(115).

A diferencia de otros estudios en los que no se han incluido los conocidos como "casos difíciles para una IA", otra de las fortalezas de este estudio es que, al realizarse en un entorno clínico real, los pólipos se incluyeron independientemente de la calidad de la imagen (pólipos a una gran distancia del endoscopio o sin centrar, cubiertos por heces

o moco, borrosos). Si se hubieran excluido estos pólipos para el análisis, el rendimiento de ATENEA habría mejorado pero la validez de los resultados se habría visto mermada.

La falta de bases de datos públicas anotadas (y la ausencia de estudios comparativos in vivo) imposibilita hacer una comparación justa entre ATENEA y otros sistemas CADx. En este sentido, la comparación de métricas realizada en los metaanálisis es difícil de entender ya que la calidad, el número y la variabilidad de clases de pólipos en las imágenes son diferentes en cada estudio y afectan a las etapas de entrenamiento y validación de los sistemas computacionales.

El otro estudio realizado en esta tesis confirma que la extracción de una serie de características de los pólipos podría servir también para identificarlos unívocamente y diferenciarlos de otros. El experimento realizado con el sistema CBIR del Polyp fingerprint permite discernir si dos imágenes pertenecen al mismo pólipo o a otro muy similar del que poder inferir su histología. Este tipo de sistemas son especialmente valiosos en la medicina porque proporcionan como salida la(s) imagen(es) más similar(es) de un gran conjunto de datos a una imagen de consulta determinada, lo que ayuda a un diagnóstico rápido y eficiente de los pacientes.

Las posibles aplicaciones clínicas del Polyp fingerprint en el terreno de la colonoscopia son múltiples. Primero, en endoscopias referidas para realización de una polipectomía compleja podría asegurar la identificación del pólipo objetivo. Sería especialmente útil en casos de poliposis colónica y cuando la ubicación del pólipo objetivo estuviese mal descrita o sin marcar con un tatuaje. En segundo lugar, podría ayudar a los endoscopistas a proporcionar una predicción de la histología del pólipo por analogía con otro pólipo ya clasificado (y heredar su categoría). También se ha hipotetizado recientemente que puede apoyar a los propios sistemas CADx, cuyo rendimiento depende del GT, el análisis histológico. Se sabe que la precisión del diagnóstico histológico y el acuerdo entre patólogos, especialmente en pólipos ≤ 3 mm(116) y para LS(99) son desgraciadamente bajos. Fernández-Esparrach et al proponen utilizar el llamado aprendizaje semisupervisado al desarrollar herramientas CADx para solucionar este problema(117). En la situación desalentadora donde las imágenes no se pudiesen etiquetar correctamente debido a diagnósticos histopatológicos inciertos, podrían clasificarse en grupos por la similitud de su apariencia.

Los estudios de esta tesis tienen las siguientes limitaciones:

1. GTCreator (y podría hacerse extensible también a ATENEA y Polyp fingerprint) no puede crear anotaciones de vídeo, lo que podría ser interesante para: 1. Aumentar el tamaño de la muestra y 2. Ampliar la robustez de los sistemas CAD, ya que los videos contienen una mayor variedad de imágenes y reducen el riesgo de sobreajuste.
2. Más concretamente de ATENEA podemos decir lo siguiente: Primero, el conjunto de datos de entrenamiento era pequeño y pobre en términos de variabilidad. Para mitigar esto se aplicaron las operaciones de 1) *transfer learning*, con ImageNet(97) y (2) *data augmentation*, que es básicamente la fabricación de casos sintéticos. La colaboración con otros centros y la disponibilidad pública de otros conjuntos de datos también podría ser de gran utilidad tanto para ampliar el conjunto de datos como para realizar validaciones multicéntricas, aumentando la robustez de los resultados. Segundo, se analizaron pólipos de diferente tamaño y el número de pólipos diminutos fue muy pequeño, por lo que no se pueden sacar conclusiones para la implementación de las estrategias PIVI. Tercero, ATENEA no consideró por separado una categoría histológica para las LSS debido a la baja cantidad de ejemplos en el conjunto de datos. Es bien sabido que las LSS son lesiones neoplásicas y no existe un método óptico ideal para su caracterización. A ello se unen las discordancias entre patólogos para distinguir las LSS de los PH. Para intentar solventar este último problema, cuando había discordancia entre el diagnóstico del endoscopista y el patólogo en las imágenes utilizadas para el entrenamiento y validación, se solicitaba una revisión por un segundo patólogo.
3. Del Polyp fingerprint, pero también de ATENEA y todos estos dispositivos inteligentes, es la necesidad de tener un gran conjunto de datos de imágenes y que las imágenes deben ser adquiridas en condiciones similares, especialmente en lo que respecta a los equipos de endoscopia (que pueden afectar la resolución de la imagen) y contenido endoluminal. Algunos elementos en la escena endoluminal que han dado lugar a errores (p. ej., reflejos especulares) son potencialmente corregibles con métodos ya descritos por nuestro grupo, lo que conllevaría un mejor desempeño del sistema.

