



HAL
open science

Développement et applications des outils de l'intelligence artificielle à la vidéocapsule endoscopique de l'intestin grêle

Romain Leenhardt

► **To cite this version:**

Romain Leenhardt. Développement et applications des outils de l'intelligence artificielle à la vidéocapsule endoscopique de l'intestin grêle. Imagerie médicale. CY Cergy Paris Université, 2023. Français. NNT: 2023CYUN1194 . tel-04318194

HAL Id: tel-04318194

<https://theses.hal.science/tel-04318194>

Submitted on 1 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse de doctorat

pour l'obtention du titre de Docteur en
Sciences et Technologies de l'Information
et de la Communication

Délivré par

l'Université de Cergy-Pontoise

Ecole doctorale n°407 EM2PSI: Sciences et Technologies de l'Information et de la
Communication

Développement et applications des outils de l'intelligence artificielle à la vidéocapsule endoscopique de l'intestin grêle

Présentée et soutenue publiquement par

Romain Leenhardt

Le 18 octobre 2023 à 10h30

Directeurs de these : Monsieur Aymeric Histace, Monsieur Xavier Dray

Jury

Pr Véronica Belmega

Présidente

Pr Emmanuel Coron

Rapporteur

Pr Xavier Dray

Directeur

Pr Aymeric Histace

Directeur

Dr Sébastien Li Thiao Té

Rapporteur

Dr Lucille Quénéhervé

Examinatrice

Remerciements

A Xavier et Aymeric pour leur confiance, leur disponibilité et leur accompagnement précieux dans ce projet.

A tous les membres du jury, d'avoir accepté d'évaluer ce travail.

A Marc pour son aide et accompagnement dans le développement des algorithmes.

A toute l'équipe d'Endoscopie digestive de l'hôpital Saint Antoine pour leur soutien inconditionnel : Ulriikka, Marine, Claire, David, Antoine et Grégoire.

A toute l'équipe de Gastroentérologie de l'hôpital Saint Antoine pour leur accompagnement et soutien dans ce travail durant mon clinicat.

A toute l'équipe d'Augmented Endoscopy pour cette belle aventure.

A ma famille, mes amis et surtout à Claire, Simon et Martin pour cette belle vie partagée.

Table des matières

1. Introduction	7
2. Epidémiologie en gastroentérologie.....	8
3. La vidéocapsule endoscopique.....	10
3.1. Présentation du dispositif de vidéocapsule endoscopique	10
3.2. La vidéocapsule endoscopique en pratique clinique : quelles indications ?	13
3.3. Rentabilité diagnostique de la vidéocapsule endoscopique.....	13
3.4. Les enjeux de la préparation de l'intestin grêle.....	14
4. L'Intelligence artificielle.....	15
4.1. Généralités sur l'intelligence artificielle	15
4.2. Terminologie en vision par ordinateur	16
4.3. L'apprentissage machine (Machine Learning)	17
4.3.1. L'apprentissage supervisé	17
4.3.2. L'apprentissage non supervisé	17
4.3.3. L'apprentissage profond (Deep Learning).....	18
5. Applications de l'intelligence artificielle en médecine	20
6. Premières tentatives d'intelligence artificielle appliquées à la lecture des VCE.....	20
7. Applications de l'intelligence artificielle en vidéocapsule endoscopique.....	21
7.1. L'intelligence artificielle en vidéocapsule : quelles perceptions et attentes ?	21
7.2. L'intelligence artificielle en vidéocapsule : quels objectifs ?	34
7.3. Les spécificités de l'intelligence artificielle appliquée à la vidéocapsule	43
7.4. Bases de données d'apprentissage	43
7.5. Evaluation de la complétude de l'examen du grêle.....	50
7.6. Evaluation de la propreté de l'intestin grêle	52
7.7. Elaboration d'une vérité terrain en vidéocapsule	58
7.7.1. Terminologie et sémiologie en vidéocapsule	58
7.7.2. Imputabilité en vidéocapsule	77
7.8. Détection des lésions vasculaires	88
7.9. Détection multiclasse.....	95
7.10. La problématique de la détection du sang.....	95
8. Valorisation industrielle	96
9. Discussion.....	98
10. Conclusion.....	103
Références.....	105

Table des figures

Figure 1 : Illustration de la réalisation d'une coloscopie (Johns Hopkins, Cancer Center, 2017)	9
Figure 2 : Schéma anatomique du tube digestif	9
Figure 3 : Modèle de vidéocapsule endoscopique de l'IG de 3ème génération (SB3 Pillcam [®] , Medtronic).....	10
Figure 4 : Vue schématique d'une vidéocapsule endoscopique de l'IG de 3ème génération (SB3 Pillcam [®] , Medtronic).	11
Figure 5 : Image d'IG captée par vidéocapsule endoscopique (SB3 Pillcam [®] , Medtronic).	12
Figure 6 : Image du viewer (SB3 Pillcam [®] , Medtronic).....	12
Figure 7 : Indications de réalisation d'une VCE et rentabilité diagnostique associée (11).	13
Figure 8 : Illustration de l'analyse d'une image par un ordinateur	16
Figure 9 : Illustration de la terminologie en vision par ordinateur	16
Figure 10 : Illustration d'un apprentissage supervisé et non supervisé	18
Figure 11 : Illustration d'un réseau de neurones artificiel. (Création personnelle)	19
Figure 12 : Illustration du fonctionnement d'une convolution. (Création Dr Charles Houdeville).....	19
Figure 13 : Projet NOCE : Nouveaux Outils en Capsule Endoscopique.....	50
Figure 14 : Illustration du temps de progression de la capsule et résultats préliminaires	52
Figure 15 : Images de VCE-G montrant une angiodysplasie (photo de gauche) et ulcération superficielle (photo de droite).	58
Figure 16 : Illustration d'images de saignement, avec de gauche à droite : trace de sang, caillots de sang, sang noir digéré	96
Figure 17 : Solution Axaro.AI.....	97
Figure 18 : présentation de la plateforme Axaro.AI	98

Listes des abréviations :

CCR : Cancer Colorectal

CNP HGE : Conseil National Professionnel d'Hépatogastroentérologie

CNRS : Centre National de la Recherche Scientifique

ENSEA : École Nationale supérieure de l'Électronique et de ses Applications

ESGE : Société Européenne d'Endoscopie digestive

ETIS : Equipes Traitement de l'Information et Systèmes

HAS : Haute Autorité de Santé

IA : Intelligence Artificielle

IRM : Imagerie par résonance magnétique

MC : Maladie de Crohn

SSIG : Suspicion de saignement de l'intestin grêle

VCE : Vidéocapsule Endoscopique IG : Intestin grêle

Résumé Français

Depuis sa présentation en 2000, la vidéo capsule endoscopique (VCE) a révolutionné l'exploration non invasive de l'intestin grêle (IG). Elle permet d'acquérir plusieurs images d'IG par seconde pendant une durée d'au moins huit heures et d'obtenir une exploration de la totalité de la muqueuse de l'IG dans plus de 90% des cas. Les images émises par cette caméra miniaturisée sont stockées dans un boîtier enregistreur par l'intermédiaire de capteurs positionnés sur l'abdomen du patient. Avec 10000 images d'IG en moyenne, l'interprétation s'avère fastidieuse, chronophage et à risque de lésions manquées et d'incertitude diagnostique. Les outils de l'intelligence artificielle (IA) sont susceptibles d'atténuer ces limites. Un sondage auprès de 333 gastroentérologues, puis un consensus Delphi auprès de 21 experts, tous deux menés à l'échelle européenne, ont confirmé l'intérêt et la confiance des médecins dans ce développement, et identifié des questions de recherche prioritaires. La création d'une première banque de 4174 vidéos et 5184 images d'intérêt a été un temps fondamental. Parallèlement, nous avons élaboré par trois processus Delphi, dans le cadre de collaborations internationales avec des experts en VCE, des vérités terrains sur la nomenclature, la description et l'imputabilité (selon le contexte clinique) des principales lésions d'intérêt de l'IG (lésions vasculaires, lésions ulcéro-inflammatoires). Sur ces bases, nous avons développé et évalué plusieurs algorithmes issus des réseaux neuronaux dédiés à la lecture automatique des VCE. Nous avons enfin établi des preuves de concept sur les bonnes performances diagnostiques de réseaux neuronaux supervisés pour la détection d'angiodysplasies intestinales (Se = 100%, Sp = 96%) et pour l'évaluation de qualité de la préparation intestinale (Se = 90,3 %, Sp 83,3%). Les résultats de ces travaux font l'objet d'une valorisation industrielle à travers le dépôt de brevets et la création d'une start-up. L'implémentation à grande échelle de l'IA dans notre société représente d'ores et déjà un bouleversement dans notre quotidien. Un changement de paradigme s'impose de fait au monde médical qui se retrouve lui aussi bousculé par l'arrivée de ces nouveaux outils et algorithmes dont le fonctionnement complexe reste peu maîtrisé par les soignants. Cette révolution technologique nous expose à des problématiques nouvelles et nécessite d'être accompagnée par les différents acteurs de santé dont les soignants afin de participer à son implémentation en pratique clinique et d'identifier les avantages et freins potentiels de ses applications en santé. Ce travail se positionne à l'interface de ce monde médico-technique complexe associant la médecine et l'analyse d'image.

Résumé Anglais

Since its inception in 2000, capsule endoscopy (CE) has revolutionized non-invasive small bowel (SB) exploration. CE is able to acquire several SB images per second during at least eight hours and can provide a full SB exploration in more than 90% of cases. The images emitted by this miniaturized camera are captured in a recording box through sensors positioned on the patient's abdomen. With an average of 10,000 SB images, CE interpretation can be tedious, time-consuming, and may carry the risk of missed lesions and diagnostic uncertainty. Artificial intelligence (AI) tools have the potential to alleviate these limitations. A survey of 333 gastroenterologists, followed by a Delphi consensus study involving 21 experts across Europe, confirmed the interest and confidence of physicians in this development and identified key research questions. The creation of a first database comprising 4,174 videos and 5,184 images of interest was a fundamental step. In parallel, through three Delphi processes, in collaboration with international CE experts, we developed a ground truth for the nomenclature, description, and imputability (according to the clinical context) of the main SB lesions of interest (vascular lesions, ulcerative-inflammatory lesions). Based on this groundwork, we developed and evaluated several algorithms from dedicated neural networks for automatic CE reading. Finally, we established proof of concept for the good diagnostic performance of supervised neural networks for the detection and characterization of SB angiodysplasia (Se = 100%, Sp = 96%) and for the evaluation of SB cleanliness (Se = 90.3%, Sp = 83.3%). The results of this work are the subject of an industrial valorization through patent filings and the creation of a start-up. The large-scale implementation of AI in our society is already a disruption in our daily lives. A paradigm shift is necessary in the medical world with the arrival of these new tools and algorithms whose complex operation remains poorly understood by healthcare professionals. This technological revolution exposes us to new challenges and requires the involvement of various healthcare stakeholders, including healthcare professionals, to participate in its implementation in clinical practice and identify potential advantages and obstacles to its applications in healthcare. This work sits at the interface of this complex medical-technical world, combining medicine and image analysis. The answers provided by the application of AI in CE through this work offer a real opportunity to promote safer and more efficient care.

1. Introduction

Les différents travaux présentés dans ce manuscrit sont le fruit d'une collaboration étroite entre le laboratoire de recherche ETIS (Equipes Traitement de l'Information et Systèmes), un centre commun au CNRS, à l'université de Cergy Pontoise et à l'École nationale supérieure de l'électronique et de ses applications (ENSEA) et l'équipe du Centre d'Endoscopie Digestive de l'hôpital Saint Antoine (Sorbonne Université, APHP, Paris, France). Cette collaboration a débuté en 2015 autour de l'analyse d'image en endoscopie digestive pour se focaliser progressivement sur l'application des outils de l'intelligence artificielle (IA) à la vidéocapsule endoscopique (VCE) de l'intestin grêle (IG) (VCE-G). L'implémentation à grande échelle de l'IA dans notre société représente d'ores et déjà un bouleversement dans notre quotidien. Un changement de paradigme s'impose de fait au monde médical qui se retrouve lui aussi bousculé par l'arrivée de ces nouveaux outils et algorithmes dont le fonctionnement complexe reste peu maîtrisé par les médecins. Cette révolution technologique nous expose à des nouvelles problématiques et nécessite d'être accompagnée par les différents acteurs de santé - dont les soignants - afin de participer à son implémentation en pratique clinique et d'identifier les avantages et freins potentiels de ses applications en santé.

Pour ses défenseurs, l'IA est la promesse de choix diagnostiques et thérapeutiques précis et rapides, avec des performances démultipliées dépassant possiblement les capacités humaines d'analyse, ouvrant de nouvelles voies de recherche. Pour ses détracteurs, l'IA représente un danger de dépersonnalisation de la médecine, de perte des fondamentaux cliniques. Ce travail de thèse applicative vient se positionner à la frontière de ce monde médicotechnique. Nous tâcherons de réaliser un état de l'art de l'utilisation de la VCE-IG en médecine pour ensuite aborder les différents axes de recherche possible autour de l'application de l'IA à la VCE-G.

2. Epidémiologie en gastroentérologie

Les affections du tube digestif représentent un enjeu médical majeur de Santé Publique. Les cancers digestifs sont fréquents et regroupent les cancers du tractus digestif (étendu de l'œsophage à l'anus) mais également du pancréas et du foie et des voies biliaires. Le cancer colorectal (CCR) est le second cancer en fréquence chez la femme et le troisième chez l'homme. Le CCR est une des principales causes de décès par cancer dans le monde avec une incidence annuelle en France estimée à 43 000 nouveau cas correspondant à un décès toutes les demi-heures (1). Les maladies inflammatoires chroniques de l'intestin sont également fréquentes, avec en moyenne 8 nouveaux cas de maladie de Crohn (MC) qui se déclarent chaque année pour 100 000 habitants. Enfin, les hémorragies digestives touchent environ 1/1000 habitants par an, avec une mortalité autour de 15%.

La prise en charge de ces maladies en France représente un marché très important, malgré de nouvelles contraintes qui apparaissent. D'après une étude récente du Conseil National Professionnel d'Hépatogastroentérologie (CNP HGE), les dépenses de santé annuelles pour les maladies digestives sont d'environ 6 milliards d'euros pour un effectif de 3400 hépatogastroentérologues en France. Cependant, d'après cette même étude, ce marché est sous tension, puisque plus de 20% des hépatogastroentérologues ont plus de 55 ans (et cesseront donc leur activité à l'horizon des années 2030), sans renouvellement ajusté des générations du fait du numerus clausus des études médicales durant ces dernières décennies. A la fréquence et à la gravité des maladies digestives, se rajoute donc un enjeu, pour la santé publique française notamment, du temps (et du coût) médical pour leur prise en charge.

L'endoscopie digestive occupe une place centrale dans la prise en charge des maladies de l'appareil digestif (dépistage, diagnostic, surveillance et pronostic, traitement curatif et palliatif). Ainsi, à l'échelle française, et d'après les données de l'Assurance Maladie, environ 3 millions d'endoscopies digestives sont réalisées annuellement par les hépatogastroentérologues. Les endoscopies dites conventionnelles permettent l'exploration du tube digestif haut (à l'aide d'un gastroscopie) jusqu'aux premières anses duodénales et du tube digestif bas (à l'aide d'un coloscope) jusqu'à la dernière anse iléale (Figure 1). L'IG est longtemps resté une frontière anatomique du fait de sa longueur importante (4 à 6m) et de son éloignement des orifices naturels (voies d'abord des endoscopies digestives) (Figure 2).

Au début des années 2000, des avancées technologiques majeures (et notamment la miniaturisation des batteries et des composants électroniques) ont permis le développement et l'avènement de la VCE. Ce dispositif représente une véritable révolution dans l'exploration du tube digestif et est aujourd'hui devenu incontournable dans la prise en charge des patients suspects ou atteints de maladies de l'IG. En France, on réalise actuellement environ 25 000 VCE-G par an. D'autres techniques existent pour l'exploration de l'IG comme l'entéroscopie (effectuée sous anesthésie générale avec un entéroscope) mais cet examen est plutôt réalisé à visée thérapeutique. Certaines explorations radiologiques comme le scanner ou l'IRM (Imagerie par résonance magnétique) permettent également de visualiser l'IG de façon complémentaire.

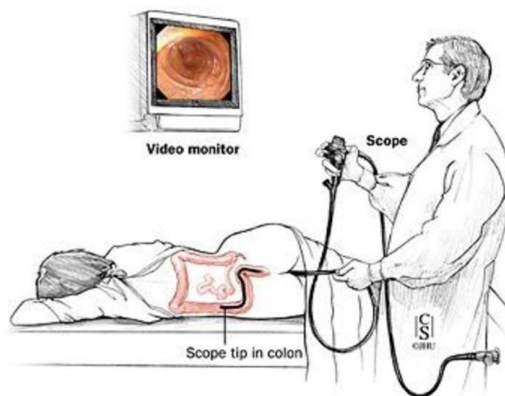


Figure 1 : Illustration de la réalisation d'une coloscopie (Johns Hopkins, Cancer Center, 2017)

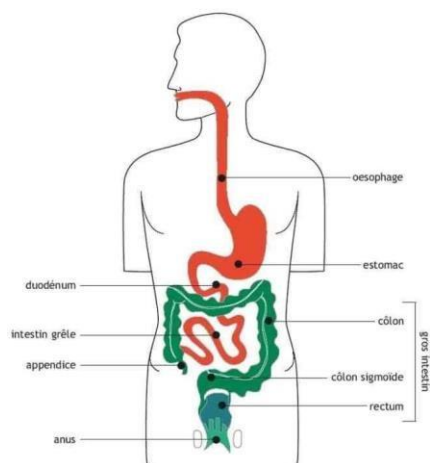


Figure 2 : Schéma anatomique du tube digestif

3. La vidéocapsule endoscopique

3.1. Présentation du dispositif de vidéocapsule endoscopique

La VCE s'est imposée comme un des examens de première ligne pour l'exploration de l'IG (2) (Figure 2).



Figure 3 : Modèle de vidéocapsule endoscopique de l'IG de 3ème génération (SB3 Pillcam[®], Medtronic)

En France, la quasi exclusivité du marché de VCE est occupé par la firme Medtronic et notamment sa troisième génération de dispositif de VCE-G. Ce dispositif est intégré dans une capsule de 26 X 11 mm et permet d'acquérir deux ou six images par seconde pendant une durée d'au moins huit heures afin d'obtenir une exploration de la totalité de la muqueuse de l'IG dans plus de 80% des cas (3) (Figure 3).

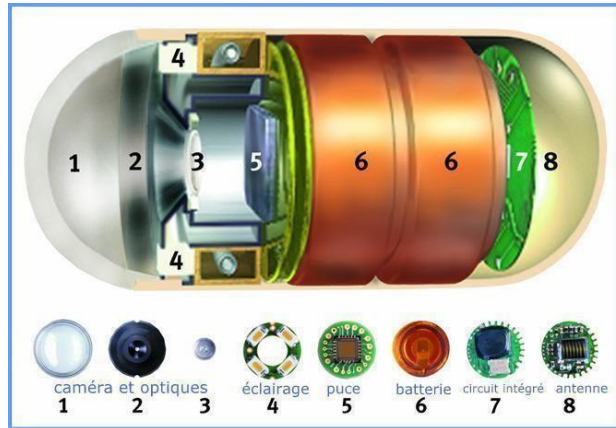


Figure 4 : Vue schématique d'une vidéocapsule endoscopique de l'IG de 3ème génération (SB3 Pillcam[®], Medtronic). 1 : Fenêtre (dome optique); 2,3. Lentille ; 4. Système d'éclairage: 4 diodes; 5: Puce (capteur d'images CMOS); 6: Système énergétique (batteries) ; 7 : Transmetteur ; 8 : Antenne

Les images émises par cette caméra miniaturisée sont enregistrées dans un boîtier par l'intermédiaire de capteurs positionnés sur l'abdomen du patient. Le développement de la VCE a été rendu possible par les avancées technologiques dans trois domaines : l'existence d'une puce électronique capable de fournir une image comparable à celle obtenue par un appareil photo, un système ASIC (*Application Specific Integrated Circuit*) qui autorise l'intégration d'un transmetteur vidéo de très petite taille, ainsi qu'un éclairage puissant miniaturisé de type LED (*light-emitting diode*). Le champ de vision obtenu est de 156° permettant une bonne visualisation de la muqueuse. La VCE-G a pour objectif de filmer l'intégralité de la muqueuse de l'IG et présente la particularité de s'ingérer par voie orale et d'être ensuite autonome dans son déplacement par le biais du péristaltisme intestinal. Contrairement à la coloscopie, elle ne nécessite ni anesthésie ni hospitalisation et est réalisée en « externe » (dans un circuit similaire à celui d'une consultation ou d'un examen radiologique). Après une période de jeûne nocturne, et en l'absence de contre-indication (situations à risque de fausse route ou de rétention) le patient ingère la capsule dans une structure de soin (cabinet médical, clinique, hôpital), boit éventuellement une préparation laxative, et retourne à son domicile, à ses activités de loisirs ou professionnelles. Lors de sa progression dans l'intestin, la capsule transmet (le plus souvent par radiofréquence) les images saisies à des capteurs reliés au boîtier d'enregistrement, portés en ceinture. La capsule

est éliminée par les voies naturelles, aux toilettes. En fin de journée, le patient rend le boîtier d'enregistrement (par lui-même, un proche ou un coursier) à partir duquel les images sont téléchargées. Elles sont ensuite lues, de façon différée, sous forme de vidéo, par le praticien (Figure 4) à l'aide d'un logiciel (*viewer*) dédié (figure 5). Un compte-rendu (avec éventuellement un diagnostic et une conduite à tenir) est établi et communiqué au patient et aux médecins correspondants.



Figure 5 : Image d'IG captée par vidéocapsule endoscopique (SB3 Pillcam®, Medtronic).



Figure 6 : Image du viewer (SB3 Pillcam®, Medtronic)

3.2. La vidéocapsule endoscopique en pratique clinique : quelles indications ?

Il est classiquement admis qu'environ 80% des hémorragies digestives ont une origine haute (avant l'angle de Treitz) (4) et 20% une origine basse (en aval de l'angle de Treitz) (5,6). Les hémorragies digestives basses sont à 80 % d'origine colo-anale. Les hémorragies digestives liées à l'IG représentent donc environ 5% des hémorragies digestives. La VCE-G est recommandée depuis 2006 par la HAS (Haute Autorité de Santé) comme examen de première intention pour l'exploration des suspicions de saignement de l'intestin grêle (SSIG) après endoscopie oesogastroduodénale et iléo-coloscopie normales. La VCE-G est également recommandée en cas de suspicion de MC de l'IG après des explorations endoscopiques et radiologiques n'ayant pas mis en évidence d'anomalie, et pour certains pour le suivi de la maladie (Figure 6). Une VCE peut être discutée dans d'autres indications plus rares telles que la maladie cœliaque résistante au régime sans gluten ou suspecte de complications (sprue réfractaire, lymphome T) ou encore les polyposes digestives (syndrome de Peutz-Jeghers, syndrome de Lynch) et autres tumeurs de l'IG (7–10). Le travail de Cortegoso Valdivia P. et al. (11) regroupant 86930 procédures de VCE offre une vision représentative globale des différentes indications de réalisation d'une VCE: SSIG (51%), explorations de symptômes digestifs (21%), suspicion de MC (21%), lésions néoplasiques (6%), maladie coeliaque (1%).

Indications de VCE	Rentabilité diagnostique
Suspicion de saignement de l'intestin grêle	59%
Suspicion de maladie de Crohn	62%
Polypose digestive (Syndrome de Lynch/ Peutz-Jeghers) / tumeur du grêle	63 %
Suivi de maladie coeliaque	52%

Figure 7 : Indications de réalisation d'une VCE et rentabilité diagnostique associée (11).

3.3. Rentabilité diagnostique de la vidéocapsule endoscopique

Les SSIG représentent environ deux tiers des indications de VCE-G. Dans cette indication, la rentabilité diagnostique est estimée entre 50 et 60 % dans la littérature (12,13).

Sa réalisation idéalement dans les 48 heures suivant le début du saignement est désormais recommandée par l'ESGE (Société Européenne d'Endoscopie digestive) afin d'optimiser la rentabilité diagnostique de l'examen (14). En contexte de SSIG, la VCE a une rentabilité diagnostique supérieure à celle de l'entéroscopie double ballon ou poussée et bien moins invasive (15–18). Par ailleurs la VCE a une sensibilité et une spécificité plus importantes que ces deux dernières modalités. Ainsi dans le travail de Saurin JC. et al. (19) la spécificité et la sensibilité dans le bilan étiologique des SSIG étaient respectivement de 92% et 80% pour la VCE, et 48% et 69% pour l'entéroscopie poussée ($p < 0,01$). La VCE-G est également indiquée en cas de suspicion de MC. Elle est complémentaire de l'entéro-IRM notamment pour son apport diagnostique des lésions jéjunales proximales et superficielles parfois difficiles à mettre en évidence en entéro-IRM en raison d'une part de la plus faible réplétion du jéjunum par rapport à l'iléon (20), d'autre part de leur caractère souvent subtil. L'essai prospectif mené par González-Suárez B. et al. (21) sur une cohorte de 47 patients atteints et/ou suspects de MC montrait ainsi une meilleure rentabilité diagnostique de la VCE par rapport à l'entéro-IRM (76,6% vs 44,7%, $p = 0,001$) dans le diagnostic des lésions de l'IG mais également dans la détection de lésions jéjunales inflammatoires (31,9% vs 6,4% $p = 0,03$). Pour autant, le faible effectif de patients inclus dans cette étude ne permet pas de généraliser ces résultats.

3.4. Les enjeux de la préparation de l'intestin grêle

Afin d'optimiser la visualisation de l'IG sur l'enregistrement, il est recommandé de réaliser une préparation associant un régime liquide et sans résidus végétaux la veille de l'examen, un jeûne nocturne. La prise (et ses modalités) d'une solution laxative est plus controversée. L'ESGE recommande l'ingestion par le patient d'une préparation avant une VCE-G (22) tandis que la Société Américaine d'Endoscopie Digestive (ASGE) recommande un régime liquide clair préalable comme attitude validée en mentionnant tout de même l'intérêt des purges (23). Il existe différents régimes et modalités de purge en ce qui concerne le type, le volume et le moment d'ingestion qui ne font pas l'objet de recommandations claires dans la littérature médicale. Pour autant, une qualité de préparation optimale est essentielle pour réaliser un diagnostic fiable en VCE. L'essai contrôlé randomisé Prepintest (24) a montré que l'ingestion ou non d'une purge n'augmente pas la rentabilité diagnostique. Le travail de Yung DE et al.

(25), auquel nous avons contribué, démontre qu'une mauvaise qualité de visualisation des images est associée à un rendement diagnostique plus faible dans les tumeurs malignes de l'IG. L'ESGE a établi des critères de qualité de la réalisation des VCE-G en fixant un objectif supérieur à 95 % de procédure avec visualisation adéquate de l'IG (26). Afin d'y parvenir, l'utilisation d'agents anti-bulles est suggérée pour réduire la présence de bulles d'air gênant la visualisation. Un critère de qualité supplémentaire établi par l'ESGE concerne le taux de complétude de l'analyse de l'IG (défini par l'analyse complète de la première à la dernière image de l'IG) supérieur à 80 % pour lequel certains traitements prokinétiques visant à améliorer la vidange gastrique peuvent être proposés au cas par cas.

4. L'Intelligence artificielle

4.1. Généralités sur l'intelligence artificielle

Le dictionnaire Larousse définit l'IA comme « *un ensemble de théories et de techniques mises en œuvre en vue de faire réaliser des machines capables de simuler l'intelligence* ». L'objectif de l'IA se résume donc à imiter les fonctions de résolution de problèmes et de prise de décision du cerveau humain. C'est lors de la conférence de Dartmouth en 1956 que le terme IA est pour la première fois prononcé (27), avec des objectifs définis et ambitieux qui prédisent une machine capable de reproduire des tâches cognitives humaines. Dans les années 60, le premier réseau de neurones artificiel, appelé perceptron, est proposé par le psychologue Rosenblatt. Le perceptron s'inspire du fonctionnement des circuits neuronaux humains comprenant une entrée et une sortie toutes deux reliées par une arborescence synaptique pondérée et prédéterminée. La mise en réseau de ces neurones artificiels ouvre la voie à l'apprentissage automatique (*Machine Learning*, ML). Malgré des perspectives initiales enthousiastes accompagnées d'investissements massifs, le développement de l'IA connaît par la suite un essoufflement en partie attribué aux capacités limitées de calcul de la micro-informatique à cette époque. Après l'âge d'or initial, cette période est décrite comme une série « d'hivers de l'IA ». A partir des années 2000, l'arrivée des processeurs graphiques GPU (*Graphics Processing Unit*) offre une puissance de calcul et de traitement rapide des données et permet à l'IA de reprendre son essor.

4.2. Terminologie en vision par ordinateur

Avant de poursuivre il apparait indispensable avant tout de clarifier le vocabulaire utilisé dans le domaine de l'analyse d'image par ordinateur. La perception d'une image par un ordinateur se réduit à un ensemble de pixels constituant un tableau de matrices correspondant à des données statistiques (Figure 7). L'œil humain en analysant cette photo reconnaît un chat avec certitude mais serions-nous capables de décrire en détail les différentes étapes du cheminement cérébral nous permettant d'identifier cet animal ?

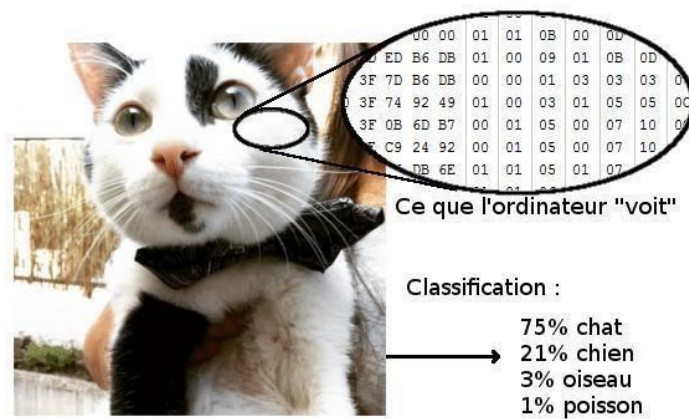


Figure 8 : Illustration de l'analyse d'une image par un ordinateur

En vision par ordinateur, le terme classification correspond au regroupement des différents pixels dans des classes spécifiques. La localisation d'un motif (*pattern*) au sein de l'image est le fait d'apposer un cadre autour du chat dans l'image d'intérêt. L'ordinateur est également capable de détecter plusieurs objets dans l'image pour réaliser une segmentation de l'image, c'est-à-dire la délimitation précise des objets au sein de l'image (Figure 8).