FUTURAS LÍNEAS DE TRABAJO

Para poder superar las limitaciones de los trabajos realizados y seguir avanzando en la implementación de la IA en colonoscopia, nos proponemos realizar los siguientes trabajos en el futuro:

- Mejora de sistemas utilizando técnicas más avanzadas de DL
- Aumentar la base de datos para:
 - Poder proporcionar pruebas más sólidas de rendimiento y seguridad (lo que facilitaría el proceso de aprobación de las organizaciones y sociedades reguladoras y de los pacientes)
 - Incluir y clasificar nuevas categorías histológicas
 - Diferenciar invasión submucosa superficial y profunda
- Integrar y validar detección y clasificación simultánea (CAdE y CAdx)
- Explorar el uso de diferentes tipos de CE virtual disponibles en la actualidad
- Estudios comparativos con otros sistemas ya desarrollados de IA

CONCLUSIONES

1. GTCreator es una herramienta que permite la creación y manipulación de anotaciones de forma rápida y precisa entre varios profesionales. Su versatilidad posibilita su uso durante las principales etapas de evaluación de sistemas inteligentes para cualquier dominio de imagen.
2. ATENEA, un sistema CADx de pólipos de colon basado únicamente en imágenes de luz blanca de alta definición, alcanza una predicción histológica precisa en vida real y un rendimiento similar a los endoscopistas a pesar de clasificar peor los no adenomas. ATENEA muestra su potencial como herramienta de soporte a los endoscopistas que realizan sus predicciones con baja confianza.
3. Polyp fingerprint, un sistema CBIR es capaz de extraer información significativa de los pólipos y hacer coincidir con precisión dos imágenes diferentes (tiempo, espacio y perspectiva) que contienen el mismo pólipo, lo que permite su identificación inequívoca y eventual clasificación por similitud.

BIBLIOGRAFÍA

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.
2. Cubiella J, Marzo-Castillejo M, Mascort-Roca JJ, Amador-Romero FJ, Bellas-Beceiro B, Clofent-Vilaplana J, et al. Clinical practice guideline. Diagnosis and prevention of colorectal cancer. 2018 Update. *Gastroenterol Hepatol*. 2018;41(9):585–96.
3. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut*. 2017;
4. Vuik FER, Nieuwenburg SAV, Bardou M, Lansdorp-Vogelaar I, Dinis-Ribeiro M, Bento MJ, et al. Increasing incidence of colorectal cancer in young adults in Europe over the last 25 years. *Gut*. 2019;
5. Rosato V, Bosetti C, Levi F, Polesel J, Zucchetto A, Negri E, et al. Risk factors for young-onset colorectal cancer. *Cancer Causes Control*. 2013;
6. Connell LC, Mota JM, Braghiroli MI, Hoff PM. The Rising Incidence of Younger Patients With Colorectal Cancer: Questions About Screening, Biology, and Treatment. *Current Treatment Options in Oncology*. 2017.
7. Gatta G, Mallone S, van der Zwan JM, Trama A, Siesling S, Capocaccia R, et al. Cancer survival in Europe 1999-2007 by country and age: Results of EURO CARE-5 - A population-based study. *Lancet Oncol*. 2014;15(1):23–34. Available from: [http://dx.doi.org/10.1016/S1470-2045\(13\)70546-1](http://dx.doi.org/10.1016/S1470-2045(13)70546-1)
8. Winawer SJ, Zauber AG, Ho MN, O'Brien MJ, Gottlieb LS, Sternberg SS, et al. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. *N Engl J Med*. 1993;
9. Netzer P, Forster C, Biral R, Ruchti C, Neuweiler J, Stauffer E, et al. Risk factor assessment of endoscopically removed malignant colorectal polyps. *Gut*. 1998;43(5):669–74.
10. Hackelsberger A, Fruhmorgen P, Weiler H, Heller T, Seeliger H, Junghanns K. Endoscopic polypectomy and management of colorectal adenomas with invasive carcinoma. *Endoscopy*. 1995;
11. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, et al. Genetic