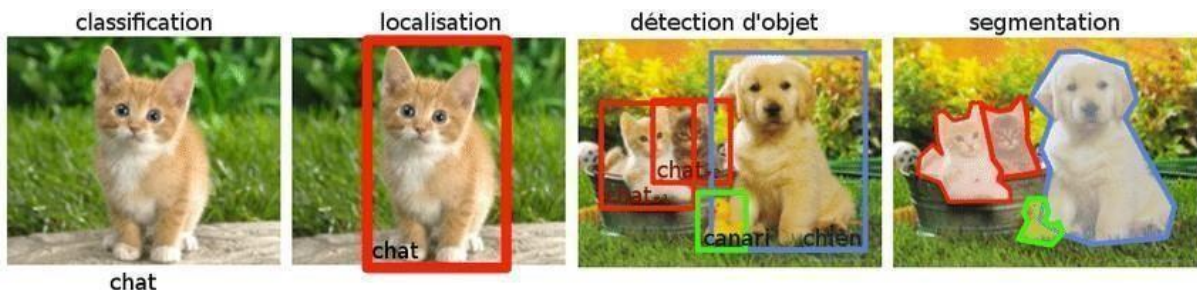


Figure 9 : Illustration de la terminologie en vision par ordinateur
De gauche à droite : classification, localisation, détections multiples et segmentation

4.3. L'apprentissage machine (Machine Learning)

Le *machine learning* est un sous domaine de l'IA. Il consiste à développer des algorithmes capables de reconnaître des motifs (*patterns*) récurrents dans un jeu de données (images, statistiques...). Les algorithmes sont capables d'apprendre mais également de s'améliorer de façon autonome dans le but de réaliser des prédictions. Le développement d'un algorithme répond à une suite systématique d'étapes avec initialement une phase d'entraînement de l'algorithme sur une base de données permettant de déterminer les paramètres du modèle minimisant le risque d'erreurs, puis des phases d'amélioration et enfin de validation du modèle sur des bases de données distinctes, indépendantes.

4.3.1. L'apprentissage supervisé

L'apprentissage supervisé est une méthode de ML nécessitant l'utilisation de données annotées et délimitées manuellement (sous forme de régions d'intérêts ou *regions of interest*, ROI). L'annotation va permettre de classer l'image dans une catégorie spécifique (anomalies vasculaire/inflammatoire) tandis que la délimitation s'applique à contourer la zone d'intérêt. L'apprentissage de l'algorithme est donc guidé par ces données et détient les prédictions de sortie du modèle permettant d'affiner les performances en diminuant l'écart entre les résultats attendus et obtenus. Cet apprentissage est exigeant notamment de par la difficulté et le caractère chronophage pour générer ces données de qualité.

4.3.2. L'apprentissage non supervisé

L'apprentissage non supervisé est une méthode de ML n'utilisant pas de données préalablement délimitées. Contrairement à l'apprentissage supervisé, les algorithmes apprennent dans ce cas à effectuer des regroupements de données en fonction de leurs similarités ou de leurs différences. Cette méthode permet de réduire l'effort chronophage et onéreux de délimitation des données.

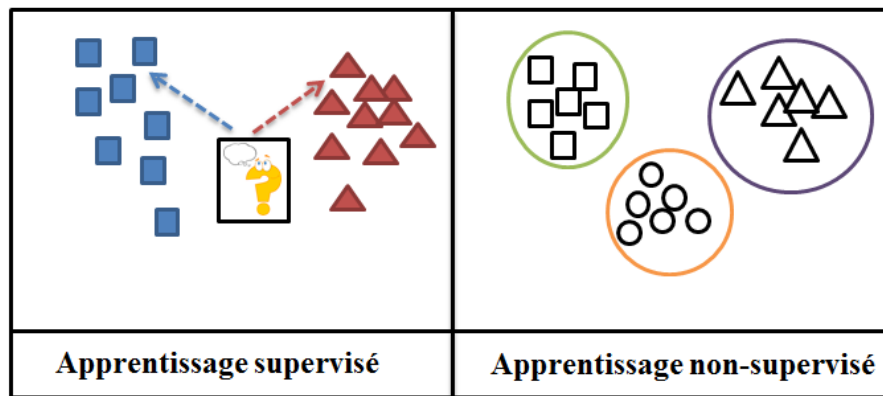


Figure 10 : Illustration d'un apprentissage supervisé et non supervisé

4.3.3. L'apprentissage profond (Deep Learning)

L'apprentissage profond (ou *Deep Learning*, DL) est une technique majeure de ML reposant sur le modèle des réseaux de neurones (RN) artificiels (Figure 11). Ces RN sont capables de mimer le fonctionnement du système nerveux humain. Ils sont composés de plusieurs couches de neurones interconnectées ayant des fonctions différentes. Les caractéristiques d'une donnée analysée par un RN sont transmises sous la forme d'informations que le neurone va recevoir et quantifier selon ses fonctions d'activation. Cette information reçue par le neurone, appelée « intrant », est quantifiée et traduite selon trois niveaux : en dessous du seuil (le neurone reste inactif), aux alentours du seuil (phase de transition), au-dessus du seuil (activation du neurone). Cette valeur seuil va servir de référence pour déclencher ou non la transmission d'une information et influencer le résultat final. Chaque couche de neurones est associée à un « poids » qui peut ou non être préalablement déterminé lors de la phase d'apprentissage en fonction des prédictions attendues de l'algorithme. La structure de RN la plus pertinente dans l'analyse d'images et notamment dans le domaine médical est le RN convolutif (*Convolutional Neural Network*, CNN) (28). Les CNN s'organisent par couches de convolutions successives. Chaque couche de convolution analyse une partie de l'image d'entrée puis multiplie les valeurs des pixels à l'aide d'une matrice prédéfinie (pour capturer des courbes, des lignes, des structures) pour émettre le résultat sur la couche suivante de neurones. Les couches de convolutions empilées successivement permettent d'extraire des informations de plus en plus simples (Figure 12).

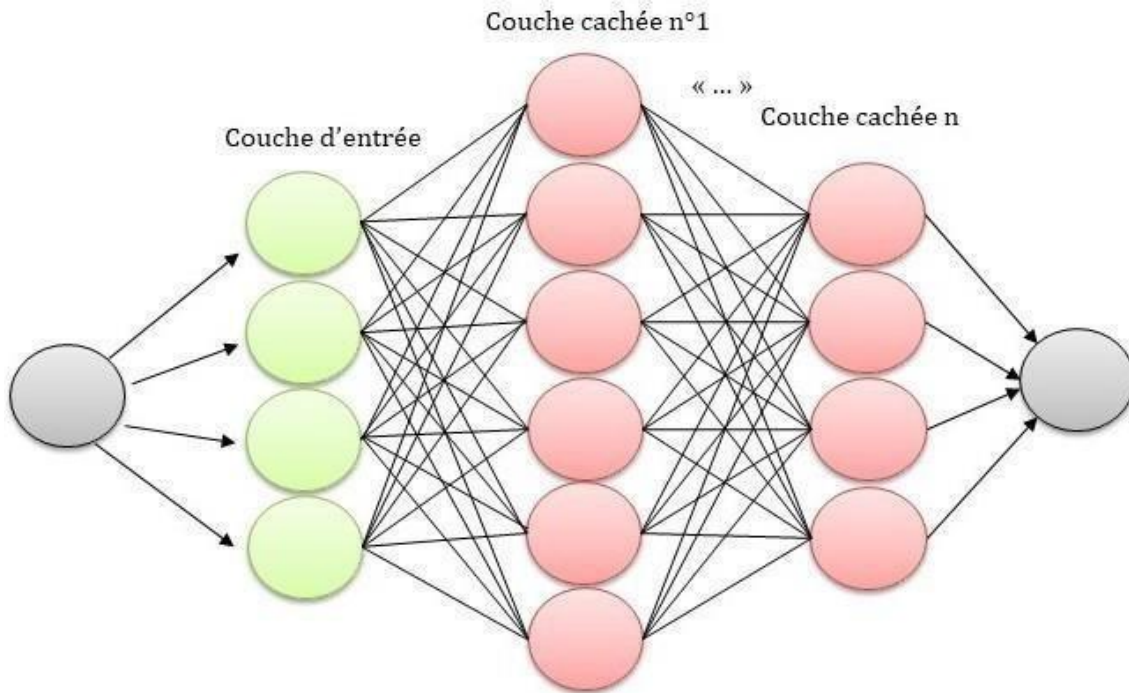


Figure 11 : Illustration d'un réseau de neurones artificiel. (Création personnelle)

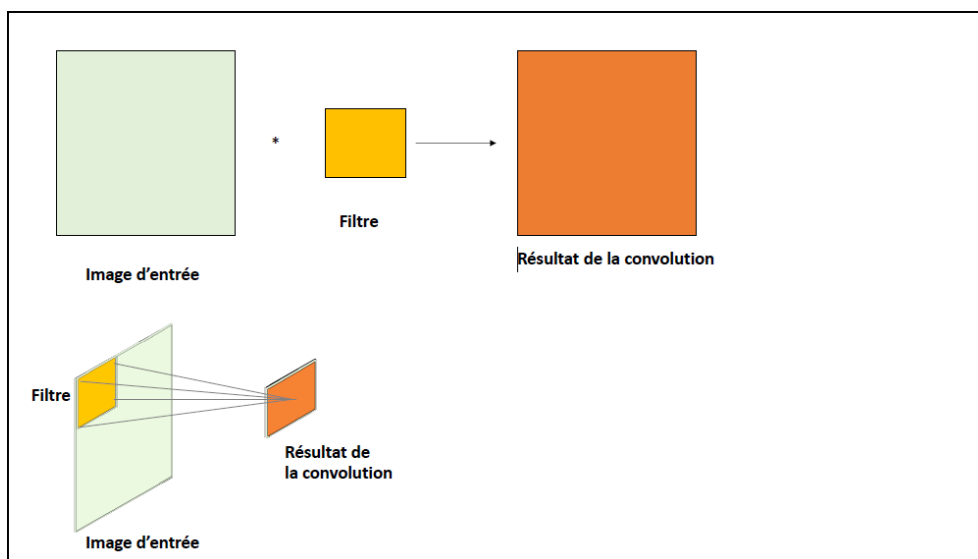


Figure 12 : Illustration du fonctionnement d'une convolution. (Création Dr Charles Houdeville)

Le principal avantage des CNN est leur capacité à extraire automatiquement les caractéristiques profondes de chaque image (texture, contours, niveau de gris...) en les compressant de façon à réduire leur taille initiale (processus de Pooling). Par la suite, une fonction d'activation (exemple fréquent : ReLU pour Rectified Linea Units) va transformer la donnée en changeant sa représentation. Dans ce cas particulier, ReLU permet de rendre le modèle non linéaire et donc plus complexe en modifiant spatialement la représentation de la donnée.

5. Applications de l'intelligence artificielle en médecine

Cette dernière décennie a connu des progrès majeurs grâce à l'avènement du ML qui a ouvert la voie à un vaste champ de développements liés à ses applications notamment dans le domaine médical (29). Les applications de l'IA en santé offrent la promesse d'une médecine personnalisée, préventive, prédictive et également participative. L'IA s'invite désormais sur des appareils grand public (téléphone, dispositif connecté, etc) et offre à titre d'exemple la possibilité pour le patient de détecter seul le caractère éventuellement malin d'une lésion dermatologique avec une fiabilité supérieure à celle de l'expert (30). Les métiers s'articulant autour de l'analyse d'images sont particulièrement sollicités par l'émergence de ces outils avec notamment la radiologie mais également l'endoscopie digestive en première ligne. L'application majeure de l'IA en endoscopie digestive se concentre sur le développement d'outils d'aide à la détection et à la caractérisation en temps réel des polypes colorectaux lors d'une coloscopie (IA assistive, en complément de la vision humaine) (31). L'implémentation de l'IA en VCE est confrontée à un défi presque opposé de par le caractère différé de la lecture d'un enregistrement de VCE, d'un scanner ou d'une IRM (IA assertive, écartant de la lecture humaine de dizaines de milliers d'images jugées non pertinentes).

6. Premières tentatives d'intelligence artificielle appliquées à la lecture des VCE

Si la VCE s'est imposée depuis plus de vingt ans comme un des examens de référence dans l'exploration de l'IG, la lecture des enregistrements demeure longue et fastidieuse pour le médecin (30 à 40 minutes). Cette lecture, chronophage et coûteuse en temps médical, est associée à un risque d'erreurs et de lésions manquées démontré dans la littérature médicale dès la deuxième lecture successive de VCE (32). De plus, le délai de lecture de

l'enregistrement diffère souvent le rendu des résultats au patient ce qui est potentiellement préjudiciable pour la prise en charge médicale de ce dernier mais également pour l'anxiété éventuelle que ces délais peuvent générer.

En 2019, au démarrage de ce travail, certaines solutions facilitant la lecture des VCE étaient d'emblée disponibles et proposées par les fournisseurs de VCE. Le système Pillcam (Medtronic) proposait déjà un premier outil informatique nommé SBI (*small bowel blood indicator*) pour l'aide à la détection des lésions à potentiel hémorragique. Pour autant, les performances diagnostiques de cet indicateur demeurent insuffisantes avec une sensibilité de 55% et une spécificité de 58 % (33) ce qui limite son utilisation en pratique clinique. Par la suite, des algorithmes de ML ont été développés de façon à supprimer les images similaires ou redondantes pour accélérer la lecture des VCE. Les principes de fonctionnement de ces algorithmes relèvent du secret industriel. Le système QuickView® en est un exemple avec des performances issues d'un essai prospective multicentrique menés sur 106 vidéos montrant une sensibilité de 85,1 % et une spécificité de 84,7% similaire à celle d'une lecture conventionnelle. Cet outil permet de réduire le temps de lecture à 11,6 minutes, en moyenne (34).

L'avènement des outils d'IA et leurs applications en médecine a d'ores et déjà bouleversé nos pratiques et offre des perspectives de développement importantes notamment dans le domaine de l'analyse d'image. L'implémentation de l'IA en VCE suscite de nombreux questionnements et comme à chaque grande révolution ou changement de paradigme, les opinions divergent sur l'intérêt d'adopter ou non ces nouvelles technologies dans notre pratique clinique.

7. Applications de l'intelligence artificielle en vidéocapsule endoscopique

7.1. L'intelligence artificielle en vidéocapsule : quelles perceptions et attentes ?

Les perceptions et attentes autour de l'implémentation de l'IA en VCE soulèvent de nombreux questionnements avec des opinions divergentes qui animent les débats. Afin d'approcher et d'évaluer au mieux ces problématiques, nous avons mené un observatoire international incluant 20 pays et 333 lecteurs de VCE (35). Les résultats de ce travail

renforcent la volonté partagée des lecteurs de VCE d'implémenter l'IA en pratique clinique afin d'améliorer le diagnostic et la caractérisation et surtout de raccourcir le temps de lecture des VCE. En revanche, une majorité d'experts s'opposent au principe de déléguer pleinement à l'IA la lecture des VCE. Un sentiment majoritaire favorable soutient l'implémentation de cet outil comme une aide à la lecture et à l'interprétation des VCE.

Article

PEACE: Perception and Expectations toward Artificial Intelligence in Capsule Endoscopy

Romain Leenhardt ^{1,2}, Ignacio Fernandez-Urien Sainz ³, Emanuele Rondonotti ⁴, Ervin Toth ⁵ ,
Cedric Van de Bruaene ⁶ , Peter Baltes ⁷, Bruno Joel Rosa ^{8,9}, Konstantinos Triantafyllou ¹⁰ ,
Aymeric Histace ² , Anastasios Koulaouzidis ¹¹ , Xavier Dray ^{1,2,*} and on behalf of the I-CARE Group [†]

¹ Endoscopy Unit, Saint Antoine Hospital, Sorbonne University, APHP, 75012 Paris, France; romain.leenhardt@aphp.fr

² ETIS UMR 8051, CY Paris Cergy University, ENSEA, CNRS, 95000 Cergy-Pontoise, France; aymeric.histace@ensea.fr

³ Gastroenterology, Hospital de Navarra, 31008 Pamplona, Spain; ifurien@yahoo.es

⁴ Gastroenterology Unit, Valduce Hospital, 22100 Como, Italy; ema.rondo@gmail.com

⁵ Department of Gastroenterology, Skane University Hospital, Lund University, 214 28 Malmo, Sweden; ervin.toth@med.lu.se

⁶ Department of Gastroenterology, Ghent University Hospital, 9000 Ghent, Belgium; Cedric.vandebruaene@ugent.be

⁷ Klinik für Innere Medizin, Agaplesion Bethesda Krankenhaus Bergedorf, 21029 Hamburg, Germany; baltes@bkb.info

⁸ Department of Gastroenterology, Hospital da Senhora da Oliveira, 4835-044 Guimarães, Portugal; bruno.joel.rosa@gmail.com

⁹ Life and Health Sciences Research Institute, School of Medicine, University of Minho, 4704-553 Braga, Portugal

¹⁰ Hepatogastroenterology Unit, Second Department of Internal Propaedeutic Medicine, Attikon University General Hospital, Medical School, National and Kapodistrian University of Athens, 10679 Athens, Greece; kttriant@med.uoa.gr

¹¹ Department of Social Medicine & Public Health, Faculty of Health Sciences, Pomeranian Medical University, 70-204 Szczecin, Poland; akoulaouzidis@hotmail.com

* Correspondence: xavier.dray@aphp.fr; Tel.: +33-149282000

† I-CARE (International Capsule endoscopy REsearch) group members.



Citation: Leenhardt, R.; Fernandez-Urien Sainz, I.; Rondonotti, E.; Toth, E.; Van de Bruaene, C.; Baltes, P.; Rosa, B.J.; Triantafyllou, K.; Histace, A.; Koulaouzidis, A.; et al. PEACE: Perception and Expectations toward Artificial Intelligence in Capsule Endoscopy. *J. Clin. Med.* **2021**, *10*, 5708. <https://doi.org/10.3390/jcm10235708>

Academic Editor: Hajime Isomoto

Received: 27 October 2021

Accepted: 22 November 2021

Published: 6 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland.

Abstract: Artificial intelligence (AI) has shown promising results in digestive endoscopy, especially in capsule endoscopy (CE). However, some physicians still have some difficulties and fear the advent of this technology. We aimed to evaluate the perceptions and current sentiments toward the use of AI in CE. An online survey questionnaire was sent to an audience of gastroenterologists. In addition, several European national leaders of the International Capsule endoscopy REsearch (I CARE) Group were asked to disseminate an online survey among their national communities of CE readers (CER). The survey included 32 questions regarding general information, perceptions of AI, and its use in daily life, medicine, endoscopy, and CE. Among 380 European gastroenterologists who answered this survey, 333 (88%) were CERs. The mean average time length of experience in CE reading was 9.9 years (0.5–22). A majority of CERs agreed that AI would positively impact CE, shorten CE reading time, and help standardize reporting in CE and characterize lesions seen in CE. Nevertheless, in the foreseeable future, a majority of CERs disagreed with the complete replacement all CE reading by AI. Most CERs believed in the high potential of AI for becoming a valuable tool for automated diagnosis and for shortening the reading time. Currently, the perception is that AI will not replace CE reading.

Keywords: small bowel capsule endoscopy; artificial intelligence; perceptions and sentiments; machine learning

1. Introduction

Artificial intelligence (AI) is a disruptive technology, especially in medical image analysis [1,2]. AI is already assisting physicians in various clinical tasks and has shown

promising results especially in digestive endoscopy [3–5]. To support and promote this technological innovation, the European Society of Gastrointestinal Endoscopy (ESGE) “suggests the possible incorporation of computer-aided diagnosis (CADx) to colonoscopy” [6]. Capsule endoscopy (CE) generates a very high amount of data involving gastroenterologists in a tedious task of reviewing full-length small-bowel (SB) CE videos. The use of AI in CE has already been studied in the relevant medical literature [4,7–9], and most authors seem optimistic and/or even ready to accept a tremendous transition. However, although AI in CE is already a commercial solution, some physicians still have difficulties accepting or some concerns related to this technology. Therefore, we aimed to evaluate the perceptions and current sentiment toward AI in CE.

2. Materials and Methods

An online and anonymized English survey questionnaire was designed by three gastroenterologists (AK, RL and XD). This questionnaire was distributed through GoogleForm® (Mountain View, CA, USA) to European national leaders of the International CAPsule endoscopy REsearch (I-CARE) group. In addition, they were asked to disseminate this online survey among their national CERs communities. Thus, participants tended to represent a diversity of training backgrounds. The survey contained 32 questions (Appendix A) divided into 6 rubrics regarding general information (physicians’ level of training and experience, practice characteristics), interest in new technologies in general, but also in endoscopy and CE, perceptions regarding the benefits of AI in endoscopy and CE, and perceptions regarding the barriers to implementing AI in CE. The term AI was defined as its general sense of “smart machines capable of performing tasks that typically require human intelligence”. Inclusion criteria for valid questionnaire returns was being a physician CER.

2.1. Agreement to Statements

For most questions, when addressing semi-quantitative questions, respondents were asked to rate their position on a numerical Likert response scale from 1 (strongly disagree) to 6 (strongly agree).

2.2. Statistical Analysis

Statistical analysis and graphical plotting were performed using Rstudio software (Version 1.2.5033, Boston, MA, USA). Qualitative variables were expressed as proportions, percentages, and raw numbers. Statistical comparisons between subgroups of responders (fellows vs. accredited gastroenterologists/below or above 40 years old, and below or above 50 years old) were performed using Chi-2 tests. Statistical significance was set at $p < 0.05$.

3. Results

3.1. Physicians’ Characteristics

A total of 380 gastroenterologists answered the questionnaire from December 2020 to March 2021 after three reminders. Most of the respondents were accredited gastroenterologist ($n = 302$; 80%), GI residents/fellows ($n = 70$; 18%), and non-medical jobs (nurse, technicians, or other healthcare providers) ($n = 8$; 2%). Overall, 20 different countries were represented. The three most active countries in terms of responders were France ($n = 62$; 16%), Spain ($n = 56$; 15%), and Italy ($n = 45$; 12%). We decided to exclude non-physician capsule readers responders ($n = 7$) from the final analysis to avoid response bias. Among responders, 333 (88%) were identified as physician CERs including 85% of accredited gastroenterologists and 15% of GI resident/fellow (Figure 1). Most CERs were between 40 and 50 years of age ($n = 109$; 33%), respectively followed by the 50–60 age group ($n = 96$; 29%), the 30–40 age group ($n = 80$; 24%), the 60–70 age group ($n = 34$; 10%), and the 20–30 age group ($n = 14$; 4%). The main practice setting was university/teaching hospital ($n = 214$; 64%), non-university hospital/district hospital ($n = 87$; 26%), and private practice ($n = 32$; 10%). The mean CE reading experience was 9.9 years (range 1–22), the current average

number of CE read per year was 67 (range 2–500), and the current average number of CE readings supervised per year was 38 (range 0–500) (Figure 2).

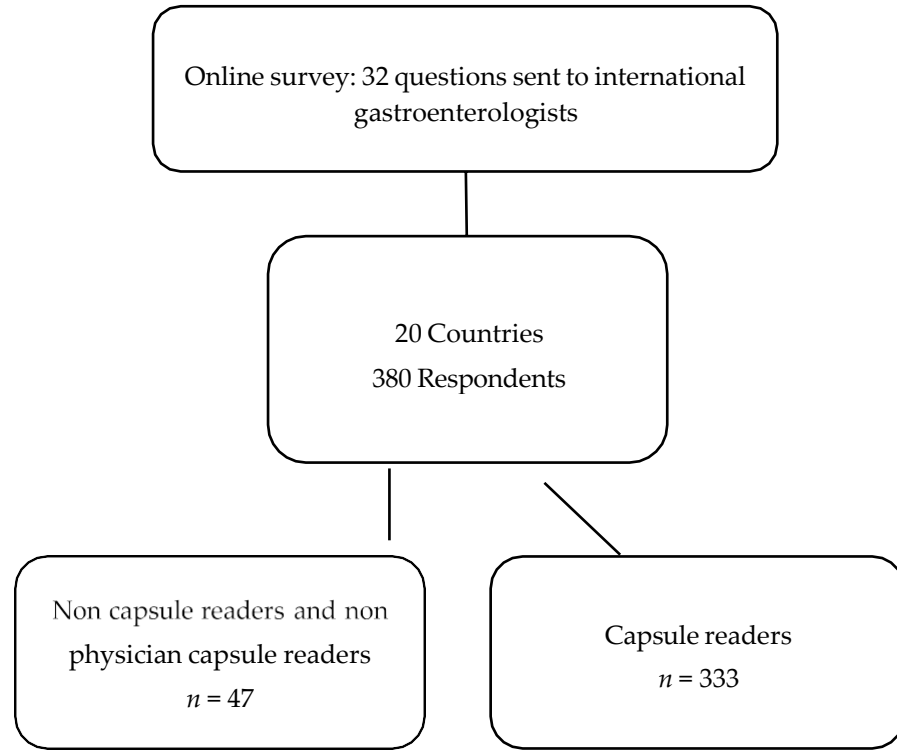


Figure 1. Flowchart of the study.



Figure 2. Cont.

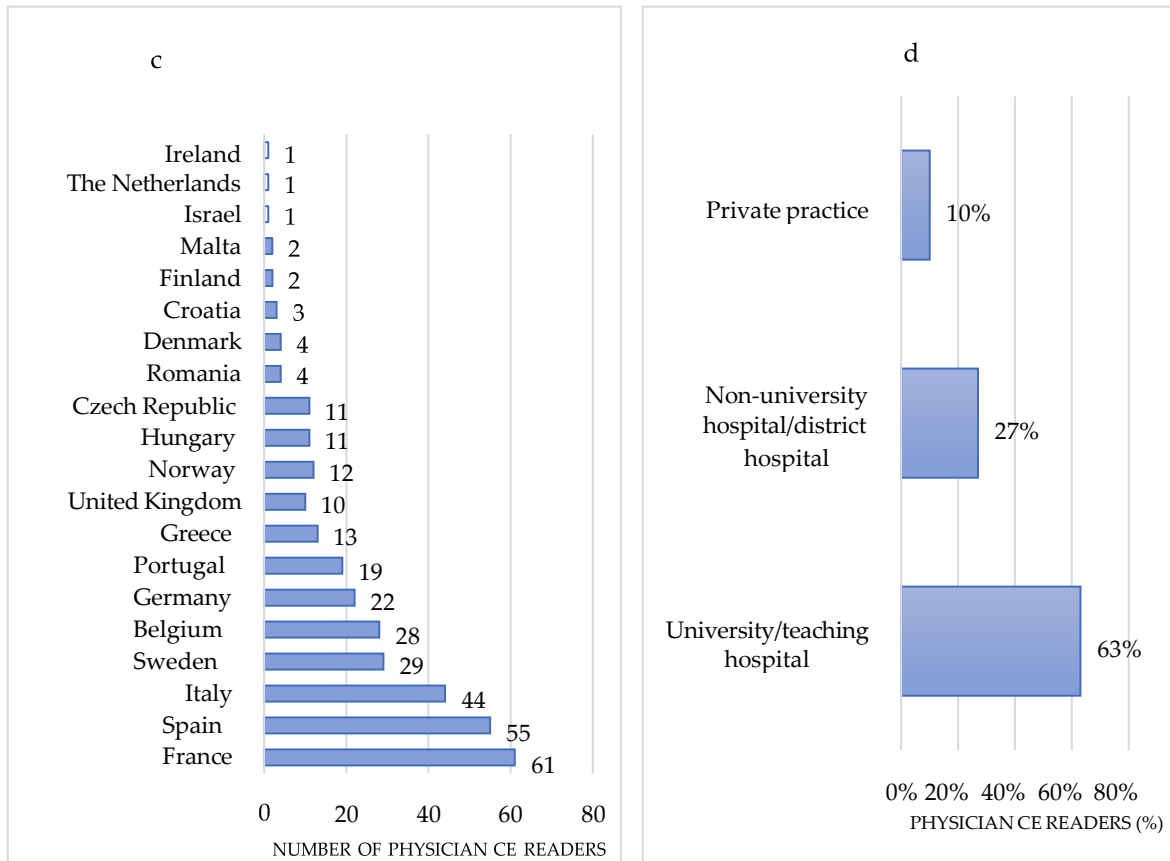


Figure 2. Capsule readers characteristics: (a) Current position of physician CE readers; (b) Age distribution of capsule readers; (c) Survey responders by country (Physician capsule readers only); (d) Main practice setting. CE: Capsule endoscopy; GI: Gastrointestinal.

3.2. Level of Information on AI

Among CERs, 38% ($n = 126$) did not consider themselves as well informed about the use of new technologies, especially informatics. Regarding CER only, 45% ($n = 151$) were aware of using any AI applications such as, for example, speech recognition, spam filters, or recommendation algorithms. Forty-six percent ($n = 152$) did not already use endoscopy AI solutions in their regular practice, but 36% ($n = 121$) planned to do so. Seventy-five percent ($n = 251$) were interested or strongly interested in receiving a generic/baseline training on AI, and 68% ($n = 226$) were interested or strongly interested in AI being part of the endoscopy training. Two-thirds (66%; $n = 221$) of responders thought that the introduction of AI would not change the relationship between physicians and patients, whereas 24% ($n = 80$) believed that this relationship would become more interactive, and 10% ($n = 30$) believed that this relationship would become more impersonal (Figure 3).