- Alterations during Colorectal-Tumor Development. 2010 Jan 14 [cited 2021 Dec 15];319(9):525–32.
12. Bosman FY, Carneiro F, Hruban RB, Theise ND. WHO classification of tumors of the digestive system. 4th edn. WHO classification of tumors. 2010.
 13. Stolte M. Vienna classification of gastrointestinal epithelial neoplasia. *Japanese J Gastroenterol.* 1999;96:A387–9.
 14. James D. Brierley MKG. TNM Classification of Malignant tumours (eighth edition). European University Institute. 2017.
 15. Toes-Zoutendijk E, Kooyker AI, Elferink MA, Spaander MCW, Dekker E, Koning HJD, et al. Stage distribution of screen-detected colorectal cancers in the Netherlands. *Gut.* 2018.
 16. Dixon MF. Gastrointestinal epithelial neoplasia: Vienna revisited. *Gut.* 2002.
 17. Shaukat A, Kaltenbach T, Dominitz JA, Robertson DJ, Anderson JC, Cruise M, et al. Endoscopic Recognition and Management Strategies for Malignant Colorectal Polyps: Recommendations of the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology.* 2020;
 18. Snover DC. Update on the serrated pathway to colorectal carcinoma. *Hum Pathol.* 2011;
 19. Leggett B, Whitehall V. Role of the Serrated Pathway in Colorectal Cancer Pathogenesis. *Gastroenterology.* 2010;
 20. Abdeljawad K, Vemulapalli KC, Kahi CJ, Cummings OW, Snover DC, Rex DK. Sessile serrated polyp prevalence determined by a colonoscopist with a high lesion detection rate and an experienced pathologist. *Gastrointest Endosc.* 2015;
 21. Nagtegaal ID, Odze RD, Klimstra D, Paradis V, Rugge M, Schirmacher P, et al. The 2019 WHO classification of tumours of the digestive system. *Histopathology.* 2020;76(2):182–8.
 22. None JGCA. Japanese Classification of Gastric Carcinoma - 2nd English Edition -. Gastric Cancer. 1998 [cited 2021 Dec 15];1(1):10–24.
 23. Axon A, Diebold MD, Fujino M, Fujita R, Genta RM, Gonvers JJ, et al. The Paris Endoscopic Classification of Superficial neoplastic lesions. *Gastrointest Endosc.* 2003;58(6):3–43.

24. Axon A, Diebold MD, Fujino M, Fujita R, Genta RM, Gonvers JJ, et al. Update on the Paris classification of superficial neoplastic lesions in the digestive tract. *Endoscopy*. 2005;
25. Ross AS, Waxman I. Flat and depressed neoplasms of the colon in western populations. *Am J Gastroenterol*. 2006;101(1):172–80.
26. Kudo S ei, Lambert R, Allen JI, Fujii H, Fujii T, Kashida H, et al. Nonpolypoid neoplastic lesions of the colorectal mucosa. *Gastrointest Endosc*. 2008;68(4 SUPPL.).
27. Van Doorn SC, Hazewinkel Y, East JE, Van Leerdam ME, Rastogi A, Pellisé M, et al. Polyp morphology: An interobserver evaluation for the Paris classification among international experts. *Am J Gastroenterol*. 2015;110(1):180–7.
28. Kudo S, Rubio CA, Teixeira CR, Kashida H, Kogure E. Pit pattern in colorectal neoplasia: Endoscopic magnifying view. *Endoscopy*. 2001;
29. Matsuda T, Fujii T, Saito Y, Nakajima T, Uraoka T, Kobayashi N, et al. Efficacy of the invasive/non-invasive pattern by magnifying chromoendoscopy to estimate the depth of invasion of early colorectal neoplasms. *Am J Gastroenterol*. 2008;
30. Tanaka S, Sano Y. Aim to unify the narrow band imaging (NBI) magnifying classification for colorectal tumors: Current status in Japan from a summary of the consensus symposium in the 79th annual meeting of the Japan gastroenterological endoscopy society. In: *Digestive Endoscopy*. 2011.
31. Hewett DG, Kaltenbach T, Sano Y, Tanaka S, Saunders BP, Ponchon T, et al. Validation of a simple classification system for endoscopic diagnosis of small colorectal polyps using narrow-band imaging. *Gastroenterology*. 2012;
32. Sano Y, Tanaka S, Kudo SE, Saito S, Matsuda T, Wada Y, et al. Narrow-band imaging (NBI) magnifying endoscopic classification of colorectal tumors proposed by the Japan NBI expert team. *Digestive Endoscopy*. 2016.
33. Kanao H, Tanaka S, Oka S, Hirata M, Yoshida S, Chayama K. Narrow-band imaging magnification predicts the histology and invasion depth of colorectal tumors. *Gastrointest Endosc [Internet]*. 2009;69(3 SUPPL.):631–6.
34. Sano Y, Ikematsu H, Fu KI, Emura F, Katagiri A, Horimatsu T, et al. Meshed capillary vessels by use of narrow-band imaging for differential diagnosis of small colorectal polyps. *Gastrointest Endosc [Internet]*. 2009;69(2):278–83.
35. Wada Y, Kudo S ei, Kashida H, Ikehara N, Inoue H, Yamamura F, et al. Diagnosis of

- colorectal lesions with the magnifying narrow-band imaging system. *Gastrointest Endosc* [Internet]. 2009;70(3):522–31.
36. Nikamf T, Saito S, Tajiri H, Ikegami M. The evaluation of histological atypia and depth of invasion of colorectal lesions using magnified endoscopy with narrow-band imaging. *Gastroenterol Endosc*. 2009;
 37. IJspeert JEG, Bastiaansen BAJ, Van Leerdam ME, Meijer GA, Van Eeden S, Sanduleanu S, et al. Development and validation of the WASP classification system for optical diagnosis of adenomas, hyperplastic polyps and sessile serrated adenomas/polyps. *Gut*. 2016;
 38. Bisschops R, Hassan C, Bhandari P, Coron E, Neumann H, Pech O, et al. BASIC (BLI Adenoma Serrated International Classification) classification for colorectal polyp characterization with blue light imaging. *Endoscopy*. 2018;
 39. Mou S, Soetikno R, Shimoda T, Rouse R, Kaltenbach T. Pathologic predictive factors for lymph node metastasis in submucosal invasive (T1) colorectal cancer: A systematic review and meta-analysis. *Surg Endosc*. 2013;27(8):2692–703.
 40. Kitajima K, Fujimori T, Fuji S, Takeda J, Ohkura Y, Kawamata H, et al. Correlations between lymph node metastasis and depth of submucosal invasion in submucosal invasive colorectal carcinoma: A Japanese collaborative study. *J Gastroenterol*. 2004;39(6):534–43.
 41. Tominaga K, Nakanishi Y, Nimura S, Yoshimura K, Sakai Y, Shimoda T. Predictive histopathologic factors for lymph node metastasis in patients with nonpedunculated submucosal invasive colorectal carcinoma. *Dis Colon Rectum*. 2005 Jan [cited 2021 Dec 15];48(1):92–100.
 42. Burgess NG, Hourigan LF, Zanati SA, Brown GJ, Singh R, Williams SJ, et al. Risk Stratification for Covert Invasive Cancer Among Patients Referred for Colonic Endoscopic Mucosal Resection: A Large Multicenter Cohort. *Gastroenterology* [Internet]. 2017 Sep 1 [cited 2022 Jun 8];153(3):732-742.e1.
 43. Bogie RMM, Veldman MHJ, Snijders LARS, Winkens B, Kaltenbach T, Masclee AAM, et al. Endoscopic subtypes of colorectal laterally spreading tumors (LSTs) and the risk of submucosal invasion: A meta-analysis. *Endoscopy* [Internet]. 2018 Mar 1 [cited 2022 Jun 8];50(3):263–82.
 44. Backes Y, Schwartz MP, Ter Borg F, Wolfhagen FHJ, Groen JN, De Vos Tot Nederveen

- Cappel WH, et al. Multicentre prospective evaluation of real-time optical diagnosis of T1 colorectal cancer in large non-pedunculated colorectal polyps using narrow band imaging (the OPTICAL study). *Gut*. 2019 Feb 1 [cited 2022 Jun 8];68(2):271–9.
45. D’Amico F, Amato A, Iannone A, Trovato C, Romana C, Angeletti S, et al. Risk of Covert Submucosal Cancer in Patients With Granular Mixed Laterally Spreading Tumors. *Clin Gastroenterol Hepatol*. 2021 Jul 1 [cited 2022 Jun 8];19(7):1395–401.
 46. Puig I, López-Cerón M, Arnau A, Rosiñol Ò, Cuatrecasas M, Herreros-de-Tejada A, et al. Accuracy of the Narrow-Band Imaging International Colorectal Endoscopic Classification System in Identification of Deep Invasion in Colorectal Polyps. *Gastroenterology*. 2019 Jan 1 [cited 2022 Jun 8];156(1):75–87.
 47. Moss A, Bourke MJ, Williams SJ, Hourigan LF, Brown G, Tam W, et al. Endoscopic mucosal resection outcomes and prediction of submucosal cancer from advanced colonic mucosal neoplasia. *Gastroenterology*. 2011;
 48. Vosko S, Shahidi N, Sidhu M, van Hattem WA, Bar-Yishay I, Schoeman S, et al. Optical Evaluation for Predicting Cancer in Large Nonpedunculated Colorectal Polyps Is Accurate for Flat Lesions. *Clin Gastroenterol Hepatol*. 2021 Nov 1 [cited 2022 Jun 8];19(11):2425-2434.e4.
 49. Rutter MD, Chattree A, Barbour JA, Thomas-Gibson S, Bhandari P, Saunders BP, et al. British society of gastroenterology/association of coloproctologists of Great Britain and Ireland guidelines for the management of large non-pedunculated colorectal polyps. *Gut*. 2015;
 50. Ferlitsch M, Moss A, Hassan C, Bhandari P, Dumonceau JM, Paspatis G, et al. Colorectal polypectomy and endoscopic mucosal resection (EMR): European Society of Gastrointestinal Endoscopy (ESGE) Clinical Guideline. *Endoscopy*. 2017.
 51. Ueno H, Mochizuki H, Hashiguchi Y, Shimazaki H, Aida S, Hase K, et al. Risk factors for an adverse outcome in early invasive colorectal carcinoma. *Gastroenterology*. 2004;127(2):385–94.
 52. Kobayashi N, Saito Y, Sano Y, Uragami N, Michita T, Nasu J, et al. Determining the treatment strategy for colorectal neoplastic lesions: Endoscopic assessment or the non-lifting sign for diagnosing invasion depth? *Endoscopy*. 2007;
 53. da Costa-Seixas JP, López-Cerón M, Arnau A, Rosiñol Ò, Cuatrecasas M, Herreros-De-Tejada A, et al. Compound Endoscopic Morphological Features for Identifying Non-