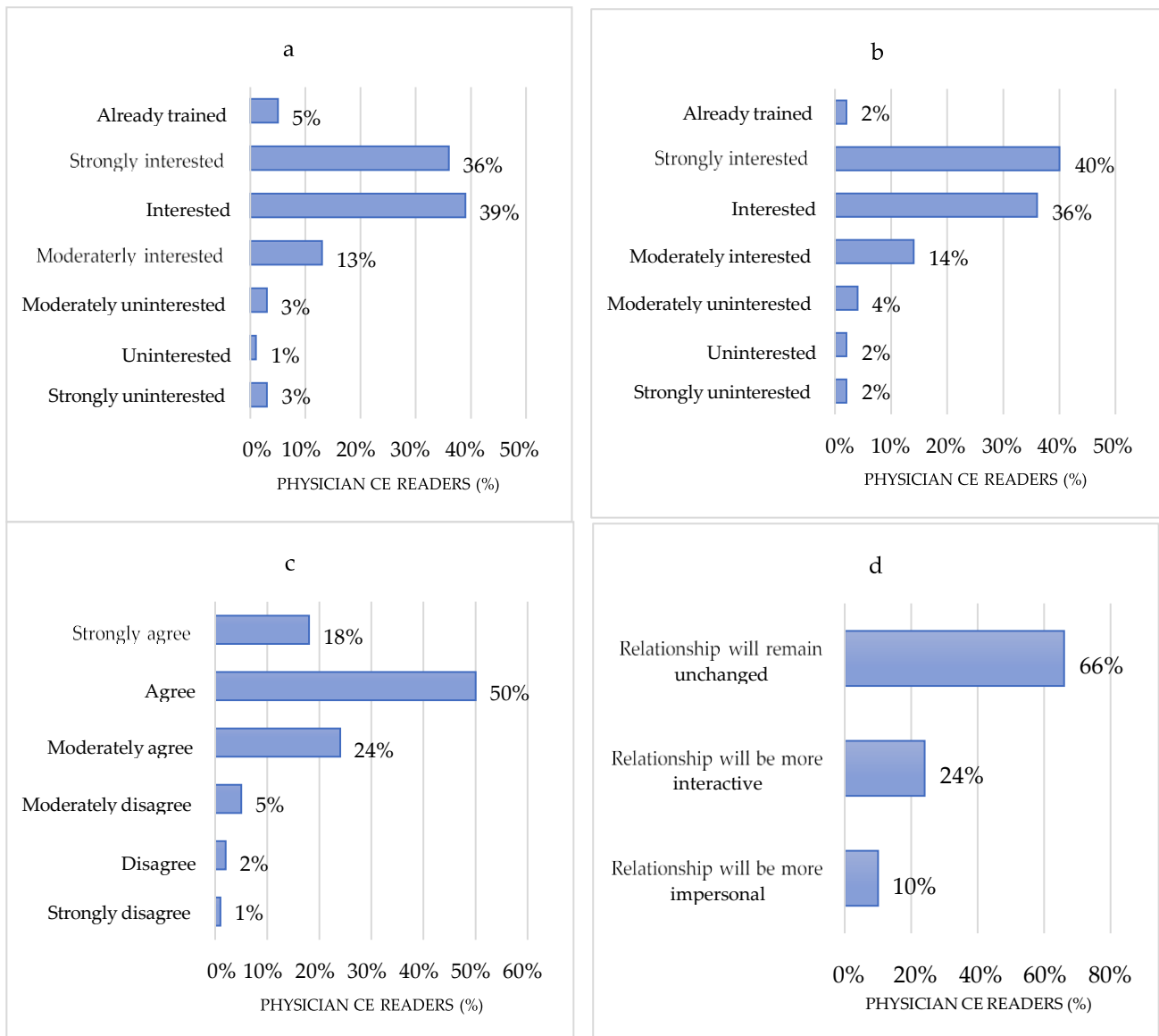


Figure 3. Level of information of capsule readers on artificial intelligence in general and in medicine. (a) How interested are you in receiving a generic/baseline training on AI?; (b) How interested are you in receiving a technically advanced training on AI?; (c) AI should be part of the endoscopy training; (d) How will the relationship between the endoscopist and the patient change with the introduction of AI?

3.3. Perceptions of AI toward Capsule Endoscopy

Eighty per cent ($n = 268$) of CERs agreed or strongly agreed that AI will positively impact CE, and 79% ($n = 265$) agreed or strongly agreed that AI would shorten the reading time in CE. Interestingly, 71% of CERs ($n = 236$) already used quick CE reading modes in their daily practice. Moreover, 74% ($n = 246$) and 68% ($n = 226$), respectively, agreed or strongly agreed that AI would help standardize reporting in CE and characterize lesions seen in CE. Nevertheless, 71% of CERs ($n = 236$) at least moderately disagreed with the idea that AI would replace them at work. However, most CERs (68%, $n = 227$) agreed or strongly agreed that AI would allow a pre-triage of CE examinations by likelihood of being normal (Figure 4). These proportions did not vary significantly, according to the professional level (fellows vs. accredited gastroenterologists) or to the age class (below or above 40 years old, and below or above 50 years old).

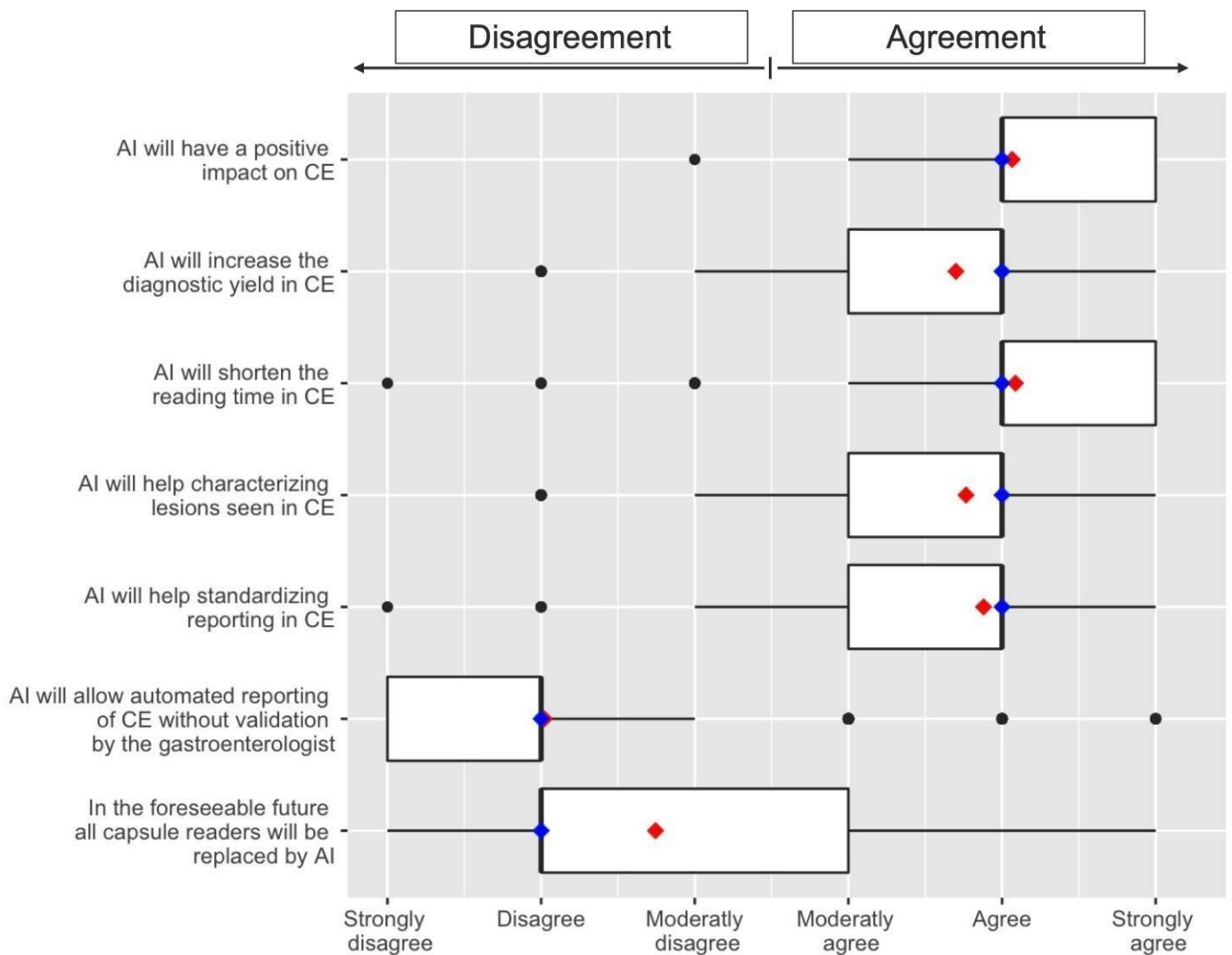


Figure 4. Perceptions of AI toward capsule endoscopy. Box plots show the distribution of responses using a Likert scale ranging from 1 to 6 (strongly disagree to strongly agree). The mean value is plotted with a red dot. The vertical blue bar represents the median value; 25th and 75th percentiles are set as box limits.

Regarding CE reports, 79% ($n = 265$) of CERs disagreed or strongly disagreed that AI would allow automated reporting without validation by the gastroenterologist. Regarding the time limit to get the results from an AI solution for CE reading, the majority of CERs (59%; $n = 196$) believed that a few minutes would be reasonable, whereas 29% ($n = 96$) would tolerate a few hours. Moreover, only 44% of CERs ($n = 157$) believed that AI would increase the number of requests for CE examinations. Finally, the maximum mean number of false positive frames per examination that CER would be ready to review was 222 (ranging from 50 to 1000).

3.4. Perceptions Regarding the Barriers for AI Implementation

Sixty percent of CR ($n = 201$) at least moderately agreed that AI implementation could lead to operator dependence. Although 51% of CER ($n = 170$) thought that patients should be aware of the use of AI in CE reading, 53% ($n = 175$) of CER believed that patients should not expressly consent to AI in CE reading. Regarding who should take the responsibility of AI-based reporting in CE, 75% of CERs ($n = 250$) believed that the responsibility remains at least with the endoscopist. In contrast, the other CERs believed that responsibility should be shared between the clinician requesting the exam, the developers of AI applications, and

insurance companies. Finally, regarding the cost of AI implementation in CE, 65% ($n = 218$) of CERs agreed or strongly agreed that this concern is of significant importance (Figure 5).

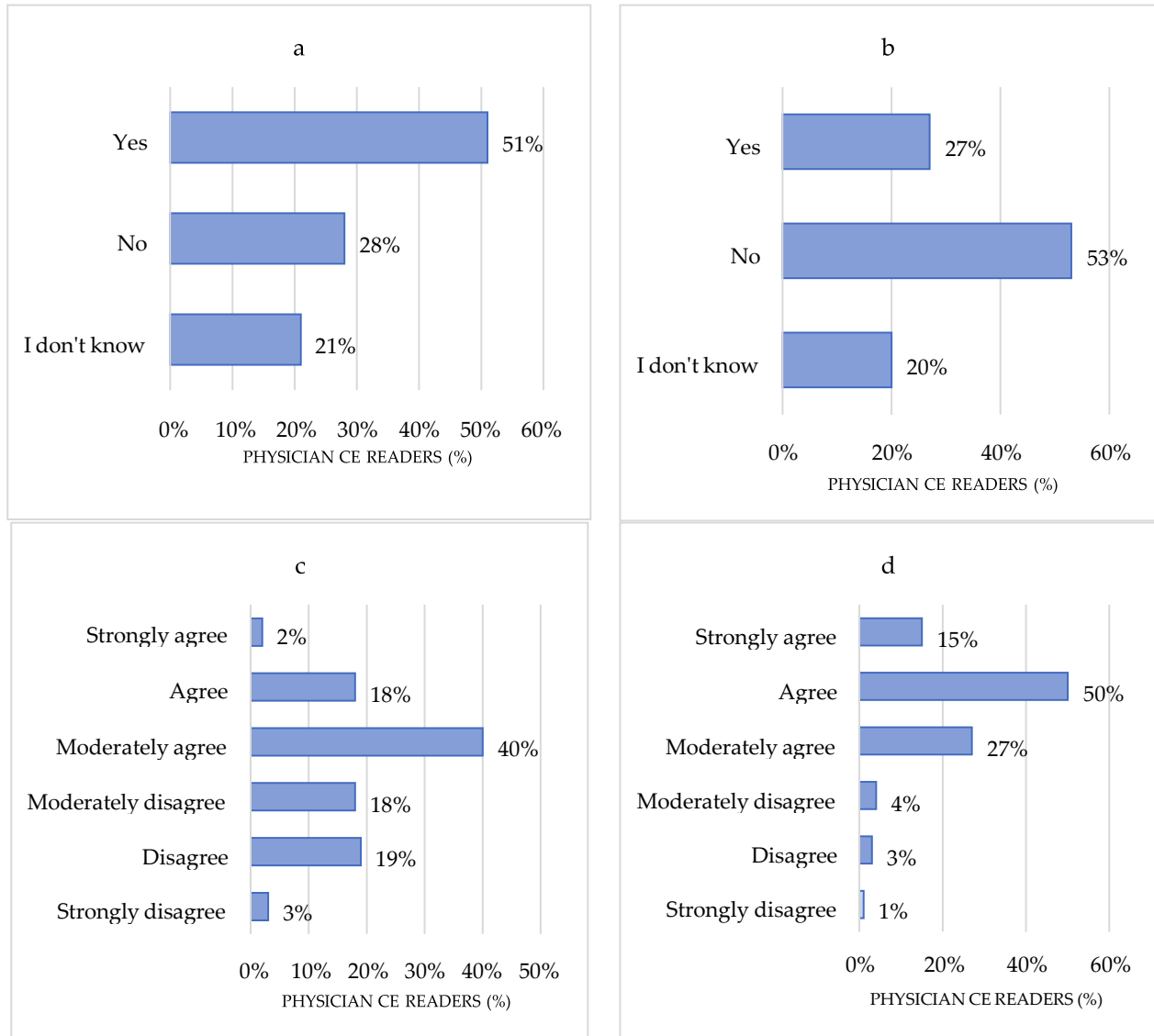


Figure 5. Perceptions regarding the barriers for AI implementation. (a) Patients should be aware about the use of AI in CE reading; (b) Patients should specifically consent to the use of AI in CE reading; (c) AI in CE will lead to operator dependence on the technology; (d) Cost is important in implementing AI in CE reading. CE: Capsule endoscopy.

4. Discussion

This study reports the first evaluation on perception and attitudes toward AI in CE involving a large group of international CERs. This survey revealed that a majority of CERs have confidence in the implementation of AI applications in CE. Our results indicate that most of the responders are interested in undergoing a learning program on AI. Moreover, 92% of CERs agree that AI should be part of the regular endoscopy training. These perceptions are per the survey of Wadhwa et al. [10], in which 86% of physicians reported interest in AI-assisted colonoscopy. Further illustrations, especially in radiology, revealed the same general enthusiasm regarding AI. For example, Waymel et al. [11] showed through an electronic survey sent to radiologists that 79.3% ($n = 214$) of respondents think AI will positively impact their future practice. Moreover, in the study of Pinto Dos Santos et al., 77% and 86% of radiologists and students, respectively, agreed that

AI would revolutionize and improve radiology [12]. These data corroborate the important expectation of AI implementation in medicine.

One of the primary functions of AI in CE is to reduce reading times while maintaining the highest possible sensitivity for abnormalities detection. The task of reviewing full-length SBCE reading is tedious, and CERs spend on average 30 to 120 min to review and interpret complete SBCE examinations encompassing thousands of images [13]. Moreover, Beg et al. showed that reader accuracy declines after reviewing one full-length SBCE recording and that accuracy is also influenced by reading speed [14]. In 2019, a large multicenter study in China showed that an AI-based algorithm could reduce the mean CE reading time to 6 min per video while maintaining a high sensitivity for abnormalities detection [4]. Although the ESGE guidelines recommend that quick CE reading modes “may be used to scan the small bowel for diffuse lesions (. . .) but should not be relied on to detect an isolated lesion” [15], our results interestingly report that 71% of CER ($n = 236$) already use quick CE reading modes in their daily practice, demonstrating a genuine desire to reduce CE reviewing time. Moreover, some could hypothesize that AI implementation would lead to an increase in CE requests, but 53% of CR ($n = 176$) suggest that no significant change is to be expected.

Regarding the potential cost of AI implementation in CE, 65% ($n = 218$) of CER agreed or strongly agreed that it is a topic of major importance. Most CE manufacturers are currently working on developing AI-based solutions for CE reporting. However, the medical literature remains scarce on the potential performances of these devices, as most available studies are retrospective, using local datasets without any external validation. These drawbacks may hamper the cautious implementation of this novel technology. Thus, evaluating the cost-effectiveness of AI solutions in CE is needed to better determine the feasibility and safety of this technology in daily practice and whether it should be implemented as a first or second reader [16]. The specific question of a cost-effectiveness threshold was not addressed in this study and should be considered as a perspective for further medico-economic studies.

Our study has several strengths. First, it is the largest survey addressing gastroenterologists' perceptions of AI toward CE. Second, this survey included various modalities of medical practice (university/teaching hospital, non-university, and private practice). Third, all participants were chosen to represent a diversity of training backgrounds. Nevertheless, we must also acknowledge a few limitations. First, the very real possibility of participation bias: CERs who voluntarily answered this survey might be interested in this topic, especially academic physicians (63%) who were a priori more prone to answering. The authors hypothesize that this higher participation could probably be due to network impact and can be responsible for participation bias. However, neither age nor professional level influenced the perceptions of AI toward CE, meaning that CERs share the same general optimism regarding AI implementation in CE. Second, our online survey did not lead to a debate, and a physical meeting would likely have enriched the discussion rather than an online survey. Third, the term “AI” was given as its general sense of “smart machines capable of performing tasks that typically require human intelligence”. Then, potential benefits of AI in CE were listed (“shorten reading time”, “help characterizing”, “help standardizing”, “automated reporting” . . .) and given to the participants in their general senses (rather than “reading in less than 10 min” or “detecting with sensitivity over 95%”, etc.). We decided so, first because standards are not defined here yet (although some research groups are working on it), second because any precise threshold may influence results where we aimed to measure qualitative outcomes (sentiments, perceptions, expectations) in the community of CER. A review on AI in SBCE from Dray et al. [9] has listed all technological bricks under development and evaluation. Finally, the authors could not calculate a participation rate, as invitations to answer this survey were spread by e-mail throughout many of countries by corresponding capsule leaders in their community of capsule readers; this could have participated in a non-response bias.

Some initiatives have focused on research questions trying to identify priorities related to AI. There is a general will for collaborative efforts to guide better and approach the barriers of AI implementation. In their paper, Vinsard et al. [17] have proposed some fundamental principles for AI system development and clinical testing to promote quality assurance of CAD and diagnosis in colonoscopy. Another example from the study of Ahmad et al. illustrates this trend as they also proposed a systematic process through a modified Delphi methodology to identify research priorities for AI in colonoscopy [18]. All these considerations regarding AI implementation have led to a current ongoing work from our group to map out research, clinical gaps, and challenges for AI in CE with some CE international leaders.

Some ethical issues also must be discussed regarding the legal aspects of AI implementation for CE reading. Depending on how AI is used, the delegation of diagnostic tasks to machines is tricky and unanswered. AI tools have demonstrated their capacity to enhance physicians' performance detecting colonic polyps or Barrett's esophagus, but AI use introduces the risk of influencing physicians' decisions. Our result illustrates that 60% of CR ($n = 201$) at least moderately agreed that AI implementation could lead to operator dependence. One thing is certain; the gastroenterologist community must contribute to the AI development for better application in clinical practice. General optimism is needed for attracting funding in AI algorithms development and for their wide adoption. Indeed, AI in CE will probably be an "adversarial" rather than "assistive" tool meaning that, as for radiologists, only selected frames will be proofread by the CER, while others will simply not be double-checked.

Moreover, human trust in AI is of tremendous importance to support its implementation. Still, AI continues to frighten a part of our society, as AI algorithms can have black boxes even to their creators. AI algorithms make decisions and predictions as humans do without being able to describe their way of operating. More transparency on these black boxes will widen its adoption. However, the community of CERs must be prepared, as AI use in CE will automatically lead to inevitable mistakes. In the new era of AI, these risks could be compared to drug accidents, and it should not slow down AI development and obscure all the potential of this new technology. Indeed, some authors already promote creating a reliable system of AI techno vigilance to solve all these issues [19] and maybe define strict liabilities for harm inflicted by AI.

5. Conclusions

Capsule readers showed general enthusiasm toward AI in CE. The majority of CR believe in the high potential of AI for shortening reading times and for the semi-automated detection of abnormalities. Although CR cannot yet rely entirely on such solutions, before high-quality trials demonstrate that this approach is highly sensitive and cost-effective, the community seems eager to welcome the technology.

Author Contributions: Conceptualization: R.L. and X.D.; methodology: R.L., X.D. and A.K.; software: A.H.; data curation, I.F.-U.S., E.R. and E.T.; investigation: R.L. and X.D.; data curation: E.R., C.V.d.B., P.B., B.J.R. and K.T.; writing: R.L.; original draft preparation: R.L., X.D. and A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: International Capsule endoscopy Research) group members are listed: Eliakim Abraham, Gunnar Baatrup, Peter Baltes, Marco Banic, Hanneke Beaumont, Mauro Bruno, Sergio Cadoni, Stefania Chetcuti Zammit, Pablo Cortegoso Valdivia, Thomas de Lange, Luca Elli, Pierre Ellul, Martin Keuchel, Uri Kopylov, Phey Shen Lee, Reuma Margalit-Yehuda, Wojciech Marlicz, Antoine Martin, Chris Mulder, Alessandro Mussetto, Deirdre McNamara, Lucian Negreanu, Ar-tur Nemeth, Peter Pak, Enrique Pérez-Cuadrado-Robles, Guillaume Perrod, Maria Elena Riccioni, Alexander Robertson, Taina Sipponen, Cristiano Spada, Ilia Tacheci, Annalisa Tortora, Gabriele Wurm Johansson. The I-CARE Group also includes: Romain Leenhardt, Ignacio Fernandez-Urien Sainz, Emanuele Rondonotti, Ervin Toth, Cedric Van de Brauene, Peter Baltes, Bruno Joel Rosa, Konstantinos

Triantafyllou, Aymeric Histace, Anastasios Koulaouzidis and Xavier Dray. Claire Amaris Hobson for her contribution in data analysis.

Conflicts of Interest: Xavier Dray and Romain Leenhardt are cofounders and shareholders of Augmented Endoscopy. Xavier Dray has acted as a consultant for Boston Scientific and Norgine and has given lectures for Fujifilm, Medtronic, MSD and Pentax. Romain Leenhardt has given a lecture for Abbvie.

Appendix A

List of the 32 questions asked to capsule readers in the questionnaire.

1. Are you a capsule endoscopy reader?
2. Your age?
3. Your country (where you practice)?
4. Your professional level?
5. What is your main practice setting?
6. Years of CE reading experience?
7. Current average number of CE read by yourself per year.
8. Are you already using any AI applications (e.g., speech recognition, spam filters, recommendation algorithms, etc.) in your everyday life?
9. Do you use any endoscopy AI solution in your regular practice already?
10. AI will revolutionize endoscopy.
11. How interested are you in receiving a generic/baseline training on AI?
12. How interested are you in receiving technically advanced training on AI?
13. How will the relationship between the endoscopist and the patient change with the introduction of AI?
14. AI should be part of endoscopy training.
15. Are you already using quick CE reading modes?
16. What will be the role(s) of endoscopists in developing/validating AI applications into CE practice?
17. What MEAN reading time should AI-based applications provide before implementation in small bowel CE practice is considered? (Minutes)
18. AI will have a positive impact on CE.
19. AI will increase the diagnostic yield in CE.
20. AI will allow automated reporting of CE without validation by the gastroenterologist.
21. AI will shorten the reading time in CE.
22. AI will help characterizing lesions seen in CE.
23. AI will help standardizing reporting in CE.
24. Patients should be aware about the use of AI in CE reading.
25. Patients should specifically consent to the use of AI in CE reading?
26. Cost is important in implementing AI in CE reading.
27. A normal small bowel CE encompasses a mean number of 12,500 frames. What maximal number of AI false positive frames would you accept to review?
28. What should be a reasonable time limit to get the results from an AI-based CE reading?
29. AI in CE will lead to operator dependence on the technology.
30. AI will increase the number of requests for CE examinations.
31. In the foreseeable future, all capsule readers will be replaced by AI.
32. Who should take the responsibility of AI-based reporting in CE?

References

1. Rajkomar, A.; Dean, J.; Kohane, I. Machine Learning in Medicine. *N. Engl. J. Med.* **2019**, *380*, 1347–1358. [[CrossRef](#)] [[PubMed](#)]
2. Schmitz, R.; Werner, R.; Repici, A.; Bisschops, R.; Meining, A.; Zornow, M.; Messmann, H.; Hassan, C.; Sharma, P.; Rösch, T. Artificial intelligence in GI endoscopy: Stumbling blocks, gold standards and the role of endoscopy societies. *Gut* **2021**. [[CrossRef](#)] [[PubMed](#)]

-
3. Wang, P.; Berzin, T.M.; Glissen Brown, J.R.; Bharadwaj, S.; Becq, A.; Xiao, X.; Liu, P.; Li, L.; Song, Y.; Zhang, D.; et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: A prospective randomised controlled study. *Gut* **2019**, *68*, 1813–1819. [[CrossRef](#)] [[PubMed](#)]
 4. Ding, Z.; Shi, H.; Zhang, H.; Meng, L.; Fan, M.; Han, C.; Zhang, K.; Ming, F.; Xie, X.; Liu, H.; et al. Gastroenterologist- Level Identification of Small-Bowel Diseases and Normal Variants by Capsule Endoscopy Using a Deep-Learning Model. *Gastroenterology* **2019**, *157*, 1044–1054.e5. [[CrossRef](#)] [[PubMed](#)]
 5. Repici, A.; Spadaccini, M.; Antonelli, G.; Correale, L.; Maselli, R.; Galtieri, P.A.; Pellegatta, G.; Capogreco, A.; Milluzzo, S.M.; Lollo, G.; et al. Artificial intelligence and colonoscopy experience: Lessons from two randomised trials. *Gut* **2021**. [[CrossRef](#)] [[PubMed](#)]
 6. Bisschops, R.; East, J.E.; Hassan, C.; Hazewinkel, Y.; Kamiński, M.F.; Neumann, H.; Pellisé, M.; Antonelli, G.; Balen, M.B.; Coron, E.; 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. [[PubMed](#)]
 7. Leenhardt, R.; Vasseur, P.; Philippe, M.; Saurin, J.C.; Rahmi, G.; Cholet, F.; Becq, A.; Marteau, P.; Histace, A.; Dray, X.; et al. A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. *Gastrointest. Endosc.* **2019**, *89*, 189–194. [[CrossRef](#)] [[PubMed](#)]
 8. Aoki, T.; Yamada, A.; Aoyama, K.; Saito, H.; Tsuboi, A.; Nakada, A.; Niikura, R.; Fujishiro, M.; Oka, S.; Ishihara, S.; et al. Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network. *Gastrointest. Endosc.* **2019**, *89*, 357–363.e2. [[CrossRef](#)] [[PubMed](#)]
 9. Dray, X.; Iakovidis, D.; Houdeville, C.; Jover, R.; Diamantis, D.; Histace, A.; Koulaouzidis, A. Artificial intelligence in small bowel capsule endoscopy—Current status, challenges and future promise. *J. Gastroenterol. Hepatol.* **2021**, *36*, 12–19. [[CrossRef](#)] [[PubMed](#)]
 10. Wadhwa, V.; Alagappan, M.; Gonzalez, A.; Gupta, K.; Brown, J.R.G.; Cohen, J.; Sawhney, M.; Pleskow, D.; Berzin, T.M. Physician sentiment toward artificial intelligence (AI) in colonoscopic practice: A survey of US gastroenterologists. *Endosc. Int. Open* **2020**, *8*, E1379–E1384. [[CrossRef](#)] [[PubMed](#)]
 11. Waymel, Q.; Badr, S.; Demondion, X.; Cotten, A.; Jacques, T. Impact of the rise of artificial intelligence in radiology: What do radiologists think? *Diagn. Interv. Imaging* **2019**, *100*, 327–336. [[CrossRef](#)] [[PubMed](#)]
 12. Pinto Dos Santos, D.; Giese, D.; Brodehl, S.; Chon, S.H.; Staab, W.; Kleinert, R.; Maintz, D.; Baeßler, B. Medical students' attitude towards artificial intelligence: A multicentre survey. *Eur. Radiol.* **2019**, *29*, 1640–1646. [[CrossRef](#)] [[PubMed](#)]
 13. ASGE Technology Committee; Wang, A.; Banerjee, S.; Barth, B.A.; Bhat, Y.M.; Chauhan, S.; Gottlieb, K.T.; Konda, V.; Maple, J.T.; Murad, F. Wireless capsule endoscopy. *Gastrointest. Endosc.* **2013**, *78*, 805–815. [[CrossRef](#)] [[PubMed](#)]
 14. Beg, S.; Card, T.; Sidhu, R.; Wronska, E.; Rangunath, K.; Ching, H.-L.; Koulaouzidis, A.; Yung, D.; Panter, S.; Mcalindon, M.; et al. The impact of reader fatigue on the accuracy of capsule endoscopy interpretation. *Dig. Liver Dis.* **2021**, *53*, 1028–1033. [[CrossRef](#)] [[PubMed](#)]
 15. Rondonotti, E.; Spada, C.; Adler, S.; May, A.; Despott, E.J.; Koulaouzidis, A.; Panter, S.; Domagk, D.; Fernandez-Urien, I.; Rahmi, G.; et al. Small-bowel capsule endoscopy and device-assisted enteroscopy for diagnosis and treatment of small-bowel disorders: European Society of Gastrointestinal Endoscopy (ESGE) Technical Review. *Endoscopy* **2018**, *50*, 423–446. [[CrossRef](#)]
 16. Piccirelli, S.; Milluzzo, S.M.; Bizzotto, A.; Cesaro, P.; Pecere, S.; Spada, C. Small Bowel Capsule Endoscopy and artificial intelligence: First or second reader? *Best Pract. Res. Clin. Gastroenterol.* **2021**, *52–53*, 101742. [[CrossRef](#)] [[PubMed](#)]
 17. Vinsard, D.G.; Mori, Y.; Misawa, M.; Kudo, S.-E.; Rastogi, A.; Bagci, U.; Rex, D.K.; Wallace, M.B. Quality assurance of computer-aided detection and diagnosis in colonoscopy. *Gastrointest. Endosc.* **2019**, *90*, 55–63. [[CrossRef](#)] [[PubMed](#)]
 18. Ahmad, O.F.; Mori, Y.; Misawa, M.; Kudo, S.-E.; Anderson, J.T.; Bernal, J.; Berzin, T.M.; Bisschops, R.; Byrne, M.F.; Chen, P.-J.; et al. Establishing key research questions for the implementation of artificial intelligence in colonoscopy: A modified Delphi method. *Endoscopy* **2021**, *53*, 893–901. [[CrossRef](#)]
 19. Cabitza, F.; Zeitoun, J.-D. The proof of the pudding: In praise of a culture of real-world validation for medical artificial intelligence. *Ann. Transl. Med.* **2019**, *7*, 161. [[CrossRef](#)]

7.2. L'intelligence artificielle en vidéocapsule : quels objectifs ?

Il apparaît nécessaire que les axes de développement autour de l'IA soient encadrés par une réflexion préalable pluridisciplinaire de façon à favoriser l'adoption et l'utilisation de ces nouveaux outils. Une réflexion en amont des développements paraît essentielle pour répondre à une problématique clinique réelle et envisager une meilleure acceptation de ces outils. Selon la dernière prise de position de l'ESGE (36), l'implémentation de l'IA en VCE s'articule autour de trois axes de recherche : l'aide à l'évaluation de la qualité de la préparation de l'IG, l'évaluation de la complétude de l'examen, et la détection d'anomalies associée à une réduction du temps de lecture. Afin d'identifier au mieux les points clés de recherche et de développement autour de l'implémentation de l'IA en VCE, nous avons mené un sondage international regroupant 21 experts lecteurs de VCE et spécialistes en analyse d'image. Cette étude offre une feuille de route guidant la recherche et le développement de l'IA pour sa bonne intégration et utilisation en VCE (37).

Key research questions for implementation of artificial intelligence in capsule endoscopy

Romain Leenhardt¹, Anastasios Koulaouzidis, Aymeric Histace, Gunnar Baatrup, Sabina Beg, Arnaud Bourreille, Thomas de Lange, Rami Eliakim, Dimitris Iakovidis, Michael Dam Jensen, Martin Keuchel, Reuma Margalit Yehuda, Deirdre McNamara, Miguel Mascarenhas, Cristiano Spada, Santi Segui, Pia Smedsrud, Ervin Toth, Gian Eugenio Tontini, Eyal Klang, Xavier Dray and Uri Kopylov, On behalf on the International Capsule endoscopy REsearch (I CARE) group

Abstract

Background: Artificial intelligence (AI) is rapidly infiltrating multiple areas in medicine, with gastrointestinal endoscopy paving the way in both research and clinical applications. Multiple challenges associated with the incorporation of AI in endoscopy are being addressed in recent consensus documents.