- Pedunculated Lesions ≥ 20 mm with Intramucosal Neoplasia. *Cancers (Basel)*. 2021 Nov 1 [cited 2022 Jun 8];13(21).
54. Matsuda T, Fukuzawa M, Uraoka T, Nishi M, Yamaguchi Y, Kobayashi N, et al. Risk of lymph node metastasis in patients with pedunculated type early invasive colorectal cancer: A retrospective multicenter study. *Cancer Sci*. 2011;
 55. Albéniz E, Pellisé M, Gimeno García AZ, Lucendo AJ, Alonso Aguirre PA, Herreros de Tejada A, et al. Guía clínica para la resección mucosa endoscópica de lesiones colorrectales no pediculadas. *Gastroenterol Hepatol* 2018 Mar [cited 2018 Sep 25];41(3):175–90.
 56. Lieberman D, Moravec M, Holub J, Michaels L, Eisen G. Polyp Size and Advanced Histology in Patients Undergoing Colonoscopy Screening: Implications for CT Colonography. *Gastroenterology*. 2008;
 57. Vleugels JLA, Hassan C, Senore C, Cassoni P, Baron JA, Rex DK, et al. Diminutive Polyps With Advanced Histologic Features Do Not Increase Risk for Metachronous Advanced Colon Neoplasia. *Gastroenterology*. 2019;
 58. Rex DK, Kahi C, O'Brien M, Levin TR, Pohl H, Rastogi A, et al. The American Society for Gastrointestinal Endoscopy PIVI (Preservation and Incorporation of Valuable Endoscopic Innovations) on real-time endoscopic assessment of the histology of diminutive colorectal polyps. *Gastrointest Endosc*. 2011;
 59. Hassan C, Pickhardt PJ, Rex DK. A resect and discard strategy would improve cost-effectiveness of colorectal cancer screening. *Clin Gastroenterol Hepatol*. 2010;
 60. Bisschops R, East JE, Hassan C, Hazewinkel Y, Kamiński MF, Neumann H, et al. Correction: Advanced imaging for detection and differentiation of colorectal neoplasia: European Society of Gastrointestinal Endoscopy (ESGE) Guideline - Update 2019 (Endoscopy DOI: 10.1055/a-1031-7657). *Endoscopy*. 2019;51(12):C6.
 61. NICE. Virtual chromoendoscopy to assess colorectal polyps during colonoscopy. 2017;(May 2017).
 62. Rutter MD, East J, Rees CJ, Cripps N, Docherty J, Dolwani S, et al. British Society of Gastroenterology/Association of Coloproctology of Great Britain and Ireland/Public Health England post-polypectomy and post-colorectal cancer resection surveillance guidelines. *Gut*. 2020;
 63. Vu HT, Sayuk GS, Hollander TG, Clebanoff J, Edmundowicz SA, Gyawali CP, et al. Resect