Objectives: In the current paper, we aimed to map future challenges and areas of research for the incorporation of AI in capsule endoscopy (CE) practice.

Design: Modified three-round Delphi consensus online survey.

Methods: The study design was based on a modified three-round Delphi consensus online survey distributed to a group of CE and AI experts. Round one aimed to map out key research statements and challenges for the implementation of AI in CE. All queries addressing the same questions were merged into a single issue. The second round aimed to rank all generated questions during round one and to identify the top-ranked statements with the highest total score. Finally, the third round aimed to redistribute and rescore the top-ranked statements.

Results: Twenty-one (16 gastroenterologists and 5 data scientists) experts participated in the survey. In the first round, 48 statements divided into seven themes were generated. After scoring all statements and rescored the top 12, the question of AI use for identification and grading of small bowel pathologies was scored the highest (mean score 9.15), correlation of AI and human expert reading-second (9.05), and real-life feasibility-third (9.0).

Conclusion: In summary, our current study points out a roadmap for future challenges and research areas on our way to fully incorporating AI in CE reading.

Keywords: capsule endoscopy, artificial intelligence, research

Received: 22 April 2022; revised manuscript accepted: 27 September 2022.

Introduction

Artificial intelligence (AI) technologies are being rapidly developed and implemented in multiple areas of medicine. In recent years, gastrointestinal (GI) endoscopy led the way, with several AI-based technologies for colonic polyp detection introduced and evaluated in clinical practice.^{1–8} Some of the systems are already being used routinely

across the world, even though real-world implementation results are still lacking.

AI is being implemented in several other areas of GI endoscopies, such as early detection of gastric neoplasia,^{9,10} Barrett's esophagus,¹¹ endoscopic ultrasound,^{12,13} and grading of mucosal inflammation in ulcerative colitis.^{14–18} An additional

Ther Adv Gastroenterol

2022, Vol. 15: 1–8

DOI: 10.1177/
17562848221132683

© The Author(s), 2022.

Article reuse guidelines:
sagepub.com/journals-
permissions

Correspondence to:
Romain Leenhardt
Centre of Digestive
Endoscopy, Sorbonne
Université, Hôpital
Saint-Antoine, 184 rue du
Faubourg Saint Antoine,
AP-HP, Paris 75012,
France

ETIS UMR 8051 (CY Paris
Cergy University, ENSEA,
CNRS), Cergy, France
**romain.leenhardt@
aphp.fr**

Xavier Dray
Sorbonne Université,
Centre of Digestive
Endoscopy, Hôpital Saint-
Antoine, AP-HP, Paris,
France

ETIS UMR 8051 (CY Paris
Cergy University, ENSEA,
CNRS), Cergy, France

Anastasios Koulaouzidis
Department of Social
Medicine and Public
Health, Pomeranian
Medical University,
Szczecin, Poland

Department of Surgery,
Odense University
Hospital, Odense,
Denmark

Department of Clinical
research, University
of Southern Denmark,
Odense, Denmark

Aymeric Histace
ETIS UMR 8051 (CY Paris
Cergy University, ENSEA,
CNRS), Cergy, France

Gunnar Baatrup
Department of Surgery,
Odense University
Hospital, Odense,
Denmark

Department of Clinical
research, University
of Southern Denmark,
Odense, Denmark

Sabina Beg
Department of
Gastroenterology, Imperial
College NHS Healthcare
Trust, London, UK

Arnaud Bourreille
Nantes Université, CHU
Nantes, Institut des
maladies de l'appareil
digestif (IMAD), Hépatogastro-
entérologie, Nantes,
France

Thomas de Lange

Department of Medicine and emergencies-
Mölnådal, Region Västra Götaland, Sahlgrenska University Hospital, Gothenburg, Sweden

Department of Molecular and Clinical and Medicine, University of Gothenburg, Sahlgrenska Academy, Gothenburg, Sweden

Rami Eliakim**Reuma Margalit Yehuda****Uri Kopylov**

Department of Gastroenterology, Sheba Medical Center and Sackler School of Medicine, Tel Aviv University, Tel-Aviv, Israel

Dimitris Iakovidis

Department of Computer Science and Biomedical Informatics, University of Thessaly, Lamia, Greece

Michael Dam Jensen

Department of Internal Medicine, Section of Gastroenterology, Lillebaelt Hospital, Vejle, Denmark

Martin Keuchel

Clinic for Internal Medicine, Agaplesion Bethesda Krankenhaus Bergedorf, Hamburg, Germany

Deirdre McNamara

Trinity Academic Gastroenterology Group, Department of Clinical Medicine, Tallaght Hospital, Trinity College Dublin, Dublin, Ireland

Miguel Mascarenhas

Department of Gastroenterology, Centro Hospitalar São João, Porto, Portugal

Cristiano Spada

Digestive Endoscopy Unit and Gastroenterology, Fondazione Poliambulanza, Brescia, Italy

Digestive Endoscopy Unit, Università Cattolica del Sacro Cuore, Rome, Italy

Santi Seguí

Department of Mathematics and Computer Science, Universitat de Barcelona, Barcelona, Spain

Pia Smedsrud

Simula Metropolitan Centre for Digital Engineering, University of Oslo, Aугere Medical AS, Oslo, Norway

Ervin Toth

Department of Gastroenterology, Skåne University Hospital, Lund University, Malmö, Sweden

field with the fast development of AI research is capsule endoscopy (CE), with several publications evaluating deep learning for automated detection of inflammatory lesions,^{19–26} vascular lesions,^{27,28} protruding and neoplastic lesions/ masses,²⁹ and scoring of bowel cleanliness.³⁰

However, there are still multiple challenges in the way of implementation of the impressive experimental performance of AI in CE in clinical practice. Some of these challenges include standardization of the results, validation of established end points, creation of common datasets and computational methodology, and correlation with clinical outcomes. These challenges are partially common to other areas of GI endoscopy and medicine in general³¹ and are being addressed by expert consensus meetings and workshops providing roadmaps into future research targets and methodologies. Recently, such priority-setting statement was published for colonoscopy⁵; the key themes were identified as the establishment of clinical trial design/end points, technological development, clinical integration, data access and annotation and regulatory approval.

The main aim of our study was to identify the top research priorities related to the implementation and further research for AI in CE.

Materials and methods

Study design

The study design was based on a modified three-round Delphi consensus online survey. The modified Delphi methodology is well known and used in medical literature^{5,32} for establishing research priorities based on expert's opinions. All rounds were distributed through GoogleForm® (Mountain View, CA, USA) and the study was conducted between September 2021 and January 2022. The first round consisted of an open and anonymized 10 queries survey proposed to a panel of CE experts. The suggested queries based on expert's opinions aimed to identify and map out key research statements and challenges for the implementation of AI in CE. All answers collected during round one were reviewed by a core group (CG) and categorized into seven singular themes. Queries addressing the same questions were merged into a single issue by the CG. The second round aimed to rank the questions generated during round 1 and identify the highest

scoring statements. The third round aimed to redistribute and rescore the top-ranked statements. All respondents were asked to rate all statements on a numerical wider scale from 1 (*very low priority*) to 10 (*very high priority*).

The core group and expert group

The CG was composed of translational CE readers and data scientists (UK, RL, AK, XD, and AH) to form a key leader opinion to conduct this study. All questionnaires were sent to a panel of CE experts, including the CG, with a diversity of backgrounds including physicians CE experts' readers and data scientists related to the CE medical field.

Results

Among the 22 experts invited to participate in this study, 21 finally answered at least one of the questionnaires. The participation rate was 90% ($n = 19$) for the first round and 95% ($n = 20$) for the second and third rounds, respectively. The 21 respondents were considered as the expert group (EG) and included physicians CE experts readers (76%, $n = 16$) and data scientists (24%, $n = 5$). Members of the EG were based in Denmark ($n = 3$), England ($n = 1$), France ($n = 4$), Germany ($n = 1$), Greece ($n = 1$), Ireland ($n = 1$), Israel ($n = 3$), Italy ($n = 2$), Norway ($n = 1$), Portugal ($n = 1$), Spain ($n = 1$), and Sweden ($n = 2$). The mean age of the experts was 49 years. The main practice setting was academic ($n = 18$; 86%) and mixed academic/private ($n = 3$; 14%). The physicians CE experts had a mean CE reading experience of 14 years [interquartile range (IQR) = 13] and a mean number of CE read annually of 154 (IQR = 150). The data scientists had a mean CE experience of 12 years.

After the review process by the CG, the first round generated 48 statements divided into seven themes (Table 1). These statements were then considered for scoring in the second round. In round two, considering the wider scale from 1 (*very low priority*) to 10 (*very high priority*), the mean score obtained for the 48 statements ranged from 4.6 to 9.2 (Table 1). Then the top 12 statements, including tied scores, were identified from three themes including performances metrics, AI in CE in clinical practice, and auditing AI systems. The final 12 statements were considered for rescoring in the third round. Results of the third round showed a mean score ranging from 7.63 to 9.15 (Table 2).

Discussion

The current study is the first attempt to prioritize and standardize the research challenges and questions in the application of AI in CE. The consensus was facilitated by a modified three-stage Delphi process through an established group of CE experts and data engineers with extensive experience on the subject.

AI is rapidly being incorporated into multiple fields in medicine, with GI endoscopy being among the leading disciplines. In colonoscopy, AI-based systems for the facilitation of polyp detection are already commercially available and have been proven successful in improving polyp detection rate by up to 40%.^{6,7} In CE, comparative research is not yet available; however, detection of multiple types of small bowel (SB) and

colonic pathologies is accurate and feasible.^{33–35} CE is perhaps an even more attractive target for AI research, as there is no need for real-time diagnosis, and the variety of pathology types is somewhat limited: thus, identification of most SB pathologies by AI was very accurate, with AUC above 90%.^{22,30,36–38} In future models of CE, strong incorporation of AI modules with automated lesion markup can be expected. However, the incorporation of AI into clinical practice and clinical trials requires a huge leap in terms of standardization, quality assessment, reproducibility, and workflow integration. In a recent large European survey encompassing 380 gastroenterologists (of them 88% experienced capsule readers), a majority of the responders agreed that AI would positively impact CE, shorten CE reading time, help standardize reporting in CE, and characterize lesions seen in CE; however, the likelihood of complete replacement of human readers by AI was deemed to be low.³⁹

We aimed to map and prioritize the main challenges for further research and integration into clinical practice. Our EG was comprised of 76% physicians and 24% data scientists, all of them with a vast track record of CE reading and research. It appears that the highest scores issued in this EG were still those referring to accuracy in detection of findings in both SB and colon, as well as the optimal threshold for accuracy of the algorithm. The next group of statements addressed the feasibility and accuracy in a real-world setting. Indeed, to date, no real-world model for the utilization of AI in CE has been published. Identification of specific lesions by AI

Table 1. Results of the first round after reviewing by the core group.

Themes	Number of statements
Performance metrics	16
Technological developments	9
AI in CE in clinical practice	8
Data (access, sharing/privacy, curation)	6
Ethical, legal and ecological issues	5
Auditing AI systems	3
Education and training	1

AI, artificial intelligence; CE, capsule endoscopy.

may be introduced into capsule reading software of any of capsule producers/vendors; nonetheless, clinical decision-making or predictive models based on AI capsule reading (either complete videos or still images) are still missing

Most of the available studies included images obtained with a specific capsule model or brand. Widespread utilization of AI will require brand/model-spanning algorithms that are still very rare.²⁸ Similarly, a clinical algorithm would be required to detect multiple types of pathologies at the same time and on the same still image, regardless of the location of the image (SB/colon). An additional issue of concern is whether AI would be able to completely replace a human reader, and what degree of human supervision/auditing will be required. This challenge is closer to those originating from the worlds of imaging and pathology; AI in colonoscopy may augment human judgment but will not replace it completely as the human is still behind the scope and is instrumental to the obtainment of quality images. CE diagnosis is completely in post-acquisition, and human intervention is not required for anything but the reading itself. Nevertheless, CE reading is a tedious and lengthy task, especially for colonic capsules; AI can shorten the reading time by at least 95%; however, real-life accuracy data that could support this model in clinical practice are completely lacking. This issue could be critical for the uptake of colonic CE that is currently hampered by long reading time and subsequent devaluation of the economic model for this potentially appealing screening modality. An additional temporary compromise could be to

Gian Eugenio Tontini
Department of Pathophysiology and Transplantation, University of Milan and Gastroenterology and Endoscopy Unit, Fondazione IRCCS Ca'Granda Ospedale Maggiore Policlinico, Milan, Italy

Eyal Klang
Sheba ARC, Sheba Medical Center and Sackler School of Medicine, Tel Aviv University, Tel-Aviv, Israel

Table 2. Results of the top 12 ranked statements third round with the top 12 ranked statements.

Rank	Statements	Total score	Mean score
1	AI for automatic detection, identification, characterization (type, size, severity), and differentiation of SB lesions: inflammatory lesions (erosion, ulceration, edema, etc.), vascular lesions, bulges, atrophy.	174	9.15
2	How do the overall results of AI and expert reading correlate (for all lesions/relevant lesions)?	172	9.05
3	Feasibility and accuracy in the real-world setting.	171	9.00
4	AI for automatic detection/identification, characterization (type, size, severity), and differentiation of colon lesions: ulcers, vascular lesions, polyps.	157	8.26
5	Auditing of CE systems after incorporation in clinical practice.	155	8.15
6	How to reduce the false-positive rate without decreasing sensitivity?	155	8.15
7	Creation of algorithms for simultaneous identification of multiple lesion types.	151	7.94
8	Adoption of AI by clinicians.	150	7.89
9	What are the optimal clinical end-points for the evaluation of AI software?	149	7.84
10	What are the optimal clinical trial design and end-points to compare different AI systems for CE?	149	7.84
11	How accurate should AI be to be incorporated in clinical trials?	147	7.73
12	What accuracy parameters are potential targets for AI utilization?	145	7.63

AI, artificial intelligence; CE, capsule endoscopy; SB, small bowel.

utilize AI to remove normal images, in similarity to the current features of ‘Quickview’ or top-100; currently, these features are insufficiently accurate^{40,41}; however, future hardware and software improvements may change that.

It is likely that in the near future we will witness several AI systems for CE, some incorporated in the reading software and others as standalone suites.⁴² There is a need to compare the accuracy of these and other forthcoming systems and a requirement for benchmarking parameters for both clinical trial and real-world use.

In inflammatory bowel diseases, CE has a potential major advantage of being able to access the panenteric inflammatory burden. The importance of mucosal healing as a therapeutic target in inflammatory bowel diseases has been well

described.^{43,44} However, the concept of mucosal healing in IBD is almost constantly addressing the colon and the terminal ileum. Nonetheless, mucosal responsiveness of different gut segments to medical treatment is not identical.^{45,46} In addition, in Crohn’s disease patients in clinical remission, some residual SB inflammation is very common⁴⁷ and has major clinical implications on the likelihood of long-term remission.⁴⁸ Surprisingly, only a handful of studies to date utilized CE for evaluation of mucosal healing prospectively.^{49,50} In the last few years, several studies reported the use of colonic capsules for panenteric evaluation^{51–54}; recently a specialized Crohn’s disease capsule (Pillcam Crohn, Medtronic, Minneapolis, MN, USA) has been released.^{55–58} AI-read CE could be a potentially safe and accurate modality for the assessment of mucosal inflammation in clinical trials in IBD,

pending further benchmarking and standardization. To date, no studies evaluating complete film AI-augmented reading were published. This challenge may require a very different analytical approach.

Our study has several limitations. Primarily, this was a Delphi survey of a predefined group of AI and CE experts. The group was limited in size however included participants with a significant track record in the field. Attitudes toward AI in CE in a larger and more representative group of gastroenterologists were previously evaluated by our group.³⁹ In addition, our objective was to raise and solicit research questions; the suggestions of the participating experts merit further research efforts in the years to come.

Conclusion

In summary, our current study points out a roadmap for future challenges and research areas on our way to fully incorporate AI in CE reading. These statements are useful not only for research but also for AI medical education.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Author contribution(s)

Romain Leenhardt: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Writing – original draft; Writing – review & editing.

Anastasios Koulaouzidis: Data curation; Formal analysis; Investigation; Methodology; Supervision; Validation; Writing – review & editing.

Aymeric Histace: Data curation; Formal analysis; Investigation; Methodology; Supervision; Validation; Writing – review & editing.

Gunnar Baatrup: Validation; Writing – review & editing.

Sabina Beg: Validation; Writing – review & editing.

Arnaud Bourreille: Validation; Writing – review & editing.

Thomas de Lange: Validation; Writing – review & editing.

Rami Eliakim: Validation; Writing – review & editing.

Dimitris Iakovidis: Validation; Writing – review & editing.

Michael Dam Jensen: Validation; Writing – review & editing.

Martin Keuchel: Validation; Writing – review & editing.

Reuma Margalit Yehuda: Validation; Writing – review & editing.

Deirdre McNamara: Validation; Writing – review & editing.

Miguel Mascarenhas: Validation; Writing – review & editing.

Cristiano Spada: Validation; Writing – review & editing.

Santi Segui: Validation; Writing – review & editing.

Pia Smedsrud: Validation; Writing – review & editing.

Ervin Toth: Validation; Writing – review & editing.

Gian Eugenio Tontini: Validation; Writing – review & editing.

Eyal Klang: Validation; Writing – review & editing.

Xavier Dray: Formal analysis; Investigation; Methodology; Supervision; Validation; Writing – review & editing.

Uri Kopylov: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Supervision; Validation; Writing – original draft; Writing – review & editing.

Acknowledgements

None.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Competing interests

Uri Kopylov: research support – Medtronic, Janssen, Takeda. Speaker/advisory fees – Abbvie, BMS, Janssen, Pfizer, Takeda, MSD, Rafa. Xavier Dray, Romain Leenhardt, and Aymeric Histace are cofounders and shareholders of Augmented Endoscopy. Xavier Dray has acted as a consultant for Boston Scientific and Norgine and has given lectures for Fujifilm, Medtronic, MSD, and Pentax. Romain Leenhardt has given a lecture for Abbvie. Martin Keuchel has acted as a consultant for Medtronic and has given lectures for Medtronic and received study support from AnXRobotics and Olympus. Arnaud Bourreille received honoraria or grant research from Abbvie, Celgene, Galapagos, Gilead, Mauna Kea Technologies, Medtronic, MSD, Hoffman la Roche, Ferring, Janssen, MSD, Pfizer, Takeda, Tillotts, and OSE immunotherapeutics. PiaSmedsrud employed and shareholder in AugereMedical a start-up company for AI in endoscopy. Medical Augere of shareholder and cofounder is Lange de Thomas. Rami Eliakim has received research support and consultant fee from Medtronic, consultant fee from Janssen and Takeda

Availability of data and materials

Not applicable.

ORCID iDs

Romain Leenhardt  <https://orcid.org/0000-0002-4641-3543>

Ervin Toth  <https://orcid.org/0000-0002-9314-9239>

Uri Kopylov  <https://orcid.org/0000-0002-7156-0588>

References

1. Pannala R, Krishnan K, Melson J, *et al.* Artificial intelligence in gastrointestinal endoscopy. *VideoGIE* 2020; 5: 598–613.
2. Tontini GE, Rimondi A, Vernero M, *et al.* Artificial intelligence in gastrointestinal endoscopy for inflammatory bowel disease: a systematic review and new horizons. *Therap Adv Gastroenterol* 2021; 14: 17562848211017730.
3. Yang YJ. The future of capsule endoscopy: the role of artificial intelligence and other technical advancements. *Clin Endosc* 2020; 53: 387–394.
4. Ahmad OF. Deep learning for colorectal polyp detection: time for clinical implementation? *Lancet Gastroenterol Hepatol* 2020; 5: 330–331.
5. 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: 893–901.
6. Repici A, Badalamenti M, Maselli R, *et al.* Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial. *Gastroenterology* 2020; 159: 512.e7–520.e7.
7. Repici A, Spadaccini M, Antonelli G, *et al.* Artificial intelligence and colonoscopy experience: lessons from two randomised trials. *Gut* 2021; 71: 757–765.
8. Topol EJ. Welcoming new guidelines for AI clinical research. *Nat Med* 2020; 26: 1318–1320.
9. Wu L, Xu M, Jiang X, *et al.* Real-time artificial intelligence for detecting focal lesions and diagnosing neoplasms of the stomach by white-light endoscopy (with videos). *Gastrointest Endosc* 2021; 95: 269–280.
10. Klang E, Barash Y, Levartovsky A, *et al.* Differentiation between malignant and benign endoscopic images of gastric ulcers using deep learning. *Clin Exp Gastroenterol* 2021; 14: 155–162.
11. Struyvenberg MR, de Groof AJ, van der Putten J, *et al.* A computer-assisted algorithm for narrow-band imaging-based tissue characterization in Barrett's esophagus. *Gastrointest Endosc* 2021; 93: 89–98.
12. Goyal H, Mann R, Gandhi Z, *et al.* Application of artificial intelligence in pancreaticobiliary diseases. *Ther Adv Gastrointest Endosc* 2021; 14: 2631774521993059.
13. Kuwahara T, Hara K, Mizuno N, *et al.* Usefulness of deep learning analysis for the diagnosis of malignancy in intraductal papillary mucinous neoplasms of the pancreas. *Clin Transl Gastroenterol* 2019; 10: 1–8.
14. Gottlieb K, Requa J, Karnes W, *et al.* Central reading of ulcerative colitis clinical trial videos using neural networks. *Gastroenterology* 2021; 160: 710.e2–719.e2.
15. Ghoshal UC, Rai S, Kulkarni A, *et al.* Prediction of outcome of treatment of acute severe ulcerative colitis using principal component analysis and artificial intelligence. *JGH Open* 2020; 4: 889–897.

16. Takenaka K, Ohtsuka K, Fujii T, *et al.* Development and validation of a deep neural network for accurate evaluation of endoscopic images from patients with ulcerative colitis. *Gastroenterology* 2020; 158: 2150–2157.
17. Yao H, Najarian K, Gryak J, *et al.* Fully automated endoscopic disease activity assessment in ulcerative colitis. *Gastrointest Endosc* 2021; 93: 728.e1–736.e1.
18. Stidham RW, Liu W, Bishu S, *et al.* Performance of a deep learning model vs human reviewers in grading endoscopic disease severity of patients with ulcerative colitis. *JAMA Netw Open* 2019; 2: e193963.
19. Klang E, Barash Y, Margalit RY, *et al.* Deep learning algorithms for automated detection of Crohn's disease ulcers by video capsule endoscopy. *Gastrointest Endosc* 2020; 91: 606.e2–613.e2.
20. Klang E, Grinman A, Soffer S, *et al.* Automated detection of Crohn's disease intestinal strictures on capsule endoscopy images using deep neural networks. *J Crohns Colitis* 2020; 15: 749–756.
21. Klang E, Kopylov U, Mortensen B, *et al.* A convolutional neural network deep learning model trained on CD ulcers images accurately identifies NSAID ulcers. *Front Med* 2021; 8: 656493.
22. Barash Y, Azaria L, Soffer S, *et al.* Ulcer severity grading in video capsule images of patients with Crohn's disease: an ordinal neural network solution. *Gastrointest Endosc* 2021; 93: 187–192.
23. Aoki T, Yamada A and Koike K. The exceptional performance of deep learning for capsule endoscopy: will such quality be maintained in clinical scenarios? *Gastrointest Endosc* 2021; 93: 365–366.
24. Aoki T, Yamada A, Aoyama K, *et al.* Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network. *Gastrointest Endosc* 2019; 89: 357.e2–363.e2.
25. Majtner T, Brodersen JB, Herp J, *et al.* A deep learning framework for autonomous detection and classification of Crohn's disease lesions in the small bowel and colon with capsule endoscopy. *Endosc Int Open* 2021; 9: E1361–E1370.
26. Mascarenhas Saraiva MJ, Afonso J, Ribeiro T, *et al.* Deep learning and capsule endoscopy: automatic identification and differentiation of small bowel lesions with distinct haemorrhagic potential using a convolutional neural network. *BMJ Open Gastroenterol* 2021; 8: e000753.
27. Leenhardt R, Vasseur P, Li C, *et al.* A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. *Gastrointest Endosc* 2019; 89: 189–194.
28. Houdeville C, Souchaud M, Leenhardt R, *et al.* A multisystem-compatible deep learning-based algorithm for detection and characterization of angiectasias in small-bowel capsule endoscopy. A proof-of-concept study. *Dig Liver Dis* 2021; 53: 1627–1631.
29. Saito H, Aoki T, Aoyama K, *et al.* Automatic detection and classification of protruding lesions in wireless capsule endoscopy images based on a deep convolutional neural network. *Gastrointest Endosc* 2020; 92: 144.e1–151.e1.
30. Leenhardt R, Souchaud M, Houist G, *et al.* A neural network-based algorithm for assessing the cleanliness of small bowel during capsule endoscopy. *Endoscopy* 2021; 53: 932–936.
31. Langlotz CP, Allen B, Erickson BJ, *et al.* A roadmap for foundational research on artificial intelligence in medical imaging: from the 2018 nih/rsna/acr/the academy workshop. *Radiology* 2019; 291: 781–791.
32. Rees CJ, Ngu WS, Regula J, *et al.* European society of gastrointestinal endoscopy – establishing the key unanswered research questions within gastrointestinal endoscopy. *Endoscopy* 2016; 48: 884–891.
33. Sullivan P, Gupta S, Powers PD, *et al.* Artificial intelligence research and development for application in video capsule endoscopy. *Gastrointest Endosc Clin N Am* 2021; 31: 387–397.
34. Tziortziotis I, Laskaratos F-M and Coda S. Role of artificial intelligence in video capsule endoscopy. *Diagnostics (Basel)* 2021; 11: 1192.
35. Ferreira JPS, de Mascarenhas Saraiva MJ da QEC, Afonso JPL, *et al.* Identification of ulcers and erosions by the novel Pillcam™ Crohn's Capsule using a convolutional neural network: a multicentre pilot study. *J Crohns Colitis* 2021; 16:169–172.
36. Aoki T, Yamada A, Aoyama K, *et al.* Clinical usefulness of a deep learning-based system as the first screening on small-bowel capsule endoscopy reading. *Dig Endosc* 2020; 32: 585–591.
37. Dray X, Iakovidis D, Houdeville C, *et al.* Artificial intelligence in small bowel capsule endoscopy—current status, challenges and future promise. *J Gastroenterol Hepatol* 2021; 36: 12–19.
38. de Maissin A, Vallée R, Flamant M, *et al.* Multi-expert annotation of Crohn's disease images of the small bowel for automatic detection using a

- convolutional recurrent attention neural network. *Endosc Int Open* 2021; 9: E1136–E1144.
39. Leenhardt R, Fernandez-Urien Sainz I, Rondonotti E, *et al.* PEACE: perception and expectations toward artificial intelligence in capsule endoscopy. *J Clin Med* 2021; 10: 5708.
 40. Arieira C, Monteiro S, Dias de Castro F, *et al.* Capsule endoscopy: is the software TOP 100 a reliable tool in suspected small bowel bleeding? *Dig Liver Dis* 2019; 51: 1661–1664.
 41. Freitas M, Arieira C, Carvalho PB, *et al.* Simplify to improve in capsule endoscopy—TOP 100 is a swift and reliable evaluation tool for the small bowel inflammatory activity in Crohn's disease. *Scand J Gastroenterol* 2020; 55: 408–413.
 42. Ding Z, Shi H, Zhang H, *et al.* Gastroenterologist-level identification of small-bowel diseases and normal variants by capsule endoscopy using a deep-learning model. *Gastroenterology* 2019; 157: 1044.e5–1054.e5.
 43. Peyrin-Biroulet L, Sandborn W, Sands BE, *et al.* Selecting therapeutic targets in inflammatory bowel disease (STRIDE): determining therapeutic goals for treat-to-target. *Am J Gastroenterol* 2015; 110: 1324–1338.
 44. Turner D, Ricciuto A, Lewis A, *et al.* STRIDE-II: an update on the selecting therapeutic targets in inflammatory bowel disease (STRIDE) initiative of the international organization for the study of IBD (IOIBD): determining therapeutic goals for treat-to-target strategies in IBD. *Gastroenterology* 2021; 160: 1570–1583.
 45. Danese S, Sandborn WJ, Colombel J-F, *et al.* Endoscopic, radiologic, and histologic healing with vedolizumab in patients with active Crohn's disease. *Gastroenterology* 2019; 157: 1007.e7–1018.e7.
 46. Pouillon L, Rousseau H, Busby-Venner H, *et al.* Vedolizumab trough levels and histological healing during maintenance therapy in ulcerative colitis. *J Crohns Colitis* 2019; 13: 970–975.
 47. Kopylov U, Yablecovitch D, Lahat A, *et al.* Detection of small bowel mucosal healing and deep remission in patients with known small bowel Crohn's disease using biomarkers, capsule endoscopy, and imaging. *Am J Gastroenterol* 2015; 110: 1316–1323.
 48. Ben-Horin S, Lahat A, Amitai MM, *et al.* Assessment of small bowel mucosal healing by video capsule endoscopy for the prediction of short-term and long-term risk of Crohn's disease flare: a prospective cohort study. *Lancet Gastroenterol Hepatol* 2019; 4: 519–528.
 49. Hall B, Holleran G, Chin J-L, *et al.* A prospective 52 week mucosal healing assessment of small bowel Crohn's disease as detected by capsule endoscopy. *J Crohns Colitis* 2014; 8: 1601–1609.
 50. Hall BJ, Holleran GE, Smith SM, *et al.* A prospective 12-week mucosal healing assessment of small bowel Crohn's disease as detected by capsule endoscopy. *Eur J Gastroenterol Hepatol* 2014; 26: 1253–1259.
 51. Boal Carvalho P, Rosa B, Dias de Castro F, *et al.* PillCam COLON 2 in Crohn's disease: a new concept of pan-enteric mucosal healing assessment. *World J Gastroenterol* 2015; 21: 7233–7241.
 52. Leighton JA, Helper DJ, Gralnek IM, *et al.* Comparing diagnostic yield of a novel pan-enteric video capsule endoscope with ileocolonoscopy in patients with active Crohn's disease: a feasibility study. *Gastrointest Endosc* 2017; 85: 196.e1–205.e1.
 53. Hall B, Holleran G and McNamara D. PillCam COLON 2(©) as a pan-enteroscopic test in Crohn's disease. *World J Gastrointest Endosc* 2015; 7: 1230–1232.
 54. Oliva S, Aloï M, Viola F, *et al.* A treat to target strategy using panenteric capsule endoscopy in pediatric patients with Crohn's disease. *Clin Gastroenterol Hepatol* 2019; 17: 2060.e1–2067.e1.
 55. Cortegoso Valdivia P, Elosua A, Houdeville C, *et al.* Clinical feasibility of panintestinal (or panenteric) capsule endoscopy: a systematic review. *Eur J Gastroenterol Hepatol* 2021; 33: 949–955.
 56. Lazaridis L-D, Tziatzios G, Toth E, *et al.* Implementation of European society of gastrointestinal endoscopy (ESGE) recommendations for small-bowel capsule endoscopy into clinical practice: results of an official ESGE survey. *Endoscopy* 2021; 53: 970–980.
 57. Tai FWD, Ellul P, Elosua A, *et al.* Panenteric capsule endoscopy identifies proximal small bowel disease guiding upstaging and treatment intensification in Crohn's disease: a European multicentre observational cohort study. *United European Gastroenterol J* 2021; 9: 248–255.
 58. Eliakim R, Yablecovitch D, Lahat A, *et al.* A novel PillCam Crohn's capsule score (Eliakim score) for quantification of mucosal inflammation in Crohn's disease. *United European Gastroenterol J* 2020; 8: 544–551.

7.3. Les spécificités de l'intelligence artificielle appliquée à la vidéocapsule

Au cours d'une endoscopie conventionnelle, le médecin manœuvre l'endoscope en temps réel de façon à examiner la muqueuse du tube digestif et détecter d'éventuelles lésions pathologiques. Le principal défi technologique de l'IA dans ce contexte est d'afficher les prédictions de l'algorithme de manière synchrone à l'examen (IA assistive, complémentaire et non compétitive de l'opérateur humain). L'application de l'IA en VCE est confrontée à un défi presque opposé. La lecture différée de l'examen efface cette contrainte temporelle du traitement des images. En revanche, dans un futur proche, l'IA sélectionnerait un échantillon d'images d'intérêt (environ 1%) et les images restantes pourraient ne pas être analysées par le médecin (environ 99%). Cette démarche s'inscrit dans un contexte « assertif », en compétition de la lecture exhaustive par l'opérateur humain, comme cela se développe par exemple pour écarter de la charge de travail d'un radiologue, des scanners (ou certaines coupes de scanner) jugés normaux par une IA. L'objectif de ce type d'IA est donc d'atteindre une sensibilité très élevée afin de ne pas méconnaître une anomalie tout en réduisant significativement le temps de lecture de l'examen par le médecin.

7.4. Bases de données d'apprentissage

En 2016, une initiative a été lancée pour la création d'une banque nationale de données en VCE. Cette banque de donnée multicentrique nommée CAD CAP (*Computer Assisted Diagnosis for Capsule Endoscopy*) comprenait 4174 enregistrements de VCE issus de 12 centres français. Parmi ces VCE, 1480 comprenaient au moins une anomalie. Ces VCE étaient pour la majorité réalisées dans un contexte de SSIG (67%). Les autres indications moins fréquentes étaient représentées par la suspicion de MC (12%) et le suivi de maladie cœliaque et de polypose de Peutz Jeghers (21%). Au total, 5124 images comprenant au moins une anomalie étaient identifiées. Chaque lésion était délimitée manuellement par une étudiante en médecine supervisée par des experts en VCE. La banque de données était composée de 3103 images de lésions vasculaires pertinentes, 2286 images de lésions vasculaires douteuses et 1370 images de lésion ulcéro-inflammatoires. La banque était également composée de 20 000 images normales servant de témoins pour l'apprentissage machine. Le point fort de cette banque de données repose sur la grande qualité des images qui y sont proposées de par leur précision de délimitation offrant un apprentissage supervisé de grande qualité. Ce travail a fait l'objet d'une publication scientifique (38).

CAD-CAP: a 25,000-image database serving the development of artificial intelligence for capsule endoscopy



Authors

Romain Leenhardt¹, Cynthia Li², Jean-Philippe Le Mouel³, Gabriel Rahmi⁴, Jean Christophe Saurin⁵, Franck Cholet⁶, Arnaud Boureille⁷, Xavier Amiot⁸, Michel Delvaux⁹, Clotilde Duburque¹⁰, Chloé Leandri¹¹, Romain Gérard¹², Stéphane Lecleire¹³, Farida Mesli¹⁴, Isabelle Nion-Larmurier¹, Olivier Romain¹⁵, Sylvie Sacher-Huvelin⁷, Camille Simon-Shane¹⁵, Geoffroy Vanbiervliet¹⁶, Philippe Marteau¹, Aymeric Histace¹⁵, Xavier Dray^{1,15}

Institutions

- 1 Sorbonne University, Endoscopy Unit
- 2 Drexel University, College of Arts & Sciences, Philadelphia, Pennsylvania, United States
- 3 Gastroenterology, Amiens University Hospital, Université de Picardie Jules Verne, Amiens, France
- 4 Georges Pompidou European Hospital, APHP, Department of Gastroenterology and Endoscopy, Paris, France
- 5 Department of Endoscopy and Gastroenterology, Pavillon L, Hôpital Edouard Herriot, Lyon, France
- 6 Digestive Endoscopy Unit, University Hospital, Brest, France
- 7 Department of Hepato-Gastroenterology, Institut des Maladies de l'Appareil Digestif, Nantes, France
- 8 Tenon Hospital, Gastroenterology Department, Paris, France
- 9 CHU Strasbourg, Gastroenterology Department, Strasbourg, France
- 10 Lomme Hospital, Gastroenterology Department, Lomme, France
- 11 Cochin Hospital Gastroenterology Department, Paris, France
- 12 CHRU Lille, Gastroenterology Department, Lille, France
- 13 CHU Rouen, Gastroenterology Department, Rouen, France
- 14 CHU Henri Mondor, Gastroenterology Department, Creteil, France
- 15 ETIS, Université de Cergy-Pontoise, ENSEA, CNRS, Cergy-Pontoise Cedex, France
- 16 CHU Nice, Gastroenterology and Endoscopy Unit, Nice, France

Corresponding author

Pr Xavier Dray, Hopital Saint-Antoine - Endoscopy Unit, 184 Rue du Faubourg Saint-Antoine, Paris 75012, France
xavier.dray@aphp.fr

ABSTRACT

Background and study aims Capsule endoscopy (CE) is the preferred method for small bowel (SB) exploration. With a mean number of 50,000 SB frames per video, SBCE reading is time-consuming and tedious (30 to 60 minutes per video). We describe a large, multicenter database named CAD-CAP (Computer-Assisted Diagnosis for Capsule Endoscopy, CAD-CAP). This database aims to serve the development of CAD tools for CE reading.

Materials and methods Twelve French endoscopy centers were involved. All available third-generation SB-CE videos (Pillcam, Medtronic) were retrospectively selected from these centers and deidentified. Any pathological frame was extracted and included in the database. Manual segmentation of findings within these frames was performed by two pre-med students trained and supervised by an expert reader. All frames were then classified by type and clinical relevance by a panel of three expert readers. An automated extraction process was also developed to create a dataset of normal, proofread, control images from normal, complete, SB-CE videos.

Results Four-thousand-one-hundred-and-seventy-four SB-CE were included. Of them, 1,480 videos (35 %) containing at least one pathological finding were selected. Findings from 5,184 frames (with their short video sequences) were extracted and delimited: 718 frames with fresh blood, 3,097 frames with vascular lesions, and 1,369 frames with inflammatory and ulcerative lesions. Twenty-thousand normal frames were extracted from 206 SB-CE normal videos. CAD-CAP has already been used for development of automated tools for angiectasia detection and also for two international challenges on medical computerized analysis.

submitted 5.4.2019

accepted after revision 16.9.2019

Bibliography

DOI <https://doi.org/10.1055/a-1035-9088> |
Endoscopy International Open 2020; 08: E415-E420
© Georg Thieme Verlag KG Stuttgart · New York
eISSN 2196-9736

Introduction

Capsule Endoscopy (CE) has rapidly become the standard minimally invasive method for visualization of the small bowel (SB). CE is considered the first-line method of investigation of SB diseases [1] and may become the leading diagnostic tool for the entire gastrointestinal tract. With a mean number of 50,000 SB frames per video, SB-CE reading is time-consuming and tedious (30 to 60 minutes per video) [2]. This entails an inherent risk of missed lesions during the reading process by physicians.

Computer-aided diagnosis (CAD) algorithms are increasingly used to assist physicians in reading and interpreting medical images. The field of CAD has rapidly expanded from radiology [3, 4] to many other settings, such as pathology, dermatology, cardiology, and ophthalmology [5 - 7]. In August 2018, the U.S. Food and Drug Administration approved the first medical device to use artificial intelligence (AI) for screening diabetic retinopathy [8]. Development of CAD has also emerged in endoscopy, mostly in the setting of colon polyp detection and characterization [5, 9, 10]. We and others [11 - 13], believe that CAD will meet the challenge of automated lesion detection in SB-CE, thus improving diagnostic accuracy and relieving the burden of SB-CE reading [14].

CAD algorithms dedicated to SB-CE combine elements of AI and computer vision. This development is mostly based on machine learning (ML) algorithms for automated detection of relevant patterns in CE still frames and videos [15, 16]. To be efficient, ML algorithms must be trained on large databases. Any

data (still frame or video) should be labeled, and any finding within frames/videos can be annotated or delimited to better guide and supervise the ML process. Only a limited number of CE images databases is available [13]. Herein, we describe a large, multicenter database named CAD-CAP (computer-assisted diagnosis for capsule endoscopy). This database aims to serve development of CAD tools, from the initial steps (ML) to the last preclinical steps (assessment and comparison of performances).

Materials and methods

CAD-CAP is a national multicenter database approved by the French Data Protection Authority. Twelve French endoscopic units participated to the database (► Table 1 and ►Table 2). A research group specialized in the computerized analysis of medical images was also involved.

Data collection and preparation of locally read SB-CE

In September 2016, all third-generation SB-CE videos (Pillcam SB3 system, Medtronic, Minnesota, United States) registered in the 12 participating endoscopy units were retrospectively collected, and de-identified. Clinical data were noted (age, gender, indication for SB-CE). Any SB icon selected by the local reader was considered as a frame of interest. Each frame of interest was extracted and included in the CAD-CAP database, together with a short adjacent video sequence that included 25 frames upstream and downstream the index frame. Two pre-

► Table 1 Characteristics of the CAD-CAP database.

Characteristic	
Total number of SB3-CE (n)	4174
Total number of SB3-CE (n) videos with at least one abnormal finding	1480
Total number of frames with abnormal findings (n)	5124
Total number of normal frames (n)	20,000
Gender ratio M/F (%)	59/41 %
Mean ± SD age (years)	64 ± 15
CAD-CAP, computer-assisted diagnosis for capsule endoscopy; SB3-CE, third-generation small bowel capsule endoscopy; SD, standard deviation	

► Table 2 CAD-CAP database details and contribution of the centers.

Centers	Number of SB3-CE provided
Brest Hospital	340
Henri Mondor Hospital, Créteil	358
Lomme Hospital	173
Lille Hospital	445
Edouard Herriot Hospital, Lyon	450
Nantes Hospital	242
Nice Hospital	471
	466
Cochin Hospital, Paris	395
HEGP, Paris	541
Saint Antoine Hospital, Paris	58
Tenon Hospital, Paris	235
Strasbourg Hospital	
SB3-CE: third-generation small bowel capsule endoscopy; HEGP: Hôpital Européen Georges Pompidou.	

med students were trained, and supervised by an expert reader (XD), to select and to delimitate any lesion found into the selected frames of interest. The delimitation process used Adobe Photoshop CS6 (Adobe Systems, United States) and GIMP softwares (GNOME Foundation, United States) with a Wacom (Wacom Co., Ltd, Japan) pen tablet connected to a laptop.

Central reading of selected frames

During several face-to-face meetings, three expert SB-CE readers (XD, SL, JPLM) screened all selected frames of interest (and associated 9-second video sequence including 25 frames upstream and 25 downstream the index still frames), and all delimitations of abnormal findings within each selected frames. Blurred or unanalyzable frames were excluded. Findings' delimitations were also reviewed, and remade when necessary.

Abnormal findings were first sorted into three different categories: (a) fresh blood and clots; (b) vascular findings; (c) ul-

cerative/inflammatory findings. Then, in these categories, all abnormal findings were sorted again according to their relevance. Relevance was defined as followed by a group of four SB-CE experts (FC, XD, GR, JCS). Fresh blood and clots [17], typical angiectasia [18] and ulcerated lesions [19] were considered "highly relevant" findings. Non-ulcerated but inflammatory findings (for instance erythema, edema, denudation) [20] were considered "moderately relevant" findings. Subtle vascular lesions (for instance erythematous patches, red spots/dots, phlebectasia, and so-called "diminutive angiectasia") were considered "poorly relevant" findings [18].

Control frames

A set of 206 normal and complete SB-CE videos was used to create a control dataset. Still frames were automatically extracted from these latter videos, every 1 % of the SB sequence (between 1 % and 100 % of the SB sequence). Thus, 100 normal frames were extracted per video. Again, for any normal extracted still frame, a short video sequence was also captured (with 25 frames upstream and downstream the index frame). All supposedly normal frames were reviewed by two readers (XD, RL). Doubtful, blurred or irrelevant frames were excluded.

Database splitting into two datasets

All the frames (and short video sequences) were randomly grouped through a computer-generated algorithm into two different, independent subsets of still frames, of equal sizes for each category (considering the presence and type of abnormal finding and the category of relevance). One dataset was dedicated to the development and training of algorithms (machine learning). The other dataset was dedicated to the testing and validation of algorithms.

Results

We collected 4,174 third-generation SB-CE from the 12 participating centers (► Table 1). We included in the database 1,480 videos (35 %) containing at least one frame of interest (with abnormal findings), and 206 normal videos (as controls). Clinical characteristics from these 1480 SB-CE were collected (59 % men, 41 % female, mean age 64). SB-CE were mainly indicated for obscure gastrointestinal bleeding (67 %), suspected Crohn's disease (CD) (12 %), and sometimes (21 %) for others various indications such as coeliac disease and Peutz-Jeghers polyposis. In total we extracted 6,013 still frames (with their adjacent short video sequences). Abnormal findings were delimited by the pre-med students within 5,124 frames, and then re-

viewed and sorted by the experts (► Fig. 1). Of the frames, 3103 contained images of vascular abnormalities (817 frames with highly relevant images of angiectasia, 2286 frames with other, poorly relevant, images of vascular lesions), 1370 frames contained images of ulcerative/inflammatory findings (1057 frames with highly relevant images of ulcerated lesions, 313 frames with moderately relevant images of inflammatory but not ulcerated lesions). Highly relevant images of fresh blood and clots were contained in 651 frames (► Fig. 2). Three-hundred-and-ninety-one frames with images of miscellaneous le-

sions (including tumors and polyps) were not delimited and not included in this first version of the database, because this number was considered too low for machine learning). Twenty-thousand normal frames (controls, with their short video sequences) were extracted from the 206 SB-CE normal and complete videos. Two subsets of 12,562 still frames (2562 with abnormal findings with an associated, delimited frame, and 10,000 normal, control, frames) were selected.

Discussion

The multicenter CAD-CAP database was created to promote the development of AI applications in the field of SB-CE. This database has several strengths, the first being the volume of collected data. Data collection was multicenter, thus providing over 4,000 SB-CE capsules videos and reports. Six thousand frames with abnormal findings and 20,000 normal frames, were provided. This makes CAD-CAP the largest published SB3-CE database so far. The second strength is the quality of the collected data. Only third-generation SB-CE were considered, which provides homogenous frames and videos (in terms of resolution, luminosity, and overall quality). CAD-CAP is composed of supervised and validated, delimited findings from SB-CE. Three levels of pertinency (relevance) of abnormal findings were defined before frames were sorted by a group of expert readers. Future CAD software could therefore be challenged at different levels (diagnosis performance for not/poorly/moderately/highly relevant lesions) by using the CAD-CAP database. It is unknown whether private CE or AI companies have such databases. Since 2017, the KID project by Koulaouzidis A et al. [13] has aimed to provide an Internet-based digital video atlas of capsule endoscopy for research purposes.

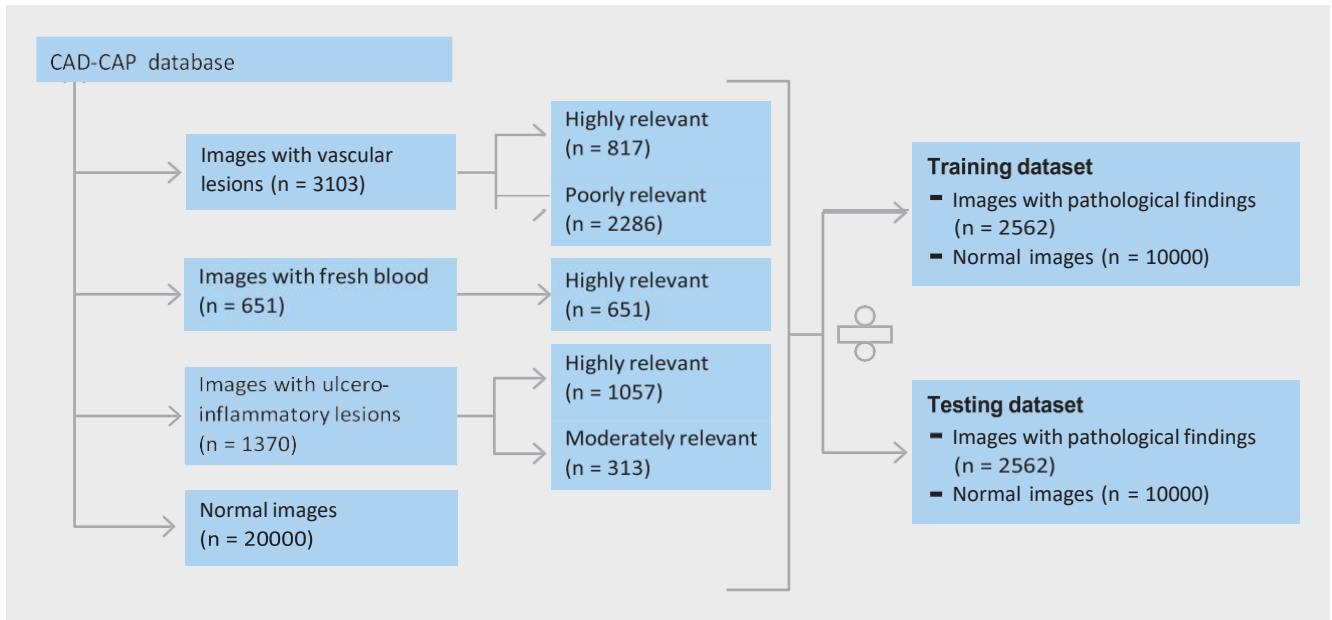
The CAD-CAP database has some limitations. First, some computer vision experts may consider the number of frames collected insufficient. For comparison, 128,175 retinal images frames were used by Gulshan et al. [7] to develop a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. Another 129,450 skin images were used by Esteva et al. [6] for development of an automated tool for classification of malignant vs benign skin lesions. However, the quality of collected data (with delimited abnormal findings within each frame of interest), and the number of normal frames used as controls may balance out this limitation. Second, the CAD-CAP database only used data from the SB-CE Pill-Cam SB3 system, thus restraining potential development of CAD algorithms with other SB-CE systems that have different image processing inner characteristics. Third, SB preparation quality was not considered in this database because of the lack of fully validated cleansing score for SB-CE [21]. Poor preparation quality should be taken into consideration as it may generate false-positive findings. Fourth, and most importantly, polyps/tumors are certainly "highly relevant findings," but such frames were not included in this first version of the CAD-CAP database because of their limited number. We plan to collect a significant number of frames with polypoid lesions from ongoing studies and include them in the next version of the database. Fifth, the clinical relevance of many SB findings is not de-



Fig. 1 Examples of native small bowel capsule endoscopy frames, with their corresponding delimitations. a Native deidentified still frame showing a vascular lesion considered "highly relevant." b Manual delimitation of the vascular lesion within frame 1a. c Native deidentified still frame showing a vascular lesion considered "poorly relevant." d Manual delimitation of the vascular lesion within frame 1c. e Native deidentified still frame showing an ulcerative lesion (considered "highly relevant"). f Manual delimitation of the ulcerative lesion within frame 1e. g Manual delimitation of the ulcerative lesion including mucosal inflammatory changes within frame 1e. h Native deidentified still frame showing an ulcero-inflammatory lesion (considered "moderately relevant"). i Manual delimitation of the ulcero-inflammatory lesion within frame 1 hour. j Native deidentified still frame showing fresh blood (considered "highly relevant"). k Manual delimitation of the fresh blood within frame 1j. l, m Two normal control frames (no delimitation needed)

termined yet in the literature. To overcome this limitation, our group is currently conducting a study based on an illustrated script questionnaire, aiming to better determine the clinical relevance of SB-CE findings. This study includes 576 frames with various types and numbers of findings, and different clinical

settings. It should provide a strong asset for a better classification of SB lesions seen in CE, thus allowing updating a second version of CAD-CAP with a more accurate findings' classification. Sixth, there was no sample size calculation. We tried to collect as many CE as possible, and to accumulate a high num-



► Fig. 2 CAD-CAP database distribution.

ber of high-quality images (i. e., with a clear delimitation of any finding within frames).

AI and machine learning processes are increasingly used for medical and research purposes. This shift provides a large variety of perspectives concerning image analysis and automated diagnostic tools. Because of the exponential growth in health-care data, new processing methods are needed for analysis of large data sets. That will completely transform our current clinical practice. Image recognition software can already make predictions as effectively as, or even better than, physicians [10, 22]. Machine learning directly from medical data could prevent clinical errors due to human cognitive biases, positively impacting patient care [23].

In medicine, AI is still in its nascent stage, but it should not be long before it becomes a part of everyday life in daily practice. For example, the first FDA-approved AI tool in medicine has already shown high sensitivity (87 %-90 %) and specificity (98 %) for detecting moderately severe retinopathy and macular edema, compared with 54 ophthalmologists and senior residents [7]. Similarly, another application has demonstrated the efficacy of an automated tool for classifying appearance of skin lesions [6]. Many others are under development.

Unsurprisingly, the AI revolution has also started in the field of endoscopy. Developed algorithms are now able to differentiate diminutive adenomas from hyperplastic polyps during colonoscopy in real time with high accuracy [5]. Other studies have shown that an automated detection tool for Barrett’s esophagus neoplasia can assist endoscopists in detecting early neoplasia on volumetric laser endomicroscopy [24]. Many more tools and studies are to come, as the field of AI is expanding exponentially [25].

Indeed, AI applications are also emerging in the field of CE. In the next few years, automated software should be able to facilitate more efficient reporting. Many studies have been con-

ducted to establish a cleansing score for SB-CE quality preparation. Different approaches have focused on the abundance of bubbles [26], red over green color [27, 28], and quality of bowel preparation in colon capsule endoscopy [29]. Algorithms based on machine learning approaches have also been developed to detect SB lesions [12, 28, 30]. Aoki et al. [16] developed a CAD algorithm for detection of erosions and ulcerations in SB-CE based on a deep convolutional neural network with sensitivity, specificity, and accuracy of 88.2 %, 90.9 % and 90.8 %, respectively. Although AI in endoscopy remains very promising, it still requires further research and clinical trials to be validated in daily clinical practice.

The CAD-CAP database has already been used for international challenges in medical image computing. During the 2017 and 2018 Medical Image Computing and Computer-Assisted Intervention (MICCAI) meetings, 14 and 20 research teams, respectively, were challenged to develop CAD tools dedicated to SB-CE [31]. We recently reported of an automated, highly sensitive, highly specific algorithm for diagnosis of SB angiectasia [12], which was developed using a subset of frames from the CAD-CAP database. This tool reached high performance with 100 % sensitivity and negative predictive value and a 96 % specificity and positive predictive value. Patent application and industrial partnership are also under consideration.

Conclusion

In conclusion, the CAD-CAP database aims to promote development of CAD reading in SB-CE. It represents an opportunity to design fully-automated software for detection of all SB lesions to facilitate and improve the future of CE reading, reviewing and reporting.

Acknowledgements

The CAD-CP database was developed with the support of the Société Nationale Française de Gastroentérologie (SNFGE, grant FARE) and with the support of MSD France.

Competing interests

Dr. Saurin is a consultant for Capsovision, Medtronic, and Intromedic. Dr. Rahmi is a consultant for Medtronic. Dr. Leenhardt is cofounder and shareholder of Augmented Endoscopy, and has given lectures for Abbvie. Dr. Dray is cofounder and shareholder of Augmented Endoscopy and has acted as a consultant for Alfasigma, Bouchara Recordati, Boston Scientific, Fujifilm, Medtronic, and Pentax. Dr. Huvelin is a consultant for Medtronic.

References

- [1] Iddan G, Meron G, Glukhovskiy A et al. Wireless capsule endoscopy. *Nature* 2000; 405: 417
- [2] McAlindon ME, Ching H-L, Yung D et al. Capsule endoscopy of the small bowel. *Ann Transl Med* 2016; 4: 369
- [3] Rajpurkar P, Irvin J, Zhu K et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225* (2017)
- [4] Chilamkurthy S, Ghosh R, Tanamala S et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *The Lancet* 2018; 392: 2388-2396
- [5] 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
- [6] Esteva A, Kuprel B, Novoa RA et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; 542: 115-118
- [7] Gulshan V, Peng L, Coram M et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* 2016; 316: 2402-2410
- [8] Commissioner O of the. Press Announcements - FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. <https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm604357.htm> [Apr 2018]
- [9] Byrne MF, Shahidi N, Rex DK. Will Computer-Aided Detection and Diagnosis Revolutionize Colonoscopy? *Gastroenterology* 2017; 153: 1460-1464.e1
- [10] Chen P-J, Lin M-C, Lai M-J et al. Accurate Classification of Diminutive Colorectal Polyps Using Computer-Aided Analysis. *Gastroenterology* 2018; 154: 568-575
- [11] Iakovidis DK, Koulaouzidis A. Automatic lesion detection in capsule endoscopy based on color saliency: closer to an essential adjunct for reviewing software. *Gastrointest Endosc* 2014; 80: 877-883
- [12] Leenhardt R, Vasseur P, Li C et al. A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. *Gastrointest Endosc* 2019; 89: 189-194
- [13] Koulaouzidis A, Iakovidis DK, Yung DE et al. KID Project: an internet-based digital video atlas of capsule endoscopy for research purposes. *Endosc Int Open* 2017; 5: E477-E483
- [14] Iakovidis DK, Koulaouzidis A. Software for enhanced video capsule endoscopy: challenges for essential progress. *Nat Rev Gastroenterol Hepatol* 2015; 12: 172-186
- [15] Lee J-G, Jun S, Cho Y-W et al. Deep Learning in Medical Imaging: General Overview. *Korean J Radiol* 2017; 18: 570-584
- [16] Aoki T, Yamada A, Aoyama K et al. Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network. *Gastrointest Endosc* 2019; 89: 357-363.e2
- [17] Saurin J-C, Delvaux M, Gaudin J-L et al. Diagnostic value of endoscopic capsule in patients with obscure digestive bleeding: blinded comparison with video push-enteroscopy. *Endoscopy* 2003; 35: 576-584
- [18] Leenhardt R, Li C, Koulaouzidis A et al. Nomenclature and Semantic Description of Vascular Lesions in Small Bowel Capsule Endoscopy: an International Delphi Consensus Statement. *Endosc Int Open* 2019; 07: E372-E379
- [19] Buisson A, Filippi J, Amiot A et al. Su1229 Definitions of the Endoscopic Lesions in Crohn's Disease: Reproducibility Study and GETAID Expert Consensus. *Gastroenterology* 2015; 148: S445
- [20] Gal E, Geller A, Fraser G et al. Assessment and validation of the new capsule endoscopy Crohn's disease activity index (CECDAl). *Dig Dis Sci* 2008; 53: 1933-1937
- [21] Yung DE, Rondonotti E, Sykes C et al. Systematic review and meta-analysis: is bowel preparation still necessary in small bowel capsule endoscopy? *Expert Rev Gastroenterol Hepatol* 2017; 11: 979-993
- [22] Mori Y, Kudo S-E, 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
- [23] Miller DD, Brown EW. Artificial Intelligence in Medical Practice: The Question to the Answer? *Am J Med* 2018; 131: 129-133
- [24] Swager A-F, van der Sommen F, Klomp SR et al. Computer-aided detection of early Barrett's neoplasia using volumetric laser endomicroscopy. *Gastrointest Endosc* 2017; 86: 839-846
- [25] Kashin S. Artificial intelligence: the rise of the machines. *United European Gastroenterology Week. 20-24 oct 2018, Vienna, Austria.*
- [26] Pietri O, Rezgui G, Histace A et al. Development and validation of an automated algorithm to evaluate the abundance of bubbles in small bowel capsule endoscopy. *Endosc Int Open* 2018; 6: E462-E469
- [27] Van Weyenberg SJB, De Leest HTJ, Mulder CJJ. Description of a novel grading system to assess the quality of bowel preparation in video capsule endoscopy. *Endoscopy* 2011; 43: 406-411
- [28] Abou AliE, Histace A, Camus M et al. Development and validation of a computed assessment of cleansing score for evaluation of quality of small-bowel visualization in capsule endoscopy. *Endosc Int Open* 2018; 6: E646-E651
- [29] Becq A, Histace A, Camus M et al. Development of a computed cleansing score to assess quality of bowel preparation in colon capsule endoscopy. *Endosc Int Open* 2018; 6: E844-E850
- [30] Fan S, Xu L, Fan Y et al. Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images. *Phys Med Biol* 2018; 63: 165001
- [31] Bernal J, Tajkbaksh N, Sanchez FJ et al. Comparative Validation of Polyp Detection Methods in Video Colonoscopy: Results From the MICCAI 2015 Endoscopic Vision Challenge. *IEEE Trans Med Imaging* 2017; 36: 1231-1249

Plus récemment, la volonté d'enrichir cette banque d'images a donné naissance au projet NOCE (Nouveaux Outils Capsule Endoscopique) pour la construction d'un large réseau de partage d'images en collaboration entre l'Assistance Publique des Hôpitaux de Paris (APHP) et le laboratoire ETIS de l'ENSEA (Protocole en annexe). L'objectif de NOCE est de colliger les données de milliers de VCE au sein de 16 unités d'endoscopie (APHP) via l'Entrepôt de Données de Santé (EDS) afin de créer une vérité terrain robuste et exhaustive pour le développement des algorithmes d'IA (ou le renforcement de ceux déjà développés avec CAD CAP) (Figure 13). Le recueil de données a commencé, avec un volume attendu de l'ordre de 2000 VCE par an, soit 10000 en 5 ans.