- and Discard Approach to Colon Polyps: Real-World Applicability Among Academic and Community Gastroenterologists. *Dig Dis Sci*. 2015;
64. Ladabaum U, Fioritto A, Mitani A, Desai M, Kim JP, Rex DK, et al. Real-time optical biopsy of colon polyps with narrow band imaging in community practice does not yet meet key thresholds for clinical decisions. *Gastroenterology*. 2013;
 65. Houwen BBSL, Hassan C, Coupé VMH, Greuter MJE, Hazewinkel Y, Vleugels JLA, et al. Definition of competence standards for optical diagnosis of diminutive colorectal polyps: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. *Endoscopy*. 2022;
 66. McClendon L, Meghanathan N. Using Machine Learning Algorithms to Analyze Crime Data. *Mach Learn Appl An Int J*. 2015;
 67. Shalev-Shwartz S, Ben-David S. Understanding machine learning: From theory to algorithms. *Understanding Machine Learning: From Theory to Algorithms*. 2013.
 68. García Rodríguez D. INDUSTRIAL IoT. MACHINE LEARNING EN LA INDUSTRIA 4.0. Univ Politec Cart Barcelona. 2020;
 69. Robledano Arillo J, Moreiro González JA. La Recuperación Documental De La Imagen Fotográfica: Perspectiva Tecnológica Y Documental. *Prim Jornadas Imagen, Cult y Tecnol* (1, 2002, Getafe, Madrid). 2002;
 70. Ketkar N. Deep Learning with Python - A Hands-on Introduction. *Deep Learning with Python*. 2017. 1–169 p.
 71. Tischendorf JJW, Gross S, Winograd R, Hecker H, Auer R, Behrens A, et al. Computer-aided classification of colorectal polyps based on vascular patterns: A pilot study. *Endoscopy*. 2010;
 72. Gross S, Trautwein C, Behrens A, Winograd R, Palm S, Lutz HH, et al. Computer-based classification of small colorectal polyps by using narrow-band imaging with optical magnification. *Gastrointest Endosc*. 2011;
 73. Takemura Y, Yoshida S, Tanaka S, Kawase R, Onji K, Oka S, et al. Computer-aided system for predicting the histology of colorectal tumors by using narrow-band imaging magnifying colonoscopy (with video). *Gastrointest Endosc*. 2012;
 74. Tamai N, Saito Y, Sakamoto T, Nakajima T, Matsuda T, Sumiyama K, et al. Effectiveness of computer-aided diagnosis of colorectal lesions using novel software for magnifying narrow-band imaging: a pilot study. *Endosc Int Open*. 2017;

75. Kominami Y, Yoshida S, Tanaka S, Sanomura Y, Hirakawa T, Raytchev B, et al. Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy. *Gastrointest Endosc.* 2016;
76. Chen PJ, Lin MC, Lai MJ, Lin JC, Lu HHS, Tseng VS. Accurate Classification of Diminutive Colorectal Polyps Using Computer-Aided Analysis. *Gastroenterology.* 2018;
77. Byrne MF, Chapados N, Soudan F, Oertel C, Pérez ML, Kelly R, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut.* 2019;
78. Van Der Zander QEW, Schreuder RM, Fonollà R, Scheeve T, Van Der Sommen F, Winkens B, et al. Optical diagnosis of colorectal polyp images using a newly developed computer-aided diagnosis system (CADx) compared with intuitive optical diagnosis. *Endoscopy.* 2021;
79. Mori Y, Kudo SE, Wakamura K, Misawa M, Ogawa Y, Kutsukawa M, et al. Novel computer-aided diagnostic system for colorectal lesions by using endocytoscopy (with videos). *Gastrointest Endosc.* 2015;
80. Mori Y, Kudo S ei, Chiu PWY, Singh R, Misawa M, Wakamura K, et al. Impact of an automated system for endocytoscopic diagnosis of small colorectal lesions: an international web-based study. *Endoscopy.* 2016;
81. Takeda K, Kudo SE, Mori Y, Misawa M, Kudo T, Wakamura K, et al. Accuracy of diagnosing invasive colorectal cancer using computer-aided endocytoscopy. *Endoscopy.* 2017;
82. Mori Y, Kudo SE, Mori K. Potential of artificial intelligence-assisted colonoscopy using an endocytoscope (with video). *Digestive Endoscopy.* 2018.
83. Mori Y, Kudo SE, Misawa M, Saito Y, Ikematsu H, Hotta K, et al. Real-time use of artificial intelligence in identification of diminutive polyps during colonoscopy a prospective study. *Ann Intern Med.* 2018;
84. Misawa M, Kudo SE, Mori Y, Nakamura H, Kataoka S, Maeda Y, et al. Characterization of Colorectal Lesions Using a Computer-Aided Diagnostic System for Narrow-Band Imaging Endocytoscopy. *Gastroenterology.* 2016;
85. Misawa M, Kudo S ei, Mori Y, Takeda K, Maeda Y, Kataoka S, et al. Accuracy of computer-aided diagnosis based on narrow-band imaging endocytoscopy for