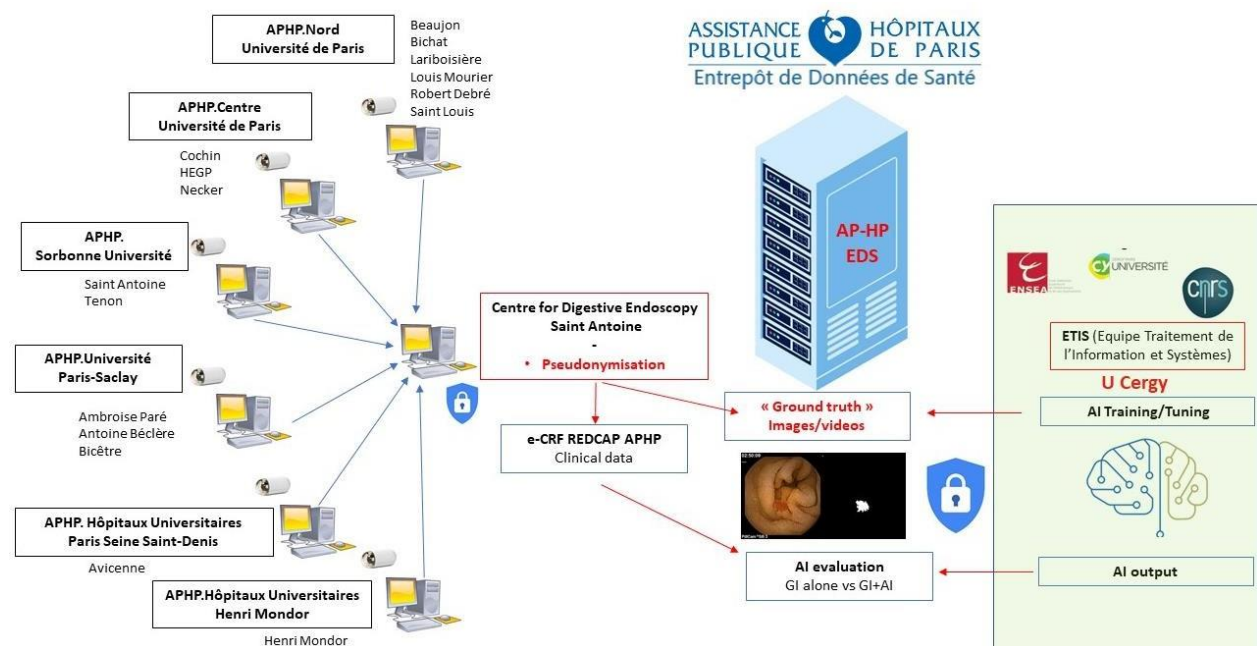
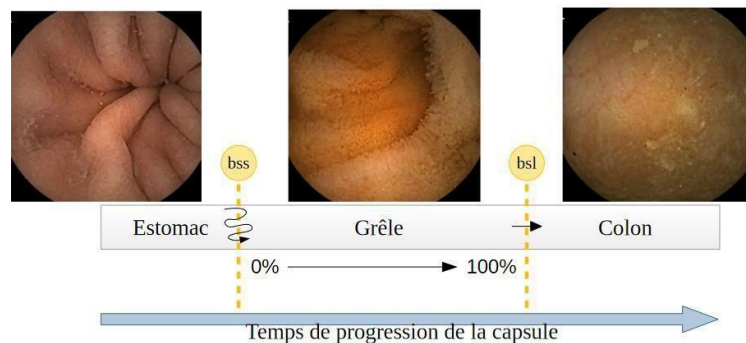


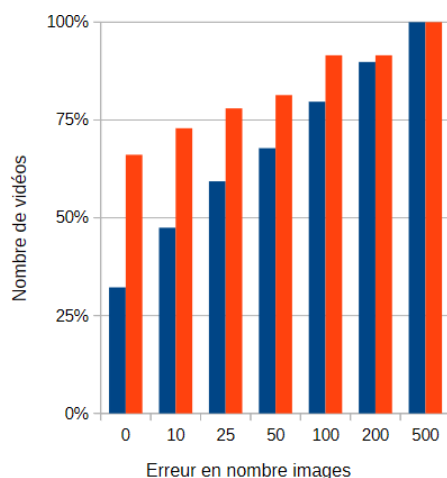
Figure 13 : Projet NOCE : Nouveaux Outils en Capsule Endoscopique

7.5. Evaluation de la complétude de l'examen du grêle

La complétude de l'exploration de l'IG est un des critères de qualité de la réalisation d'une VCE. Elle se définit par un enregistrement vidéo allant de la première à la dernière image de l'IG. Le taux de complétude de l'exploration de l'IG par VCE établi dans la littérature médicale est de 85.2% (11). Il paraît nécessaire d'approcher cette question de recherche afin de se positionner dans la perspective d'une capsule panentérique dont l'évaluation des différents segments du tube digestif pourra être facilitée par l'IA. Pour cela, nous avons mené

un travail d'apprentissage machine à l'échelle de l'image grâce à une banque d'images regroupant 10350 images gastrique, 27000 images d'IG et 13400 images du côlon. L'utilisation d'un classifieur unique pour la distinction des repères anatomiques entre l'estomac, l'IG et le côlon était source de trop nombreuses erreurs. Nous avons donc entraîné séparément deux classifieurs distincts estomac/grêle et grêle/côlon. La méthodologie utilisait des CNN en se basant sur des occurrences d'épisodes définis grâce à une fenêtre temporelle glissante le long de la séquence d'évènement. Pour chaque classifieur, l'analyse se faisait en moyennant la sortie du classifieur dans la fenêtre temporelle glissante. Chaque instance de la fenêtre glissante contenant un épisode était considérée comme une occurrence. On retournait alors la première occurrence où l'on dépassait le seuil critique. Une évaluation préliminaire à l'échelle de 60 vidéos a été réalisée et a permis d'établir un placement correct des repères anatomiques dans 90 % des vidéos avec environ 200 images d'erreurs pour la première image de l'IG et environ 100 images d'erreur pour la dernière image de l'IG (Figure 14). Ces résultats préliminaires sont prometteurs et seront poursuivis afin d'améliorer les performances. Ce travail est en cours (dans le cadre d'un Master de Sciences), à partir de la nouvelle collection de données (NOCE).





En bleu : la première image du grêle

En rouge : la dernière image du grêle

Figure 14 : Illustration du temps de progression de la capsule et résultats préliminaires

7.6. Evaluation de la propreté de l'intestin grêle

La qualité de visualisation de la muqueuse de l'IG est variable selon les examens et la préparation entourant l'ingestion de la VCE. Elle représente une des limites de la VCE. En effet elle peut être réduite notamment en cas de préparation insuffisante et/ou de non-suivi du régime préalable. L'appréciation globale de la propreté de l'IG doit faire partie du compte rendu de l'examen mais cet indicateur reste très dépendant de l'appréciation du lecteur de VCE. En cas de visualisation insuffisante de la muqueuse, la réalisation d'un nouvel examen peut être discutée si par exemple une suspicion forte de MC persiste. Il existe une variété de scores disponibles dans la littérature pour évaluer la qualité de la visualisation (39). Ces scores ont le défaut d'être subjectifs car dépendants de l'opérateur humain. A titre d'exemple, le score développé par l'équipe de Brotz C. et al. (40), a l'inconvénient de présenter une reproductibilité inter-observateur médiocre (41). Il existe donc un besoin réel autour du développement d'un outil automatique et reproductible d'évaluation de la propreté en VCE. Nous avons mené un travail sur ce sujet qui a fait l'objet d'une publication scientifique (42).

A neural network-based algorithm for assessing the cleanliness of small bowel during capsule endoscopy

Authors

Romain Leenhardt^{1, 2}, Marc Souchaud², Guy Houist³, Jean-Philippe Le Mouel⁴, Jean-Christophe Saurin⁵, Franck Cholet⁶, Gabriel Rahmi⁷, Chloé Leandri⁸, Aymeric Histace^{2,*}, Xavier Dray^{1, 2,*}

Institutions

- 1 Sorbonne University, Center for Digestive Endoscopy, Saint Antoine Hospital, APHP, Paris, France
- 2 ETIS, Université de Cergy-Pontoise, ENSEA, CNRS, Cergy-Pontoise, France
- 3 Gastroenterology Department, Centre Hospitalier Sud Francilien, Corbeil-Essonnes, France
- 4 Gastroenterology, Amiens University Hospital, Université de Picardie Jules Verne, Amiens, France
- 5 Gastroenterology and Endoscopy Unit, Edouard Herriot Hospital, Lyon, France
- 6 Endoscopy Unit, CHU La Cavale Blanche, Brest, France
- 7 Department of Gastroenterology and Digestive Endoscopy, Georges-Pompidou European Hospital, APHP, Paris, France
- 8 Gastroenterology Department, Cochin Hospital, APHP, Paris, France

submitted 14.6.2020

accepted after revision 30.10.2020

published online 2.11.2020

Bibliography

Endoscopy

DOI 10.1055/a-1301-3841

ISSN 0013-726X

© 2020. Thieme. All rights reserved.

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

Corresponding author

Xavier Dray, MD, PhD, Center for Digestive Endoscopy, Saint Antoine Hospital, 184 rue du Faubourg Saint Antoine, 75012 Paris, France
xavier.dray@aphp.fr

ABSTRACT

Background Cleanliness scores in small-bowel capsule endoscopy (SBCE) have poor reproducibility. The aim of this study was to evaluate a neural network-based algorithm for automated assessment of small-bowel cleanliness during capsule endoscopy.

Methods 600 normal third-generation SBCE still frames were categorized as “adequate” or “inadequate” in terms of cleanliness by three expert readers, according to a 10-point scale, and served as a training database. Then, 156 third-generation SBCE recordings were categorized in a consensual manner as “adequate” or “inadequate” in terms of cleanliness; this testing database was split into two independent 78-video subsets for the tuning and evaluation of the algorithm, respectively.

Results Using a threshold of 21% “adequate” still frames per video to achieve the best performance, the algorithm yielded a sensitivity of 90.3%, specificity of 83.3%, and accuracy of 89.7%. The reproducibility was perfect. The mean calculation time per video was 3 (standard deviation 1) minutes.

Conclusion This neural network-based algorithm allowing automatic assessment of small-bowel cleanliness during capsule endoscopy was highly sensitive and paves the way for automated, standardized SBCE reports.

Introduction

Capsule endoscopy has become the mainstay of small-bowel examination. Bile, bubbles, residues, and liquids can alter the visualization of the small-bowel mucosa and decrease the diagnostic performance of small-bowel capsule endoscopy (SBCE) examinations. Several cleanliness scores have been proposed for SBCE, but none had undergone external validation [1]. Recently, the Overall Adequacy Assessment (OAA) score of small-bowel cleanliness during capsule endoscopy proposed by Brotz

et al. [2] was demonstrated to have moderate interobserver reproducibility [3].

Overall, although the use of a cleanliness score is recommended for SBCE [4, 5], no reliable scale is available for clinical practice. Thus, any assessment of preparation regimens for SBCE remains challenging. The European Society of Gastrointestinal Endoscopy (ESGE) and United European Gastroenterology (UEG) have recently called for “the development of software for assessment of the quality of small-bowel preparation” [4]. The aim of this study was to develop and assess a deep

* These authors contributed equally.

Methods

Deep learning algorithm

Machine learning is a type of artificial intelligence technique that allows the analysis of full-length videos [6]. Machine learning approaches based on convolutional neural networks have demonstrated good performance for the automated detection of small-bowel lesions during capsule endoscopy [7 - 9]. The neural network-based algorithm proposed herein was trained, tuned, and evaluated using a customized 16-layer Visual Geometry Group architecture [10], inspired by a previous survey [11]. The neural network algorithm included convolutional layers at different scales, a ReLu unit, and MaxPool layers for the extraction of features.

Training dataset at the still image level

A total of 600 frames from SBCE examinations were first used to train the neural network-based algorithm at the still image level. The dataset, which has been described elsewhere [11], included 600 frames that were randomly extracted from 30 normal, complete, de-identified videos of SBCE using third-generation capsules (Pillcam SB3; Medtronic, Minneapolis, Minnesota, USA). These frames were analyzed independently by three experts in order to assess small-bowel cleanliness according to the 5-item, 10-point, Quantitative Index (QI) by Brotz et al. [2]. A still frame was categorized as having adequate cleanliness when the mean score of the three experts' scores was $\geq 7/10$. This threshold was based on the distribution curve of the quantitative score developed by Brotz et al. [2], and in our previous demonstration the threshold produced high 0.81 and 0.87 kappa coefficients between experts [11]. Data augmentation was used to increase the robustness of the training process and led us to consider a pool of 3000 images. Receiver operating characteristic (ROC) curves were obtained from the comparison of the outputs of various versions of the neural network-based algorithms with the experts evaluation. The optimal algorithm, defined as one that could distinguish adequate cleanliness from inadequate cleanliness in small-bowel still frames, with the best trade-off between sensitivity and specificity, was selected by computing the operating point of the best-fitting ROC curve.

Tuning and evaluation datasets at the video level

Following preliminary training using still frames, the neural network-based algorithm was then tuned to categorize videos. For this purpose, we used 156 complete, de-identified, third-generation capsule (Pillcam SB3) video recordings from the PREPINTTEST multicenter randomized controlled trial (NCT01267981) [12]. This trial aimed to compare the diagnostic yield of second- and third-generation capsule endoscopes according to three different preparation regimens. The 156 complete third-generation video recordings retrieved from the five most active centers of the PREPINTTEST trial were edited in a universal video format (mpeg) from the first to the last frame of small-bowel examination. All videos were independently reviewed at an accelerated speed ($\times 32$) by three expert readers, and categorized as "adequate" or "inadequate" in terms of

small-bowel cleanliness using the OAA scale (► Table 1), as described by Brotz et al. [2]. Each video categorized as "adequate" by two or three experts was considered to be consensually "adequate"; similarly, "inadequate" categorization by at least two experts was considered to be consensually "inadequate". The experts' consensual adequate/inadequate classification was considered to be the gold standard.

The 156-video dataset was then randomly split into two independent subsets, each with 78 video clips and an equal proportion of videos classified as adequate and inadequate accord-

ing to the expert standard (► Fig. 1). Within each subset, each video was processed to extract the full set of constituent images. Each image of each video was then categorized by the neural network-based algorithm, and a computed score (from 0 and 100 %) was established based on the proportion of "ade-

quate" images per video (► Fig. 2). Using the first subset of 78 videos (tuning subset), the optimal score was tuned, and a corresponding ROC curve was plotted. The operating point was used to determine the optimal proportion of "adequate" still frames per video for the best prediction of the experts' consensual classification. The performance of the best-fitting neural network-based algorithm was then evaluated on the second random subset of 78 videos (evaluation subset), using the experts' consensual classification (adequate vs. inadequate) as reference.

End points and statistics

The primary end point was the sensitivity of the algorithm for predicting adequate cleanliness of SBCE at the video level. The specificity, positive (PPV) and negative (NPV) predictive values, reproducibility, and calculation times were considered secondary end points. Data were reported as percentages with 95 % confidence interval (CI) and mean with standard deviation (SD), as appropriate.

Results

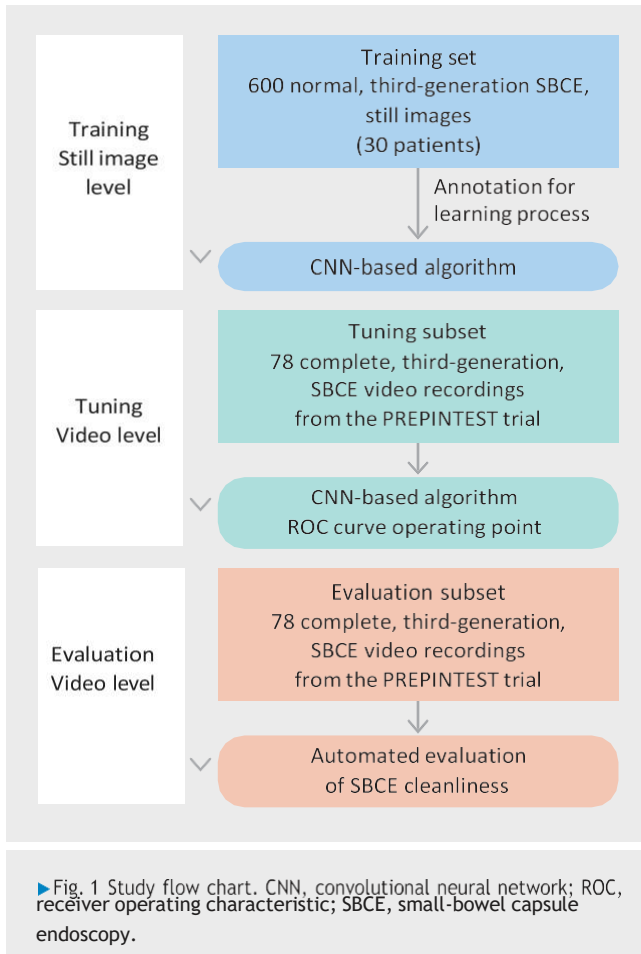
After data augmentation and initial training on still frames, the best-fitting neural network-based algorithm demonstrated a sensitivity of 94.7 % (95 %CI 92.2 % - 97.2 %), a specificity of 94.0 % (95 %CI 91.3 % - 96.7 %), a PPV of 87.8 % (95 %CI 84.1 % - 91.5 %), an NPV of 97.5 % (95 %CI 95.7 % - 99.3 %), and an overall accuracy of 95.7 % (95 %CI 93.4 % - 98.0 %) (► Fig. 3a).

In the second database, 144 videos (92.3 %) were categorized as adequate by the experts. The same proportion (72 out of 78, 92.3 %) of "adequate" SBCE videos were randomly dis-

tributed into the tuning and evaluation subsets (► Table 2). The optimal proportion of "adequate" still frames per video for the prediction of the experts' consensual classification was 21%* (► Fig. 3b).

In the evaluation subset (► Table 2), 65 videos were categorized as adequate by the neural network-based algorithm among 72 videos classified as adequate by the experts, providing a sensitivity of 90.3 % (95 %CI 83.7 % - 96.9 %). Conversely, the algorithm categorized 5 videos as inadequate among 6 videos classified as inadequate by the experts, yielding a specificity of 83.3 % (95 %CI 75.0 % - 91.6 %). PPV was 98.5 % (95 %CI

*Version corrigée de l'article: seuil corrigé à 21 % au lieu de 79%

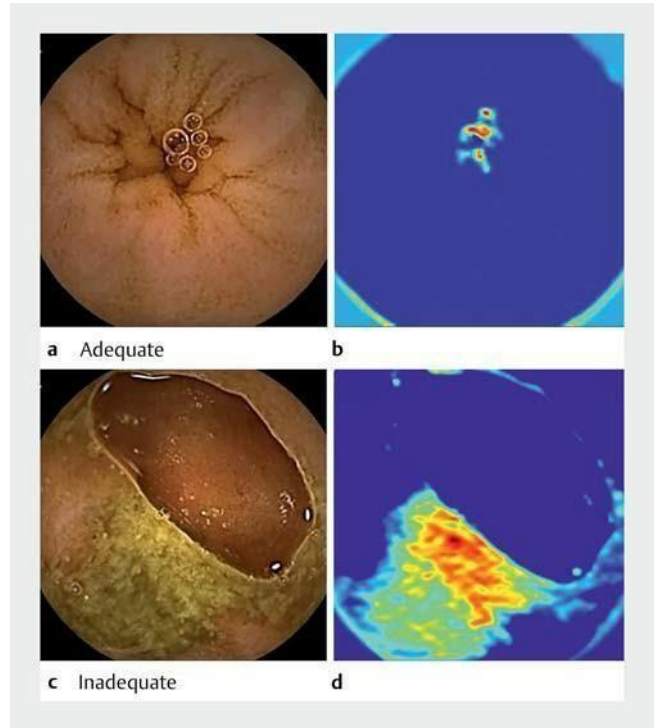


95.8 % - 100.0 %), NPV was 41.7 % (95 %CI 30.8 % - 52.6 %), and diagnostic accuracy was 89.7 % (95 %CI 82.9 % - 96.4 %). Reproducibility was perfect. Calculation time per video was 3 (SD 1) minutes.

Discussion

The neural network-based algorithm developed in the current study allowed a rapid and reproducible automated assessment of the cleanliness of the small bowel during capsule endoscopy. The algorithm demonstrated high sensitivity (90.3 %, PPV 98.5 %) but moderate specificity (83.3 %, NPV 41.7 %).

The assessment and reporting of small-bowel cleanliness is listed in various guidelines as a performance measure for SBCE [4, 5]; however, preparation scales have poor reproducibility [2]. Among the most popular scores in this setting, the OAA score proposed by Brotz et al. [2] has an interobserver kappa coefficient of only 0.41 (95 %CI 0.31 - 0.50). Our own data (kappa 0.52 - 0.71) confirmed this drawback [3]. Thus, a quality initiative process by the ESGE and UEG recently called for “the development of software for assessment of the quality of small-bowel preparation” [4]. Previous studies based on still frames have demonstrated the high diagnostic performances of hand-crafted computerized algorithms in assessing small-bowel cleanliness during capsule endoscopy [11, 13]. It has since

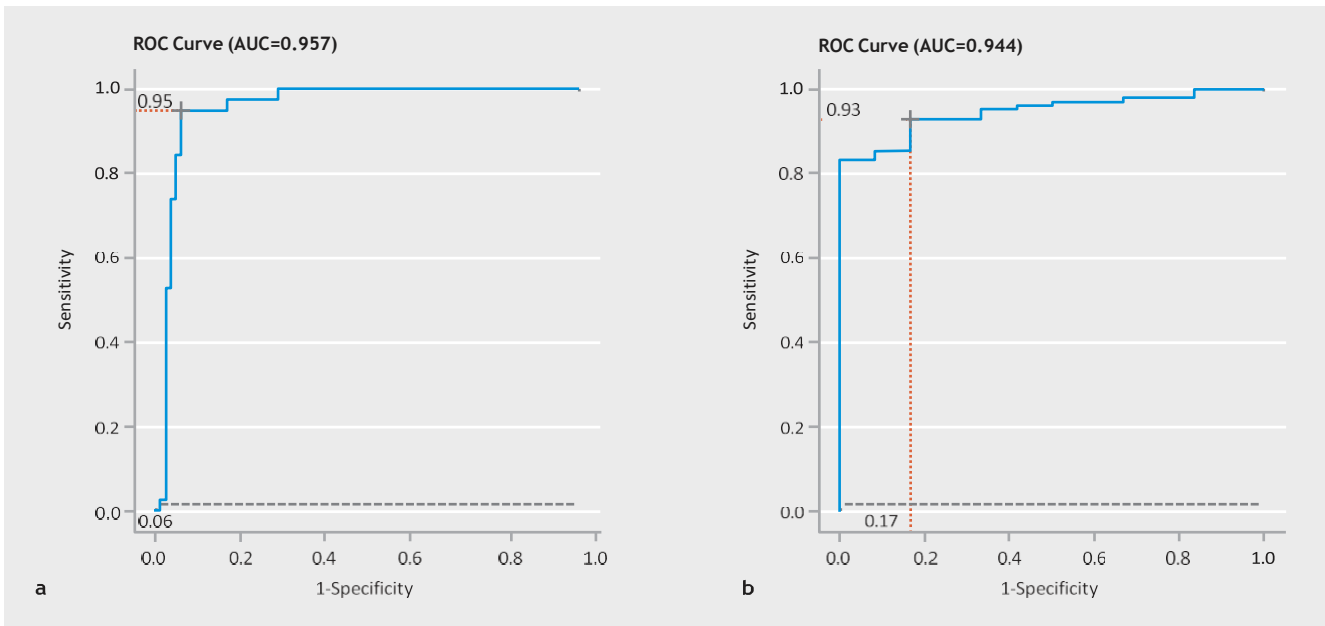


▶ Fig. 2 Heat map illustrating final decisions of “adequate” versus “inadequate” cleanliness, as taken at the image level by the neural network-based algorithm. The left column (a and c) displays native images, labeled with their final decision according to the neural network-based algorithm. The right column (b and d) shows the output of the neural network-based algorithm as a related probabilistic heat map for the corresponding native image. Redness highlights areas contributing to an “inadequate” decision, whereas blueness highlights areas contributing to an “adequate” decision. Such heat maps illustrate that the neural network-based algorithm is able to learn specific features related to the physicians’ visual examination.

been demonstrated that neural network-based algorithms can significantly reduce the reading time of capsule endoscopy videos [7 - 9, 14], but they have not yet been used to rate small-bowel cleanliness. In our current work, a neural network-based algorithm was shown to be a valuable tool for better standardization and quality reporting in capsule endoscopy.

The strengths of this study include the video databases, which originated from a multicenter, randomized trial [12] with three different bowel preparations, and thus offered some variability in small-bowel cleanliness in the recordings. In addition, the assessment by three expert readers circumvented the issue of the moderate reproducibility of the OAA score by Brotz et al. [2].

Some limitations should also be mentioned, most of which relate to the retrospective design of the study. All of the videos from the PREPINTEST trial were performed in cases of obscure gastrointestinal bleeding and did not include patients with Crohn’s disease. Moreover, our algorithm tool would be of little help in the setting of active bleeding. The results cannot be generalized because only third-generation capsule recordings were included. Another limitation is the small number of “in-



► Fig. 3 Receiver operating characteristic (ROC) curves from the comparison of outputs from the neural network-based algorithm and the experts' evaluation. a The learning subset. b The tuning subset.

Table 1 Overall Adequacy Assessment (OAA) and Qualitative Evaluation (QE) scale from Brotz et al. [2].

OAA	QE	Criteria
Adequate	Excellent	Visualization of $\geq 90\%$ of mucosa; no, or minimal, fluid and debris, bubbles, and bile/chyme staining; no, or minimal, reduction of brightness
	Good	Visualization of $\geq 90\%$ of mucosa; mild fluid and debris, bubbles, and bile/chyme staining; mildly reduced brightness
Inadequate	Fair	Visualization of $< 90\%$ of mucosa; moderate fluid and debris, bubbles, and bile/chyme staining; moderately reduced brightness
	Poor	Visualization of $< 80\%$ of mucosa; excessive fluid and debris, bubbles, and bile/chyme staining; severely reduced brightness

adequately” cleansed small-bowel examinations in both the tuning and evaluation datasets, according to the experts. However, the proportion (7.7 %) of inadequately cleansed small bowel during capsule endoscopy is close to that recommended ($< 5\%$) by the ESGE and UEG [4], but lower compared with the

literature (10 % - 40 %) [15]. An explanation may be that patients who participated in the PREPINTEST trial might have been more compliant with the preparation modalities than patients in real-life conditions. Nevertheless, this small number had a high impact on any calculation of performance involving true negative cases as references (mostly the specificity and NPV). Overall, a raw “inadequate” result given by the algorithm should

► Table 2 Performances of the neural network-based algorithm outputs for the assessment of small-bowel cleanliness, compared with the consensual overall adequacy assessment [2] by the experts.

Videos	Algorithm output		Total
	Adequate	Inadequate	
Experts' consensual evaluation	Adequate	65	72
	Inadequate	1	6
Total	66	12	78

better be rendered as “doubtful” to capsule endoscopy

readers, with a final evaluation of cleanliness belonging to the reader. Overall, the reader’s opinion on cleanliness would be necessary in 10 % of cases only; in the remaining 90 % of cases, only validation would be required.

Conclusion

A neural network-based algorithm allowing the automated assessment of small-bowel cleanliness during capsule endoscopy was demonstrated to be rapid, reproducible, and highly sensitive. This tool paves the way for automated, standardized cap-

sule endoscopy reports. This neural network-based algorithm could also allow valid comparisons of different preparation regimens, and capsule endoscope diagnostic yield according to cleanliness.

Acknowledgment

This study has been made possible with the support of Norgine Pharma, France, and the support of the Société Nationale Française de Gastroentérologie (FARE grant).

Competing interests

Xavier Dray, Romain Leenhardt, and Aymeric Histace are cofounders and shareholders of Augmented Endoscopy. Xavier Dray has acted as a consultant for Boston Scientific and Norgine, and has presented lectures for Fujifilm, Medtronic, and Pentax. Romain Leenhardt has presented lectures for Abbvie. All other authors declare that they have no conflicts of interest.

References

- [1] Niv Y. Efficiency of bowel preparation for capsule endoscopy examination: a meta-analysis. *World J Gastroenterol* 2008; 14: 1313-1317
- [2] Brotz C, Nandi N, Conn M et al. A validation study of 3 grading systems to evaluate small-bowel cleansing for wireless capsule endoscopy: a quantitative index, a qualitative evaluation, and an overall adequacy assessment. *Gastrointest Endosc* 2009; 69: 262-270
- [3] Dray X, Houist G, Le Mouel JP et al. Prospective evaluation of third-generation small bowel capsule endoscopy videos by independent readers demonstrates poor reproducibility of cleanliness classifications. *Clin Res Hepatol Gastroenterol*. (in press)
- [4] Spada C, McNamara D, Despott EJ et al. Performance measures for small-bowel endoscopy: a European Society of Gastrointestinal Endoscopy (ESGE) Quality Improvement Initiative. *United Eur Gastroenterol J* 2019; 7: 614-641
- [5] Enns RA, Hookey L, Armstrong D et al. Clinical practice guidelines for the use of video capsule endoscopy. *Gastroenterology* 2017; 152: 497-514
- [6] Le Berre C, Sandborn WJ, Aridhi S et al. Application of artificial intelligence to gastroenterology and hepatology. *Gastroenterology* 2020; 158: 76-94
- [7] Ding Z, Shi H, Zhang H et al. Gastroenterologist-level identification of small bowel diseases and normal variants by capsule endoscopy using a deep-learning model. *Gastroenterology* 2019; 157: 1044-1054
- [8] Leenhardt R, Vasseur P, Li C et al. A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. *Gastrointest Endosc* 2019; 89: 189-194
- [9] Aoki T, Yamada A, Aoyama K et al. Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network. *Gastrointest Endosc* 2019; 89: 357-363
- [10] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. Presented at the International Conference on Learning Representations; 2015 May 7-9; San Diego, California <https://arxiv.org/pdf/1409.1556.pdf>
- [11] Oumrani S, Histace A, Abou Ali E et al. Multi-criterion, automated, high-performance, rapid tool for assessing mucosal visualization quality of still images in small bowel capsule endoscopy. *Endosc Int Open* 2019; 7: E944-E948
- [12] Cholet F, Rahmi G, Gaudric M et al. Does polyethylene glycol cleansing purge improve video capsule endoscopy diagnostic yield in obscure gastrointestinal bleeding? *Endoscopy* 2018; 50: S18
- [13] Van Weyenberg SJB, De Leest HTJ, Mulder CJJ. Description of a novel grading system to assess the quality of bowel preparation in video capsule endoscopy. *Endoscopy* 2011; 43: 406-411
- [14] Saito H, Aoki T, Aoyama K et al. Automatic detection and classification of protruding lesions in wireless capsule endoscopy images based on a deep convolutional neural network. *Gastrointest Endosc* 2020; 92: 144-151
- [15] Lapalus M-G, Ben Soussan E, Saurin J-C et al. Capsule endoscopy and bowel preparation with oral sodium phosphate: a prospective randomized controlled trial. *Gastrointest Endosc* 2008; 67: 1091-1096

7.7. Elaboration d'une vérité terrain en vidéocapsule

7.7.1. Terminologie et sémiologie en vidéocapsule

A l'issue de la lecture d'une VCE, le médecin rédige un compte rendu décrivant les différents éléments constatés. La standardisation des comptes rendus est un des critères de qualité de la réalisation des VCE. Pour autant, le vocabulaire utilisé dans ces comptes rendus de VCE diffère souvent d'un lecteur à l'autre. Une initiative dans les années 2000 a consisté à élaborer un lexique appelée Capsule Endoscopy Structured Terminology (CEST) (43). Cet effort a permis de retenir des mots-clés. Mais ces mots-clés n'étaient associés à aucune description ou définition, limitant la portée de ce travail à un lexique, là où un dictionnaire illustré serait attendu. Ainsi, 20 ans plus tard, l'interprétation des lésions visibles en VCE souffre encore d'une importante variabilité inter et intra observateur. En effet, une méta-analyse regroupant 52 études, démontrait qu'une concordance «bonne ou parfaite» inter et intra observateurs étaient obtenues seulement dans 23 % et 36 % des cas respectivement (44).