- diagnosing colorectal lesions: comparison with experts. *Int J Comput Assist Radiol Surg.* 2017;
86. Kudo S ei, Misawa M, Mori Y, Hotta K, Ohtsuka K, Ikematsu H, et al. Artificial Intelligence-assisted System Improves Endoscopic Identification of Colorectal Neoplasms. *Clin Gastroenterol Hepatol.* 2020;
 87. Barua I, Wieszczy P, Kudo S, Misawa M, Holme Ø, Gulati S, et al. Real-Time Artificial Intelligence–Based Optical Diagnosis of Neoplastic Polyps during Colonoscopy. *NEJM Evid [Internet].* 2022 Apr 13 [cited 2022 May 16]; Available from: <https://evidence.nejm.org/doi/full/10.1056/EVIDoa2200003>
 88. André B, Vercauteren T, Buchner AM, Krishna M, Ayache N, Wallac MB. Software for automated classification of probe-based confocal laser endomicroscopy videos of colorectal polyps. *World Journal of Gastroenterology.* 2012.
 89. Ștefănescu D, Streba C, Cârțână ET, Săftoiu A, Gruionu G, Gruionu LG. Computer aided diagnosis for confocal laser endomicroscopy in advanced colorectal adenocarcinoma. *PLoS One.* 2016;
 90. Rath T, Tontini GE, Vieth M, Nägel A, Neurath MF, Neumann H. In vivo real-time assessment of colorectal polyp histology using an optical biopsy forceps system based on laser-induced fluorescence spectroscopy. *Endoscopy.* 2016;
 91. Aihara H, Saito S, Inomata H, Ide D, Tamai N, Ohya TR, et al. Computer-aided diagnosis of neoplastic colorectal lesions using “real-time” numerical color analysis during autofluorescence endoscopy. *Eur J Gastroenterol Hepatol.* 2013;
 92. Komeda Y, Handa H, Watanabe T, Nomura T, Kitahashi M, Sakurai T, et al. Computer-Aided Diagnosis Based on Convolutional Neural Network System for Colorectal Polyp Classification: Preliminary Experience. *Oncol.* 2017;
 93. Sánchez-Montes C, Sánchez F, Bernal J, Córdova H, López-Cerón M, Cuatrecasas M, et al. Computer-aided prediction of polyp histology on white light colonoscopy using surface pattern analysis. *Endoscopy [Internet].* 2018;2–6.
 94. Yang YJ, Cho BJ, Lee MJ, Kim JH, Lim H, Bang CS, et al. Automated classification of colorectal neoplasms in white-light colonoscopy images via deep learning. *J Clin Med.* 2020;
 95. Bang CS, Lee JJ, Baik GH. Computer-aided diagnosis of diminutive colorectal polyps in endoscopic images: Systematic review and meta-analysis of diagnostic test accuracy.

- Journal of Medical Internet Research. 2021.
96. Shahidi N, Rex DK, Kaltenbach T, Rastogi A, Ghalehjegh SH, Byrne MF. Use of Endoscopic Impression, Artificial Intelligence, and Pathologist Interpretation to Resolve Discrepancies Between Endoscopy and Pathology Analyses of Diminutive Colorectal Polyps. *Gastroenterology*. 2020;
 97. Li L-J, Li K, Li FF, Deng J, Dong W, Socher R, et al. ImageNet: a Large-Scale Hierarchical Image Database Shrimp Project View project hybrid intrusion detection systems View project ImageNet: A Large-Scale Hierarchical Image Database. 2009 IEEE Conf Comput Vis Pattern Recognit. 2009;
 98. Lin T-Y, Maire M, Belongie S, Bourdev L, Girshick R, Hays J, et al. Microsoft COCO: Common Objects in Context. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*. 2015;
 99. Wong NACS, Hunt LP, Novelli MR, Shepherd NA, Warren BF. Observer agreement in the diagnosis of serrated polyps of the large bowel. *Histopathology*. 2009;
 100. Mori Y, Kudo SE. Detecting colorectal polyps via machine learning. *Nature biomedical engineering*. 2018.
 101. Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology*. 2018.
 102. Carter D. Regulation and ethics in artificial intelligence and machine learning technologies: Where are we now? Who is responsible? Can the information professional play a role? *Bus Inf Rev*. 2020;
 103. Tiple V. Recommendations on the European Commission's WHITE PAPER on Artificial Intelligence - A European approach to excellence and trust, COM(2020) 65 final (the 'AI White Paper'). *SSRN Electron J*. 2020;
 104. Hann A, Troya J, Fitting D. Current status and limitations of artificial intelligence in colonoscopy. *United European Gastroenterology Journal*. 2021.
 105. Repici A, Badalamenti M, Maselli R, Correale L, Radaelli F, Rondonotti E, et al. Efficacy of Real-Time Computer-Aided Detection of Colorectal Neoplasia in a Randomized Trial. *Gastroenterology*. 2020;
 106. Liu P, Wang P, Glissen Brown JR, Berzin TM, Zhou G, Liu W, et al. The single-monitor trial: an embedded CAde system increased adenoma detection during colonoscopy: a prospective randomized study. *Therap Adv Gastroenterol*. 2020;