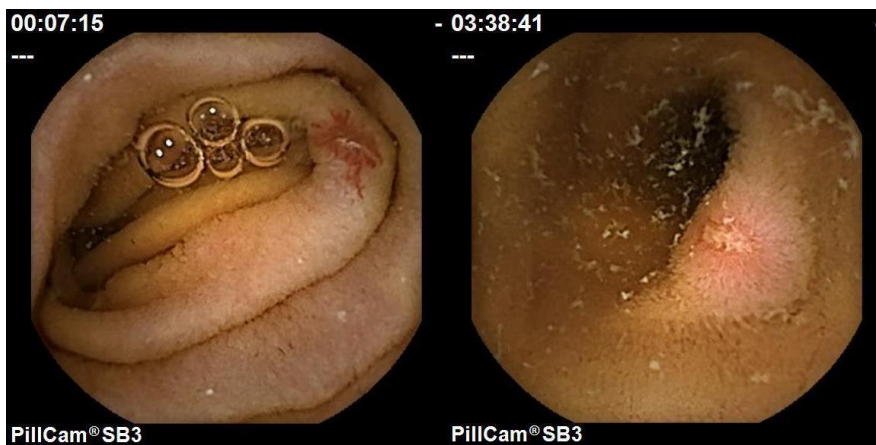


Figure 15 : Images de VCE-G montrant une angiodysplasie (photo de gauche) et ulcération superficielle (photo de droite).

Deux travaux visaient à harmoniser les pratiques en proposant une terminologie et des descriptions consensuelles des lésions les plus fréquemment visibles en VCE. En cas de SSIG, la lésion vasculaire la plus fréquemment mise en évidence et responsable du saignement est une angiodysplasie intestinale ou angiectasie (AGD). En cas de suspicion de MC de l'IG, les lésions inflammatoires à type d'ulcérations profondes, superficielles et d'érosions aphtoïdes sont recherchées. Ainsi, une première étude s'est attachée à définir et décrire les lésions vasculaires de l'IG (45). Un groupe de 18 experts a été sollicité pour établir de façon consensuelle selon la méthodologie Delphi une nomenclature illustrée des lésions vasculaires les plus fréquentes. Une deuxième étude regroupant 32 experts et utilisant la même méthodologie Delphi a permis d'établir de façon consensuelle une nomenclature illustrée des lésions inflammatoires les plus fréquentes de l'IG (46). Des études similaires ont été ensuite réalisées par d'autres membres de notre groupe de travail, pour d'autres types de lésions : diverticules de l'IG (47), zones d'atrophie muqueuse (48), auxquels nous avons contribué, et lésions sous muqueuse (49).

Nomenclature and semantic description of vascular lesions in small bowel capsule endoscopy: an international Delphi consensus statement*



Authors

Romain Leenhardt¹, Cynthia Li^{1, 2}, Anastasios Koulaouzidis³, Flaminia Cavallaro⁴, Franck Cholet⁵, Rami Eliakim⁶, Ignacio Fernandez-Urien⁷, Uri Kopylov⁶, Mark McAlindon⁸, Artur Németh⁹, John N. Plevis³, Gabriel Rahmi¹⁰, Emanuele Rondonotti¹¹, Jean-Christophe Saurin¹², Gian Eugenio Tontini¹³, Ervin Toth⁹, Diana Yung³, Philippe Marteau¹, Xavier Dray¹

Institutions

- 1 Sorbonne Université & Department of Hepatogastroenterology, APHP, Saint Antoine Hospital, Paris, France
- 2 Drexel University, College of Arts & Sciences, Philadelphia, Pennsylvania, United States
- 3 Centre for Liver and Digestive Disorders, The Royal Infirmary of Edinburgh, Edinburgh, United Kingdom
- 4 Gastroenterology & Digestive Endoscopy Unit, IRCCS Policlinico San Donato, San Donato Milanese, Milan, Italy
- 5 Department of Gastrointestinal Endoscopy, Brest Teaching Hospital, Brest, France
- 6 Department of Gastroenterology, Chaim Sheba Medical Center, Sackler School of Medicine, Tel-Aviv University, Tel-Aviv, Israel
- 7 Department of Gastroenterology - NHC, Pamplona, Spain
- 8 Academic Unit of Gastroenterology and Hepatology, Sheffield Teaching Hospitals NHS Trust, Sheffield, UK
- 9 Department of Gastroenterology, Skåne University Hospital, Malmö, Lund University, Sweden
- 10 Department of Gastroenterology and Digestive Endoscopy, Georges-Pompidou European Hospital, Paris, France
- 11 Gastroenterology Unit, Valduce Hospital, Como, Italy
- 12 Gastroenterology and Endoscopy Unit, Edouard Herriot Hospital, Lyon, France
- 13 Gastroenterology & Endoscopy Unit, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy

submitted 7.5.2018

accepted after revision 9.8.2018

Bibliography

DOI <https://doi.org/10.1055/a-0761-9742> |

Endoscopy International Open 2019; 07: E372-E379

© Georg Thieme Verlag KG Stuttgart · New York

ISSN 2364-3722

Corresponding author

Leenhardt Romain, Hopital Saint-Antoine - Gastroenterology and Endoscopic Department, 184 Rue du Faubourg Saint-Antoine, Paris 75012, France
Phone: +0033149282000
romain.leenhardt@gmail.com

ABSTRACT

Background and study aims Nomenclature and descriptions of small bowel (SB) vascular lesions in capsule endoscopy (CE) are scarce in the medical literature. They are mostly based on the reader's opinion and thus differ between experts, with a potential negative impact on clinical care, teaching and research regarding SBCE. Our aim was to better define a nomenclature and to give a description of the most frequent vascular lesions in SBCE.

Methods A panel of 18 European expert SBCE readers was formed during the UEGW 2016 meeting. Three experts constructed an Internet-based four-round Delphi consensus, but did not participate in the voting process. They built questionnaires that included various still frames of vascular lesions obtained with a third-generation SBCE system. The 15 remaining participants were asked to rate different proposals and description of the most common SB vascular lesions. A 6-point rating scale (varying from 'strongly disagree' to 'strongly agree') was used successive rounds. The consensus was reached when at least 80 % voting members scored the statement within the 'agree' or 'strongly agree'. **Results** Consensual terms and descriptions were reached for angiectasia/angiodysplasia, erythematous patch, red spot/dot, and phlebectasia. A consensual description was reached for more subtle vascular lesions tentatively named "diminutive angiectasia" but no consensus was reached for this term.

Conclusion An international group has reached a consensus on the nomenclature and descriptions of the most frequent and relevant SB vascular lesions in CE. These terms and descriptions are useful in daily practice, for teaching and for medical research purposes.

* Meeting presentations: 25th United European Gastroenterology Week 2017

Introduction

Capsule endoscopy (CE) is a minimally invasive technique that allows complete visualization of the small bowel (SB) [1]. During its development in the 1990s, CE was immediately perceived as a disruptive technology. Since its clinical inception in 2000, CE has rapidly become the first-line diagnostic tool to explore the SB, with a diagnostic yield of approximately 60 % in the setting of obscure gastrointestinal bleeding (OGIB), and 50 % in suspected Crohn's disease [2].

Because CE offers enhanced imaging of the SB mucosa, an important effort has been made during the first few years of its clinical use to design a structured lexicon for use in training, clinical practice, and for research purposes [3 - 5]. The CE Structured Terminology (CEST) emerged through this work, but it did not provide definitions nor full descriptions of the most common SB lesions. As a matter of fact, no clear definitions/descriptions of vascular lesions have been widely accepted yet. A typical example is "angiodysplasia" (AGD) that remains the most common vascular and obvious lesion found in SBCE in the setting of OGIB [6]; however, there are controversies regarding the name, description, and interobserver agreement of this diagnosis. Some experts argue that the suffix "dysplasia" is inappropriate from a pathological point of view, and that the term "ectasia" should be preferred [7]. Nevertheless, in clinical practice, more than half of reports use other terms, such as, "angioma," "ectasia," "arteriovenous malformation," "vascular ectasia," or "vascular malformation" [4].

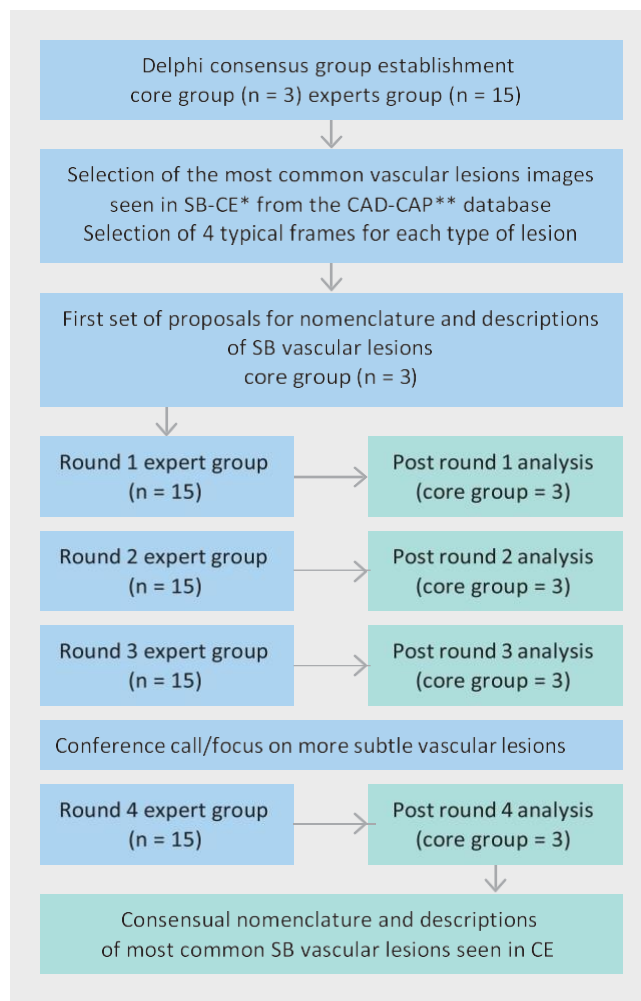
In the medical literature, semantic descriptions of AGD are scarce [6, 8], and are even more rare for more subtle vascular lesions. In addition, in most studies aiming to evaluate interobserver in describing CE findings, the "gold standard" for diagnosis is based on the experience of expert investigators, without any description or definition of vascular lesions [9, 10]. Overall, names, descriptions, and diagnosis of vascular SB lesions remain subjective. Still, when a putative diagnosis of SB vascular lesion is made with CE, ablation therapy is first-line treatment in most cases, pharmacological treatments are discussed, but endoscopic sampling or surgery followed by pathological examination are rarely performed to confirm the diagnosis [6]. Overall, interobserver agreement for all lesions seen in CE is moderate (kappa coefficient = 0.6) in the medical literature [10], and seems even lower for more subtle lesions.

Universal nomenclature and descriptions of SB vascular are needed for CE reporting [11], medical education, and research. The aim of this study was to achieve consensus on the nomenclature and descriptions of SB vascular lesions seen in CE.

Materials and methods

Design

The consensus process used a Delphi method, an established method for obtaining answers to questions that cause uncertainty even between experts [12, 13]. The study design was based on alternative feedback rounds of a core group (CG) and an expert group (EG) until consensus was achieved (► Fig. 1). Several Internet-based illustrated questionnaires were thus



► Fig. 1 Delphi Consensus Study flowchart. *SB-CE: Small bowel capsule endoscopy, **CAD-CAP: Computer-assisted Diagnosis for Capsule endoscopy [14].

from the CG to the EG for rating and comments. All questionnaires were structured in a similar way, with proposals for a nomenclature and a semantic description, and calling for free comments. Nomenclature and descriptions were revised by the CG following each feedback round based on rating and of the EG.

Working groups and feedback rounds

Working groups were formed aside the 24th United European Gastroenterology Week (October 15 - 19, 2016, Vienna, Austria). The three investigators of the CG worked in Sorbonne University & APHP, Saint Antoine Hospital, Paris, France. The 15 international investigators of the EG were based in England (n = 1), France (n = 3), Israel (n = 2), Italy (n = 3), Scotland (n = 3), Spain (n = 1) and Sweden (n = 2). Mean age of the EG was 47 years. Mean CE reading experience of the EG was 13 years, with a mean number of 100 CEs read annually.

► Table 1 Six-point numerical scale to rate proposals during the 4th round.

1	Strongly disagree
2	Disagree
3	Moderately disagree
4	Moderately agree
5	Agree
6	Strongly agree

Proposals and revisions of names and semantic descriptions

The multicenter French database CAD-CAP (Computer-assisted Diagnosis for Capsule endoscopy) was used. CAD-CAP is composed of 20,000 normal CE still frames and 6013 still frames of abnormal CE findings from third-generation SBCE system videos (Pillcam SB3 system, Medtronic), including 2946 vascular lesions [14]. The CG selected an initial set of 100 still frames of most common vascular lesion images seen in SBCE. Based on this image dataset and on their knowledge, the members of the CG categorized 5 different types of SB vascular lesions. For each type of SB vascular lesion, the CG selected 4 typical frames (with adequate mucosal visualization and brightness) and proposed a putative name and a semantic description, to be sent to the EG through an Internet-based questionnaire. At each round, and for each type of lesion, members of the EG were asked to rate proposals of the CG, on a numerical scale from 1

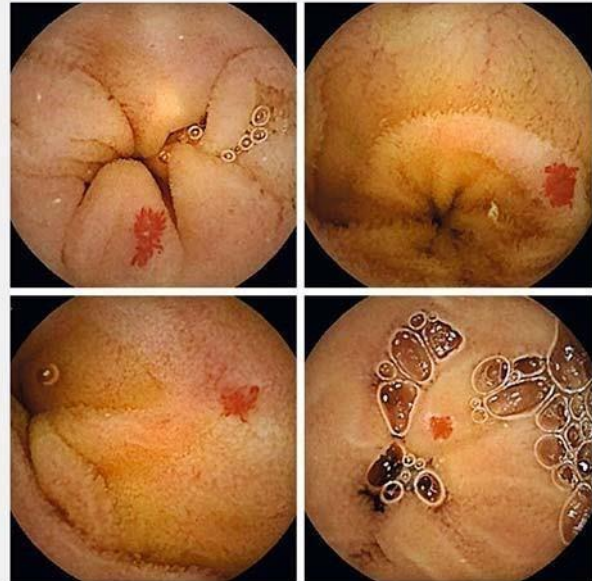
(strongly disagree) to 6 (strongly agree) (► Table 1), and to provide comments and suggestions for substantive revisions. After each round, all suggestions from EG respondents were used by the CG to improve names/descriptions.

Analysis

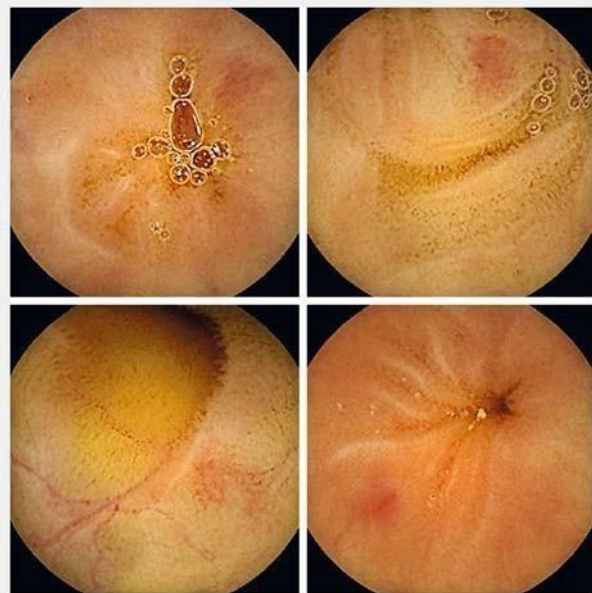
Consensus was reached when at least 80 % of voting members rated with a score of 5 (agree) or 6 (strongly agree) within four rounds. Rounds were stopped for any lesion for which consensual name and description were found. For any other lesion, votes and comments were analyzed and discussed during several face-to-face meetings by the three participants of the CG in between the three first rounds. When consensual name and description were not achieved after the third round, a teleconference was held between the CG and the EG. Then, the CG made final revisions and provided a fourth questionnaire to the EG, for a final vote. The process was stopped thereafter, whether a consensus was reached or not.

Results

The CG selected five types of common SB vascular lesions from the CAD-CAP database. For each type of vascular lesions, a set of 4 still frames (► Fig. 2, ► Fig. 3, ► Fig. 4, ► Fig. 5, ► Fig. 6) was sent, using Internet-based questionnaires, to the EG with tentative names and descriptions. Further feedback rounds between



► Fig. 2 Four typical frames of “angiectasia/angiodysplasia,” as presented to the expert group for voting.



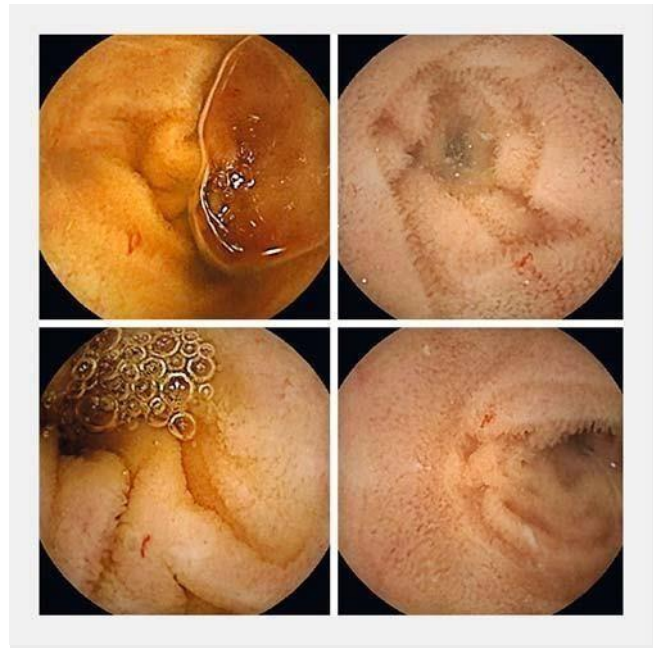
► Fig. 3 Four typical frames of “erythematous patch,” as presented to the expert group for voting.

the CG and the EG were scheduled until consensus was achieved or until round 4 was reached. Consensual names and

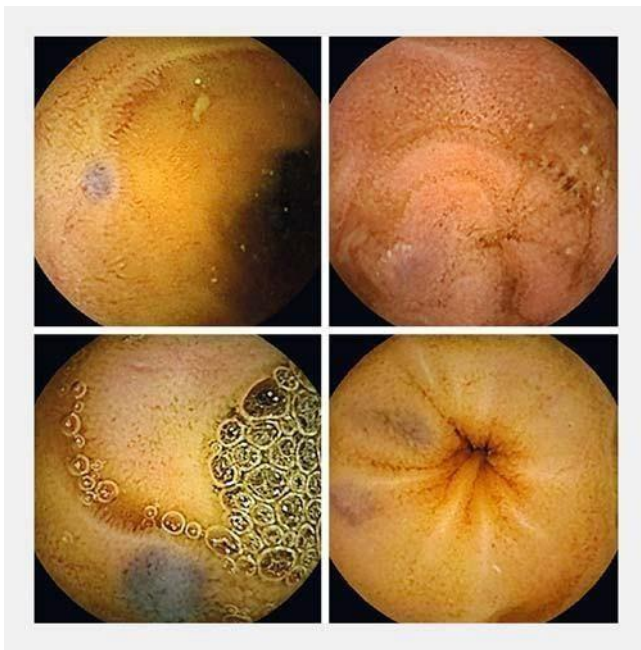
definitions of “angiectasia/angiodysplasia” (► Fig. 2), “erythematous patch” (► Fig. 3), “red spot/dot” (► Fig. 4), and “phlebectasia” (► Fig. 5) were achieved within 3 rounds (► Table 2). However, a conference call and a final fourth round did not result in consensus on the name of more subtle vascular mucosal abnormalities, tentatively named “diminutive angiectasia”



► Fig. 4 Four typical frames of “red spot/dot,” as presented to the expert group for voting.



► Fig. 6 Four typical frames of “diminutive angiectasia,” as presented to the expert group for voting.



► Fig. 5 Four typical frames of “phlebectasia,” as presented to the expert group for voting.

(► Fig. 6), although a consensus was reached on their semantic description. Numbers of voting experts for names and descriptions are described more precisely for the last round in

► Table 3 and ► Table 4. Details on the votes for nomenclatures are given in ► Supplementary Table 1.

Angiectasia/angiodyplasia

Regarding ► Fig. 2, the terminology voted for after one round with a 100.0 % rate of agreement was an “angiodyplasia.” A 93.3 % rate of agreement was found after Round 3 on the following description: “a clearly demarcated, bright-red, flat lesion, consisting of tortuous and clustered capillary dilatations within the mucosal layer (surrounded by intestinal villi). Can be small (few mm) to large (few cm).” However, there were controversies regarding the name “angiodyplasia” versus “angiectasia.” During post-Round 1 analysis, five experts mentioned that they would have preferred the term “angiectasia.” Thus, the core group decided to put the question to the vote during the second round. After Round 2, no agreement was found as 60.0 % of the experts preferred the term “angiectasia,” whereas 40.0 % preferred the term “angiodyplasia.”

Erythematous patch

Regarding ► Fig. 3, the terminology voted for after Round 2 with an 86.6 % rate of agreement was an “erythematous patch.” An 80.0 % rate of agreement was found on the following description: “small (few mm) and flat reddish area, without any vessel appearance within the mucosal layer (surrounded by intestinal villi).”

Red spot/dot

Regarding ► Fig. 4, the terminology voted for after Round 1 with 93.3 % rate of agreement was a “Red spot/dot.” An 80.0 % rate of agreement was found on the following description: “a miniscule (less than 1 mm), punctuate, and flat lesion with a bright-red area, without linear or vessel appearance within the mucosal layer (surrounded by intestinal villi).” However, there were controversies regarding the term “red dot” versus “red spot” and

► Table 2 International Delphi Consensus on the nomenclature and descriptions of the most frequent small bowel vascular lesions.

Nomenclature	Semantic Description	Nomenclature/Description % of agreement or strong agreement
Angiectasia/angiodysplasia	A clearly demarcated, bright-red, flat lesion, consisting of and clustered capillary dilatations, within the mucosal layer (surrounded by intestinal villi). Can be small (few mm) to large (few cm).	tortuous 100%/93 %
Erythematous patch	A small (few mm) and flat reddish area, without any vessel appearance, within the mucosal layer (surrounded by intestinal villi).	87%/80 %
Red spot/dot	A miniscule (less than 1 mm), punctuate, and flat lesion with a bright-red area, without linear or vessel appearance, within the mucosal layer (surrounded by intestinal villi).	93%/80 %
Phlebectasia	A small (few mm), flat to slightly elevated, bluish venous dilatation running below the mucosa (covered by intestinal villi).	93%/87 %
Diminutive angiectasia*	A clearly demarcated, linear, bright-red lesion, consisting of tiny non-clustered capillary dilatations, within the mucosal layer (surrounded by intestinal villi).	73%/87 %

* Consensus was not reached

► Table 3 Experts' rating regarding the nomenclature of the most frequent small bowel vascular lesions.

Nomenclature	Numerical scale/Number of voting expert						% of agreement or strong agreement	of rounds passed
	0	1	2	3	4	5		
Angiectasia/angiodysplasia	0	0	0	0	1	14	100 %	1
Erythematous Patch	1	0	1	0	3	10	87 %	2
Red spot/dot	1	0	0	1	4	9	93 %	1
Phlebectasia	0	0	0	1	3	11	93 %	1
"Diminutive angiectasia"	0	3	0	1	8	3	73 %*	4

* Consensus was not reached

► Table 4 Experts' rating regarding descriptions of the most frequent small bowel vascular lesions

Description	Numerical scale/Number of voting expert						% of agreement or strong agreement	of rounds passed
	0	1	2	3	4	5		
Angiectasia/angiodysplasia	0	0	0	1	5	9	93 %	3
Erythematous Patch	2	0	0	1	9	3	80 %	2
Red spot/dot	1	1	1	1	6	6	80 %	1
Phlebectasia	0	0	1	1	4	9	87 %	1
"Diminutive angiectasia"*	0	1	0	1	8	5	87 %	4

* Consensus was not reached

the CG decided to put this question to a vote. After Round 2, no agreement was found, as 53.3 % preferred the name “red spot,” whereas 46.7 % preferred the term “red dot.”

Phlebectasia

Regarding ► Fig. 5, the terminology voted for after Round 1 with a 93.3 % rate of agreement was a “phlebectasia.” An 86.6 % rate of agreement was found in Round 1 for the following description: “a small (few mm), flat to slightly-elevated, bluish venous dilatation running below the mucosa (covered by intestinal villi).”

‘Diminutive angiectasia’

Regarding ► Fig. 6, the terminology voted for after Round 4 with only 73.3 % rate of agreement was a “diminutive angiectasia.” This lesion was the most difficult to name and to describe. Several proposals for terminology were discussed during the three first rounds, among them “dilated capillaries,” “small angiodyplasia,” “telangiectasia,” “engorged capillaries,” “small arterial vascular malformation (AVM),” “atypical AVM,” and “atypical AGD.” Experts failed to achieve a consensus after Round 3. According to protocol, a conference call was set up, for names and descriptions to be debated before a final vote. At Round 4, there was an 86.6 % rate of agreement on the following description: “a clearly demarcated, linear, bright-red lesion, consisting of tiny non-clustered capillary dilatations, within the mucosal layer (surrounded by intestinal villi).”

Discussion

This study established nomenclature and descriptions for some of the most common vascular lesions seen in SB-CE. Through a Delphi consensus conducted with a significant number of CE expert readers, four names and their semantic descriptions were validated with more than an 80 % rate of agreement: “angiectasia/angiodyplasia” (93.3 %), “erythematous patch” (80.0 %),

“red spot/dot” (80.0 %), and “phlebectasia” (86.6 %) (► Table 2). Since the early 2000s, CE offers a new insight into SB pathology. Although an effort has been made initially to create a lexicon to describe SBCE findings [4], there has been no dictionary (with an entry name and a corresponding description) of SB vascular lesions since. This makes it difficult to attempt to standardize SBCE reading, reports, and research. For example, to date, most studies aiming to evaluate interobserver agreement in the description of SBCE findings have used the expert investigators’ opinion as gold standard for diagnosis [10], rather than any validated name or description of SB vascular lesions.

Consensus methods are being used increasingly to solve problems in medicine. Their main purpose is to define levels of agreement on controversial subjects. As a mean of achieving consensus on an issue, the Delphi methodology has become widely used in healthcare research [15, 16]. The literature on this technique is expanding, mainly addressing what it is and how it should be used [13, 17]. Given the need for a SB vascular lesions dictionary, we considered that Delphi methodology a good approach for large-scale research. However, Delphi methodology has some limitations. Therefore, we paid special

attention to clearly separate the CG from the EG, with different settings and different tasks. The CG was a small group of three physicians from the same center who had experience in SB-CE reading, teaching, and research. Members of the CG were able to attend consensus meetings and aimed to make proposals and revisions based on the EG votes and comments but they did not vote. The EG comprised 15 European opinion leaders in SB-CE. They were chosen based on their long experience (mean 13 years) and on their high volume of SB-CE reading (mean 100 cases a year). This study has other strengths. All Internet-based answers were anonymous, to allow the experts to express any contradictory view or even to change their mind during the voting rounds. The participation rate for all questionnaires reached 100 % in all rounds.

Some limitations should be mentioned as well. First, the CG of three physicians from the same institution made the initial image dataset and proposals for names and descriptions, possibly misleading the EG from the very start. Other names, terms, and definitions could have emerged from a larger or more heterogeneous group. Second, questionnaires and votes were Internet-based, except one conference call. It is likely that a conference call or physical meeting would have created more debate, controversy, and enrichment of vocabulary for names and descriptions than comments on a website without any live interaction. However, free comments were still possible through the Internet-based questionnaire, and at the time of the conference call after Round 3. Third, some rare, protruding or syndrome-related vascular lesions were not addressed in this study (hemangioma, varices, Kaposi sarcoma, blue rubber bleb syndrome, for example). We believe that there was an urge to build a consensus on the most frequent vascular lesions seen in SB-CE. However, many other less frequent vascular lesions will also need to be better defined. Fourth, some differences in descriptions of vascular lesions are based on the lesion size. Although a team has proposed measuring the size of SB polyps using reference granules in patients with Peutz-Jeghers syndrome [18], there is currently no validated system for size measurement in SBCE. Fifth, this consensus aimed to brush up the nomenclature and semantic descriptions of SB vascular lesions, not to provide their clinical relevance in terms of bleeding. Our group is now strongly considering doing another Delphi consensus on the significance of the SB vascular lesions named and described herein.

Some issues have not been fully solved by this study. Although their description was consensual, there were some controversies regarding the choice for a name between “angiodyplasia” and “angiectasia.” The core group decided to put that question to the vote during Round 2 but no agreement was found as 60.0 % of the experts preferred the term “angiectasia,” whereas 40.0 % preferred the term “angiodyplasia.” Thus, in the experts’ view, there would be a slight preference for the term “angiectasia” for daily practice. For some experts, the term “angiodyplasia” is widely used in clinical practice and in research, and refers adequately to the above-described vascular lesion. However, some other experts argue that the suffix “dysplasia” is inappropriate because the lesion is not an epithelial anomaly of growth and differentiation, which for patholo-

gists usually indicates a precancerous lesion. However, in Ancient Greek, “*dys*” means difficult and “*plasso*” means formed, and “angiodyplasia” would therefore etymologically mean “vascular malformation,” with no reference to precancerous change in cells and tissues [19]. Indeed, the term “angiodyplasia” is used in other benign conditions, such as hip, kidney or bone dysplasia. On the other hand, “*ektasis*” means “dilation” [19], and seems very appropriate to describe mucosal capillary dilations seen by pathologists in the rare cases in which such lesions are sampled. Overall, some slight confusion may still ensue with the different terms “angiectasia” (also sometimes written “angioectasia”) and “angiodyplasia,” although they relate to the very same type of lesion and may be considered as synonyms. Indeed, in the study validating the Capsule Endoscopy Standardized Terminology, the term “angiectasia” was used as an “obvious synonym” for the term “angiodyplasia” in 435 out of 486 occurrences for the experts [4]. For even more subtle lesions, such as so-called “diminutive angiectasia,” we were not able to find a consensus name (only a 73.3 % agreement was reached), although a consensus description was made.