107. Schmitz R, Werner R, Repici A, Bisschops R, Meining A, Zornow M, et al. Artificial intelligence in GI endoscopy: Stumbling blocks, gold standards and the role of endoscopy societies. *Gut*. 2022.
108. Iakovidis DK, Goudas T, Smailis C, Maglogiannis I. Ratsnake: A versatile image annotation tool with application to computer-aided diagnosis. *Sci World J*. 2014;
109. Russell BC, Torralba A, Murphy KP, Freeman WT. LabelMe: A database and web-based tool for image annotation. *Int J Comput Vis*. 2008;
110. Dutta A, Zisserman A. The VIA annotation software for images, audio and video. In: *MM 2019 - Proceedings of the 27th ACM International Conference on Multimedia*. 2019.
111. Abràmoff MD, Magalhães PJ, Ram SJ. Image processing with imageJ. *Biophotonics International*. 2004.
112. Van Der Sommen F, De Groof J, Struyvenberg M, Van Der Putten J, Boers T, Fockens K, et al. Machine learning in GI endoscopy: Practical guidance in how to interpret a novel field. *Gut*. 2020;
113. Rodriguez-Diaz E, Baffy G, Lo WK, Mashimo H, Vidyarthi G, Mohapatra SS, et al. Real-time artificial intelligence–based histologic classification of colorectal polyps with augmented visualization. In: *Gastrointestinal Endoscopy*. 2021.
114. García-Rodríguez A, Tudela Y, Sánchez F, Córdova H, Garcés-Durán R, Cuatrecasas M, et al. HISINVIA: A HYBRID SOLUTION FOR COLONIC POLYP HISTOLOGY PREDICTION IN WHITE LIGHT COLONOSCOPY IMAGES COMBINING ARTIFICIAL INTELLIGENCE AND CLINICAL INFORMATION. In: *ESGE Days*. 2020.
115. Lui TKL, Guo CG, Leung WK. Accuracy of artificial intelligence on histology prediction and detection of colorectal polyps: a systematic review and meta-analysis. *Gastrointestinal Endoscopy*. 2020.
116. Ponugoti P, Rastogi A, Kaltenbach T, MacPhail ME, Sullivan AW, Thygesen JC, et al. Disagreement between high confidence endoscopic adenoma prediction and histopathological diagnosis in colonic lesions ≤ 3 mm in size. *Endoscopy*. 2019;
117. Mori Y. Gastrointestinal Endoscopy Artificial Intelligence for Disease Diagnosis - the Gold Standard Challenge.

DECLARACIÓN DE CÓDIGO ÉTICO Y BUENAS PRÁCTICAS

Los estudios se realizan siguiendo los principios expresados en la declaración de Helsinki y han sido aprobado por el Comité Ético del Hospital Clínic. Todos los datos referentes a pacientes permanecieron en el anonimato y los pacientes firmaron un consentimiento informado.

DECLARACIÓN DE LA DOCTORANDA

LA DOCTORANDA, ANA GARCÍA-RODRÍGUEZ,

CERTIFICA:

Que la memoria que lleva por título *"Inteligencia artificial como herramienta de soporte para la caracterización de pólipos colorrectales en imágenes de luz blanca de alta definición"*, para optar al grado de doctora en Medicina, no contiene plagio. Conozco y consiento que la tesis podrá ser sometida a procedimiento para comprobar su originalidad.

Y para que quede constancia a los efectos oportunos, firmo la presente en 13 de julio de 2022.

Ana García-Rodríguez



DECLARACIÓN DE LA DIRECTORA Y TUTORA DE TESIS

LA DRA. MARÍA GLORIA FERNANDEZ ESPARRACH, INVESTIGADORA SENIOR DEL INSTITUT D'INVESTIGACIONS BIOMÈDIQUES AUGUST PI I SUNYER (IDIBAPS) DE BARCELONA,

CERTIFICA:

Que la memoria que lleva por título *"Inteligencia artificial como herramienta de soporte para la caracterización de pólipos colorrectales en imágenes de luz blanca de alta definición"*, presentada por Ana García-Rodríguez para optar al grado de Doctora en Medicina, ha sido realizada bajo mi dirección y tutela, cumpliendo con los códigos éticos y de buenas prácticas, y sin conocimiento de que se haya producido ningún plagio.

Y para que quede constancia a los efectos oportunos, firmo la presente en 13 de julio de 2022.



Dra. María Gloria Fernández Esparrach

DECLARACIÓN DEL DIRECTOR DE TESIS

EL DR. JORGE BERNAL DEL NOZAL, PROFESSOR AGREGAT DEL DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN DE LA UNIVERSIDAD AUTÓNOMA DE BARCELONA E INVESTIGADOR ADSCRITO AL CENTRO DE VISIÓN POR COMPUTADOR,

CERTIFICA:

Que la memoria que lleva por título *"Inteligencia artificial como herramienta de soporte para la caracterización de pólipos colorrectales en imágenes de luz blanca de alta definición"*, presentada por Ana García-Rodríguez para optar al grado de Doctora en Medicina, ha sido realizada bajo mi dirección y tutela, cumpliendo con los códigos éticos y de buenas prácticas, y sin conocimiento de que se haya producido ningún plagio.

Y para que quede constancia a los efectos oportunos, firmo la presente en 13 de julio de 2022.



Dr. Jorge Bernal del Nozal