Conclusion

In conclusion, the effort carried out by an international group through this consensus statement on nomenclature and descriptions of the most frequent SB vascular lesions represents a step forward in standardization of SB-CE reading, teaching, and medical research on this topic.

Acknowledgements

The Computer-Assisted Diagnosis for Capsule Endoscopy database is endorsed by the Société Française d’Endoscopie Digestive. It has been made possible with the support of the Société Nationale Française de Gastroentérologie and with the support of MSD France.

Competing interests

Dr. Dray has acted as a consultant for Boston Scientific, Fujifilm, Medtronic, and Pentax.

References

- [1] Iddan G, Meron G, Glukhovskiy A et al. Wireless capsule endoscopy. *Nature* 2000; 405: 417
- [2] Pennazio M, Spada C, Eliakim R et al. Small-bowel capsule endoscopy and device-assisted enteroscopy for diagnosis and treatment of small-bowel disorders: European Society of Gastrointestinal Endoscopy (ESGE) Clinical Guideline. *Endoscopy* 2015; 47: 352 - 376
- [3] Korman LY, Delvaux M, Gay G et al. Capsule endoscopy structured terminology (CEST): proposal of a standardized and structured terminology for reporting capsule endoscopy procedures. *Endoscopy* 2005; 37: 951 - 959
- [4] Delvaux M, Friedman S, Keuchel M et al. Structured terminology for capsule endoscopy: results of retrospective testing and validation in 766 small-bowel investigations. *Endoscopy* 2005; 37: 945 - 950
- [5] Aabakken L, Rembacken B, LeMoine O et al. Minimal standard terminology for gastrointestinal endoscopy - MST 3.0. *Endoscopy* 2009; 41: 727 - 728
- [6] Becq A, Rahmi G, Perrod G et al. Hemorrhagic angiodysplasia of the digestive tract: pathogenesis, diagnosis, and management. *Gastrointest Endosc* 2017; 86: 792 - 806
- [7] Wassef M, Blei F, Adams D et al. Vascular anomalies classification: recommendations from the International Society for the Study of Vascular Anomalies. *Pediatrics* 2015; 136: e203 - 214
- [8] Yano T, Yamamoto H, Sunada K et al. Endoscopic classification of vascular lesions of the small intestine (with videos). *Gastrointest Endosc* 2008; 67: 169 - 172
- [9] Pezzoli A, Cannizzaro R, Pennazio M et al. Interobserver agreement in describing video capsule endoscopy findings: a multicentre prospective study. *Dig Liver Dis* 2011; 43: 126 - 131
- [10] Jang BI, Lee SH, Moon J-S et al. Inter-observer agreement on the interpretation of capsule endoscopy findings based on capsule endoscopy structured terminology: a multicenter study by the Korean Gut Image Study Group. *Scand J Gastroenterol* 2010; 45: 370 - 374
- [11] Micic D, Semrad CE. Small bowel endoscopy. *Curr Treat Options Gastroenterol* 2016; 14: 220 - 235
- [12] Dalkey N, Helmer O. An experimental application of the DELPHI Method to the use of experts. *Management Science* 1963; 9: 458 - 467
- [13] Hsu C-C, Sandford BA. The Delphi technique: making sense of consensus. *Pract Assess Res Eval* 2007; 12: 1 - 8
- [14] Leenhardt R, Li C, Le Mouel JP et al. CAD-CAP: a 26000 images database serving the development of artificial intelligence for capsule endoscopy. *United European Gastroenterology Week, Vienna, 22-24 oct 2018*
- [15] Katragadda C, Finnane A, Soyer HP. Technique standards for skin lesion imaging: a delphi consensus statement. *JAMA Dermatol* 2017; 153: 207 - 213
- [16] Collins S, Yen P-Y, Phillips A et al. Nursing informatics competency assessment for the nurse leader: The Delphi Study. *J Nurs Adm* 2017; 47: 212 - 218
- [17] Okoli C, Pawlowski SD. The Delphi method as a research tool: an example, design considerations and applications. *Info Manag* 2004; 42: 15 - 29
- [18] Rácz I, Jánoki M, Kovács V. Measurement of small-bowel polyp size in patients with Peutz-Jeghers syndrome by using reference granules during video capsule endoscopy. *Endoscopy* 2007; 1: E41
- [19] Brandt LJ. Terminology for vascular lesions of the GI tract. *Gastrointest Endosc* 2018; 87: 1595 - 1596, comment to Becq A, et al. Terminology for vascular lesions of the GI tract. *Gastrointest Endosc* 2018; 87: 1596

► Supplementary Table 1 Proposed nomenclature/agreement for five different types of small bowel vascular lesions.

Vascular lesions images	Round 1 Proposed term % of agreement	Round 2 Proposed term % of agreement	Round 3 Proposed term % of agreement	Round 4 Proposed term % of agreement
► Figure 2	Angiodysplasia 100.0	Angiodysplasia 40.0 vs. Angiectasia 60.0	/	/
► Figure 3	Red spot 53.3	Erythematous patch 86.6	/	/
► Figure 4	Red dot 93.3	Red dot 46.7 vs. Red spot 53.3	/	/
► Figure 5	Phlebectasia 93.3	/	/	/
► Figure 6	Telangiectasia 66.7	Angiectasia 66.7	Small angiodysplasia 53.3 Atypical angiodysplasia 0.0 Engorged capillaries 20.0 Small AVM* 20.0 Atypical AVM 13.3 Unclassified small-bowel subtle vascular lesions 20.0 Undetermined small-bowel subtle vascular lesions 33.3	Diminutive angiectasia 73.3 Dilated capillaries 26.7

* AVM=Arterial vascular malformation

Nomenclature and semantic descriptions of ulcerative and inflammatory lesions seen in Crohn's disease in small bowel capsule endoscopy: An international Delphi consensus statement

Romain Leenhardt¹, Anthony Buisson², Arnaud Bourreille³, Philippe Marteau¹, Anastasios Koulaouzidis⁴, Cynthia Li^{1,5}, Martin Keuchel⁶, Emmanuele Rondonotti⁷, Ervin Toth⁸, John N Plevris⁴, Rami Eliakim⁹, Bruno Rosa¹⁰, Konstantinos Triantafyllou¹¹, Luca Elli¹², Gabriele Wurm Johansson⁸, Simon Panter¹³, Pierre Ellul¹⁴, Enrique Pérez-Cuadrado Robles^{15,25}, Deirdre McNamara¹⁶, Hanneke Beaumont¹⁷, Cristiano Spada¹⁸, Flaminia Cavallaro¹⁹, Franck Cholet²⁰, Ignacio Fernandez-Urien Sainz²¹, Uri Kopylov⁹, Mark E McAlindon²², Artur Németh⁸, Gian Eugenio Tontini²³, Diana E Yung⁴, Yaron Niv²⁴, Gabriel Rahmi²⁵, Jean-Christophe Saurin²⁶ and Xavier Dray¹

Abstract

Background: In the medical literature, the nomenclature and descriptions (ND) of small bowel (SB) ulcerative and inflammatory (U-I) lesions in capsule endoscopy (CE) are scarce and inconsistent. Inter-observer variability in interpreting these

¹Sorbonne Université, Endoscopy Unit, Hôpital Saint-Antoine, AP-HP, Paris, France

²Dept. of Gastroenterology, CHU Estaing Clermont-Ferrand, Clermont-ferrand, France

³Institut des Maladies de l'Appareil Digestif (IMAD), Dept of Gastroenterology, CHU Nantes, University of Nantes, Nantes, France

⁴Centre For Liver & Digestive Disorders, The Royal Infirmary of Edinburgh, Edinburgh, United Kingdom

⁵College of Arts & Sciences, Drexel University, Philadelphia, USA

⁶Klinik für Innere Medizin, Bethesda Krankenhaus Bergedorf, Hamburg, Germany

⁷Gastroenterology Unit, Valduce Hospital, Como, Italy

⁸Department of Gastroenterology, Skåne University Hospital, Lund University, Malmö, Sweden

⁹Dept. of Gastroenterology, Sheba Medical Center, Ramat Gan, and Sackler School of Medicine, Tel-Aviv University, Tel-Aviv, Israel

¹⁰Departamento de Gastroenterologia, Universidade do Minho, Hospital Senhora da Oliveira, Guimarães, Portugal

¹¹Hepatogastroenterology Unit, Second Department of Internal Medicine - Propaedeutic Research Institute and Diabetes Center, National and Kapodistrian University of Athens, Medical School, Attikon University General Hospital, Athens, Greece

¹²Center for Prevention and Diagnosis of Celiac Disease, Fondazione IRCCS Ca Granda, Milano, Italy

¹³Gastroenterology, South Tyneside Hospital, South Shields, United Kingdom

¹⁴Department of Medicine, Mater Dei Hospital, Msida, Malta

¹⁵Department of Gastroenterology, Cliniques Universitaires Saint-Luc, Bruxelles, Belgium

¹⁶Trinity Academic Gastroenterology Group, Departement of Clinical Medicine, Tallaght Hospital, Trinity College Dublin, Ireland

¹⁷Department of Gastroenterology, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

¹⁸Digestive Endoscopy Unit and Gastroenterology, Fondazione Poliambulanza, Brescia, Italia; Digestive Endoscopy Unit, Università Cattolica del Sacro Cuore, Roma, Italia

¹⁹Gastroenterology and Endoscopy Unit, IRCCS Policlinico San Donato, San Donato Milanese, Milan, Italy

²⁰Endoscopy unit, CHU La Cavale Blanche, Brest, France

²¹Gastroenterology, Hospital de Navarra, Pamplona, Spain

²²Dept. of Gastroenterology, Royal Hallamshire Hospital, Sheffield, United Kingdom

²³Department of Pathophysiology and Transplantation, University of Milan - Gastroenterology and Endoscopy Unit, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy

²⁴Rabin Medical Center, Dept. of Gastroenterology, Petach Tikva, Israel, and Sackler School of Medicine, Tel-Aviv University, Tel-Aviv, Israel

²⁵Department of Gastroenterology and Digestive Endoscopy, Georges-Pompidou European hospital, Paris, France

²⁶Gastroenterology and Endoscopy Unit, Edouard Herriot Hospital, Lyon, France

Corresponding author:

Xavier Dray, Hôpital Saint-Antoine, 184 rue du Faubourg Saint-Antoine Paris, 75012 France.

Email: xavier.dray@aphp.fr

findings remains a major limitation in the assessment of the severity of mucosal lesions, which can impact negatively on clinical care, training and research on SB-CE.

Objective: Focusing on SB-CE in Crohn's disease (CD), our aim is to establish a consensus on the ND of U-I lesions.

Methods: An international panel of experienced SB-CE readers was formed during the 2016 United European Gastroenterology Week meeting. A core group of five CE and inflammatory bowel disease (IBD) experts established an Internet-based, three-round Delphi consensus but did not participate in the voting process. The core group built illustrated questionnaires, including SB-CE still frames of U-I lesions from patients with documented CD. Twenty-seven other experts were asked to rate and comment on the different proposals for the ND of the most frequent SB U-I lesions. For each round, we used a 6-point rating scale (varying from 'strongly disagree' to 'strongly agree'). The consensus was reached when at least 80% of the voting members scored the statement within the 'agree' or 'strongly agree' categories.

Results: A 100% participation rate was obtained for all the rounds. Consensual ND were reached for the following seven U-I lesions: *aphthoid erosion, deep ulceration, superficial ulceration, stenosis, edema, hyperemia* and *denudation*.

Conclusion: Considering the most frequent SB U-I lesions seen in CE in CD, a consensual ND was reached by the international group of experts. These descriptions and names are useful not only for daily practice and medical education, but also for medical research.

Keywords

Capsule endoscopy, small-bowel, Crohn's disease, Delphi consensus, inflammatory bowel diseases, nomenclature

Received: 31 August 2019; accepted: 24 November 2019

Introduction

Considering its high ability to visualise the entire small bowel (SB) mucosa, capsule endoscopy (CE) has rapidly been considered as one of the first-line diagnostic tools for SB examination.¹ CE has an approximate diagnostic yield of 50% in suspected Crohn's disease (CD) cases² and plays a significant role in its management.^{3,4} The Capsule Endoscopy Crohn's Disease Activity Index (CECDAI) was developed⁵ and validated⁶ to grade the severity of SB-CE findings in patients with CD. The Lewis score also provides a common language to quantify mucosal inflammatory variations.⁷ Both scores are commonly used both in clinical practice and in research because they display a strong correlation.⁸ However, the nomenclature and descriptions (ND) of ulcerative and inflammatory (U-I) lesions seen in CE and that are used in these scores are far from established. For instance, in the study by Gralnek et al., the 'definitions of endoscopic lesions were a priori specified and reviewed by all readers prior to prospective testing', yet no definitions were consensually agreed and provided.⁷ As CE reviewing is strongly related to the reader's opinion and experience, which differ among readers,^{9,10} it is crucial to search for a common international ND to define the most common U-I lesions seen on SB-CE in patients with CD.

Indeed, in the medical literature, studies assessing the inter-observer variability have measured a mean kappa value of around 0.5.¹¹ We believe that the inter-observer variability in interpreting the findings may be (at least partly) due to the lack of a common language, and that it

is a major limitation for assessing the severity of mucosal lesions observed in inflammatory bowel diseases (IBD), with a potential negative impact on patients' healthcare and on the evaluation of therapeutics. Therefore, it seems necessary to standardise and use a unified language for all major U-I lesions seen during SB-CE in IBD. A similar approach has been used by our group through the establishment of an international consensus on the ND of the most frequent vascular lesions seen in SB-CE.¹² Hence, the aim of the current study was to establish a consensual ND of the most common U-I lesions seen on SB-CE in patients with CD.

Patients, materials and methods

Design

The Delphi method is one of the most accurate and validated process for answering questions that cause uncertainty even among experts.^{13,14} The present study used a Delphi consensus model based on alternative feedback rounds between a core group (CG) and an expert group (EG) until a consensus was reached (Figure 1). The CG built several Internet-based illustrated questionnaires and asked the EG to rate the descriptions and add comments to inform the next round. The questionnaires were structured similarly, presenting images related to several propositions for the nomenclature and the semantic descriptions of U-I lesions seen in CD patients. All members of the EG were invited to add comments for each proposal. Based on the rating and on the comments received from the

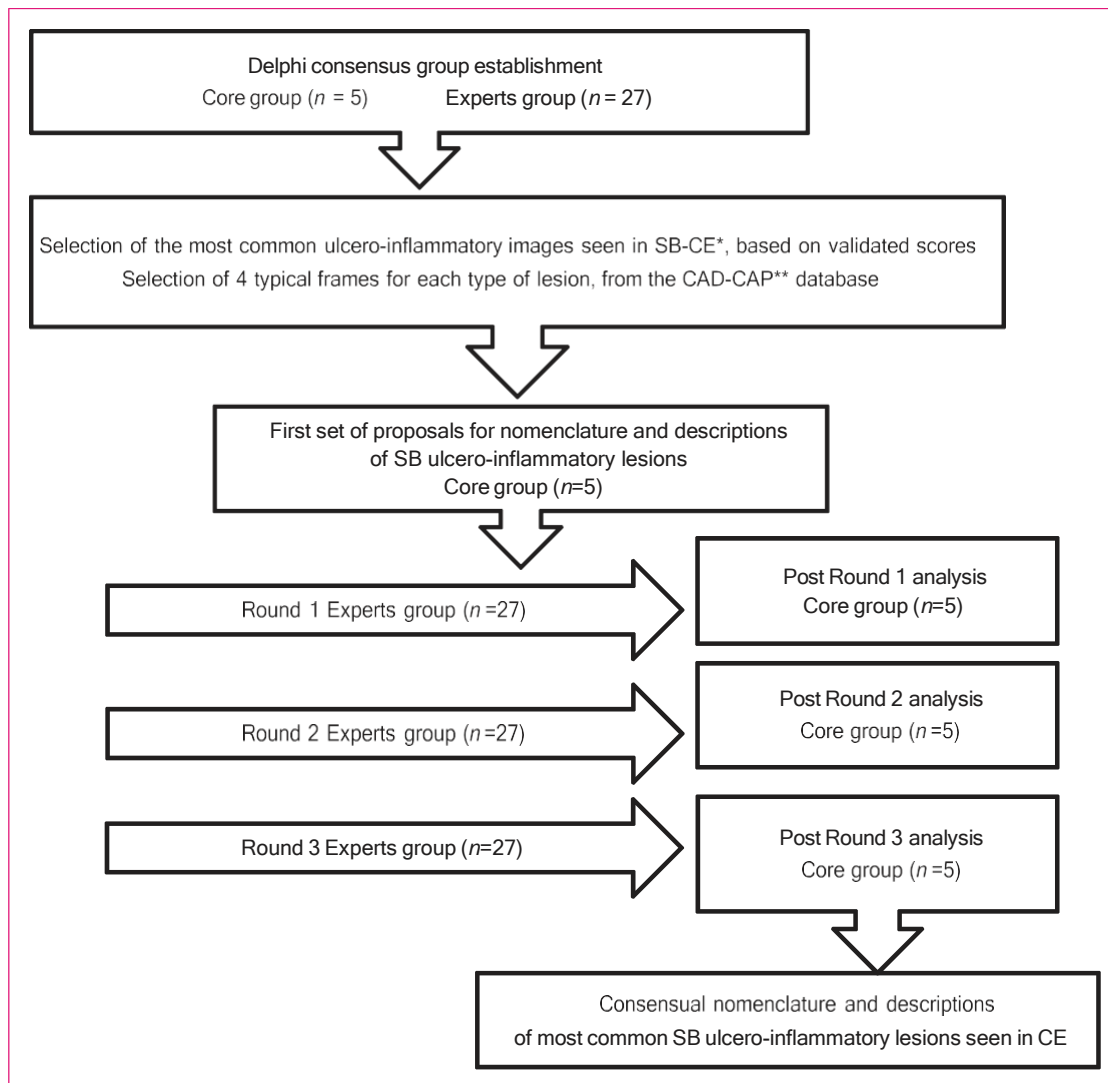


Figure 1. Delphi consensus study flowchart. *SB-CE: small bowel capsule endoscopy. **CAD-CAP: Computer-aided Diagnosis for capsule endoscopy database (19).

experts, the name and the description of each lesion were then reviewed by the CG after each feedback round.

Working groups and feedback rounds

The working groups were constituted during the 24th United European Gastroenterology Week (15–19 October 2016, Vienna, Austria). The CG was composed of five French IBD specialists, with three of them also being SB-CE expert readers. The CG members did not participate in the EG. The 27 members of the international EG were based in England ($n = 2$), France ($n = 3$), Germany ($n = 1$), Greece ($n = 1$), Ireland ($n = 1$), Israel ($n = 3$), Italy ($n = 5$), Malta ($n = 1$), the Netherlands ($n = 1$), Portugal ($n = 1$), Scotland ($n = 3$), Spain ($n = 2$) and Sweden ($n = 3$). The mean age of the EG members was 47.6 years.

The mean CE reading experience of the EG members was 12.4 ± 4.6 years, with a mean number of 136 ± 85 CEs read annually.

Proposals and revisions of nomenclature and semantic descriptions

To design the first round, the CG first reviewed relevant U-I terms described in the SB-CE IBD scores^{5,7} or in the papers of the digestive endoscopic terminology.^{15–17} Moreover, to avoid confusion, the CG also considered the terms and definitions of U-I lesions found during colonoscopy of IBD patients, as proposed by *Groupe d'Etudes Thérapeutiques des Affections Inflammatoires du Tube Digestif* (GETAID).¹⁸

The CG proposed a selection of four typical SB-CE image frames, a nomenclature and a semantic

description per specific type of SB U-I lesion. All selected frames were extracted from the multicentre French database CAD-CAP (computer-assisted diagnosis for capsule endoscopy). This database is composed of 20,000 normal CE still frames and 5184 still-frame images of abnormal CE findings from third-generation SB-CE system videos (Pillcam®SB3 system, Medtronic), including 1369 images with U-I lesions from patients with IBD.¹⁹ It has already been used to standardise and illustrate a common lexicon for vascular lesions seen in SB-CE¹² and to develop CAD tools for capsule reading.²⁰ An Internet-based survey was used to send all the propositions to the EG.

For each round, the EG members were invited to rate the propositions made for each type of lesion, using a numerical scale varying from 1 (strongly disagree) to 6 (strongly agree) (Table 1). They were also asked to provide comments and suggestions for subsequent revisions. To improve the lesions ND, the CG reviewed the EG's suggestions after each round.

Predefined stop criterion and analysis

We considered a consensual agreement was reached when over 80% of the voting members rated with a score of 5 (agree) or 6 (strongly agree) within three rounds. The Delphi consensus process was stopped when the consensual name and description of any lesion was established. Meetings were organised between the five participants of the CG within the first three rounds to discuss the votes and comments. Subsequently, whether a consensus was reached or not, the process was interrupted.

Results

The CG identified the seven following lesions to be proposed to the EG for definition and description: aphthoid erosion, deep ulceration, superficial ulceration, stenosis, edema, denudation and hyperemia. A 100% participation rate was obtained from the EG for all three rounds. The consensual ND, as finally obtained at the end of the Delphi process, are given in Table 2. The experts' ratings are detailed in Tables 3 and 4 for the nomenclature and descriptions, respectively.

Aphthoid erosion

Regarding Figure 2(a), the nomenclature voted on for this type of lesion after one round with an 85.2% agreement rate was 'aphthoid erosion'. After round 2, the description 'Diminutive loss of epithelial layering with a whitish centre and a red halo, surrounded by normal mucosa' obtained a 96.3% agreement rate.

Table 1. Six-point numerical scale to rate proposals during the Delphi rounds.

1	Strongly disagree
2	Disagree
3	Moderately disagree
4	Moderately agree
5	Agree
6	Strongly agree

Deep ulceration

Regarding Figure 2(b), the nomenclature 'deep ulceration' was voted on for this type of lesion after one round with a 96.3% rate of agreement. The description 'Frankly, deep loss of tissue compared with the surrounding swollen/edematous mucosa, with a whitish base' reached an 85.2% rate of agreement after round 3.

Superficial ulceration

Regarding Figure 2(c), the nomenclature 'superficial ulceration' was voted on for this type of lesion after one round with an 81.5% rate of agreement. The rate of agreement for the description 'Mildly depressed loss of tissue with a whitish bottom, whose features fit neither with that of aphthoid erosion nor with that of deep ulceration, as previously defined' was of 85.2% after round 2. Although referring to the descriptions of 'deep ulceration' and 'aphthoid erosion', a consensual description for 'superficial ulceration' was obtained before this, during round 2.

Stenosis

Regarding Figure 3(a), after one round the nomenclature 'stenosis' was voted on for this type of lesion with a 100.0% rate of agreement. During round 1 the EG asked for videos to further discuss the description. A 9-second video sequence (25 frames upstream, and 25 downstream the index still frame) was added to the questionnaire during round 2, as well as a special mention for the need of a dynamic sequence to establish a diagnosis. After round 2, the rate of agreement of the following description: 'Narrowing of the intestinal lumen withholding or delaying the passing of the video-capsule (therefore, to be evaluated on a video)' was of 88.9%.

Edema

After round 1, the EG asked for modifications in the selection of the SB-CE illustrations. Regarding

Table 2. International Delphi consensus on the nomenclature and descriptions of ulcerative and inflammatory lesions in SB-CE.

Nomenclature	Description	% Nomenclature/ description
Aphthoid erosion	Diminutive loss of epithelial layering with a whitish center and a red halo, surrounded by normal mucosa	85.2/96.3
Deep ulceration	Frankly deep loss of tissue compared to the surrounding swollen/edematous mucosa, with a whitish base	96.3/85.2
Superficial ulceration	Mildly depressed loss of tissue with a whitish bottom, whose features fit neither with that of aphthoid erosion nor with that of deep ulceration, as previously defined	81.5/85.2
Stenosis	Narrowing of the intestinal lumen withholding or delaying the passing of the videocapsule (therefore, to be evaluated on a video)	100.0/88.9
Edema	Enlarged / swollen / engorged villi	85.2/81.5
Hyperemia	Area of reddish villi	96.3/81.5
Denudation	Reddish (but not whitish) mucosal area where villi are absent	81.5/81.5

Table 3. Experts' rating regarding the nomenclature of the selected ulcerative and inflammatory SB lesions.

Nomenclature	Numerical scale / Number of expert voting						% of 'agree' or 'strongly agree'	Number of rounds passed
	1	2	3	4	5	6		
Aphthoid erosion	2	–	2	–	12	11	85.2	1
Deep ulceration	1	–	–	–	5	21	96.3	1
Superficial ulceration	–	–	2	3	11	11	81.5	1
Stenosis	–	–	–	–	9	18	100.0	1
Edema	–	1	1	2	12	11	85.2	2
Hyperemia	–	–	1	–	21	6	96.3	3
Denudation	–	1	–	4	10	12	81.5	1

Table 4. Experts' rating on the description of the selected ulcerative and inflammatory SB lesions.

Description	Numerical scale / Number of expert voting						% of 'agree' or 'strongly agree'	Number of rounds passed
	1	2	3	4	5	6		
Aphthoid erosion	–	–	–	1	13	13	96.3	2
Deep ulceration	–	1	–	3	14	9	85.2	3
Superficial ulceration	–	–	–	4	15	8	85.2	2
Stenosis	–	–	1	2	13	11	88.9	2
Edema	–	1	–	4	14	8	81.5	2
Hyperemia	–	1	1	3	14	8	81.5	3
Denudation	–	1	–	4	10	12	81.5	1

Figure 3(b), the nomenclature voted on for this type of lesion after two rounds was an 'edema' with an 85.2% rate of agreement. A rate of agreement of 81.5% was reached after round 2 for the following description: 'Enlarged/swollen/engorged villi'.

Hyperemia

A hyperemia is a nonspecific finding often associated with other common U-I lesions. During the first two rounds the EG asked the CG for modifications in the selection of the SB-CE images. Thus, two frames were



Figure 2. Frames with various types of loss of tissue. (a) Four typical frames of 'Aphthoid erosion'; (b) Four typical frames of 'Deep ulceration'; (c) Four typical frames of 'Superficial ulceration'.

removed from the initial selection. The regions of interest within the two remaining frames were delimited (Figure 3(d)). The nomenclature voted on for Figure 3(d) after three rounds was a 'hyperemia' with a 96.3% rate of agreement. An 81.5% rate of agreement was reached after round 3 for the description: 'Area of reddish villi'.

Denudation

Regarding Figure 3(c), an 81.5% rate of agreement was reached for the nomenclature 'denudation' after one round. After round 1 the following description:

'Reddish (but not whitish) mucosal area where villi are absent' obtained an 81.5% rate of agreement.

Discussion

In the current study, the ND of seven common U-I lesions seen in SB-CE in IBD patients were validated by experts who reached a more than 80% consensus rate. The agreements were reached using a Delphi consensus process conducted with 27 expert CE readers.

Since its clinical inception in 2000, CE has revolutionised the examination and description of the SB.

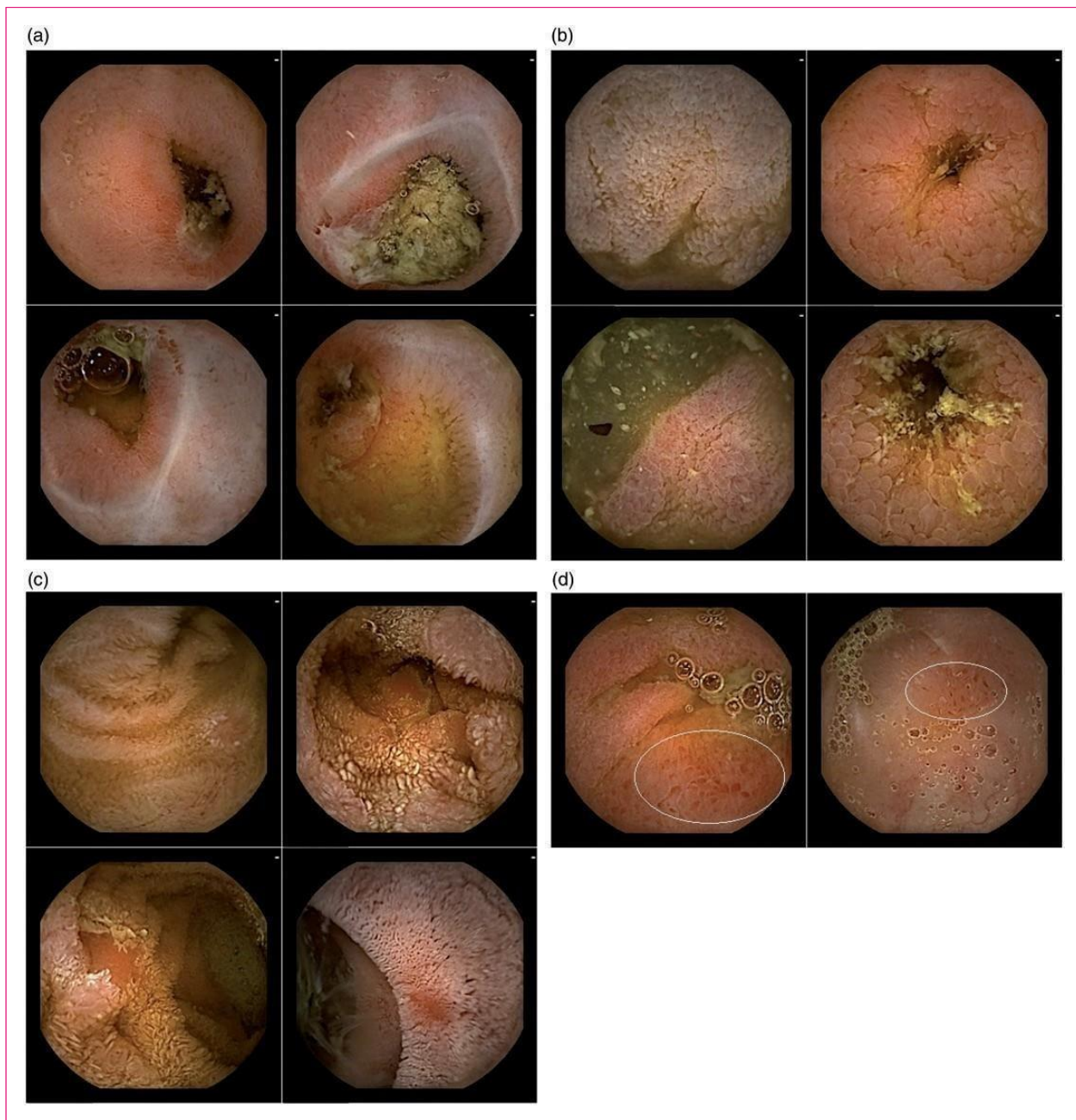


Figure 3. Frames with stenosis and with various types of inflammatory, non-ulcerated, lesions.

(a) Four typical frames of ‘Stenosis’; (b) Four typical frames of ‘Edema’; (c) Four typical frames of ‘Denudation’; (d) Two typical frames of ‘Hyperemia’.

Although a terminology index exists for SB-CE,¹⁵ a consensual ND of SB findings has not been clearly established. This lack of ND makes it difficult to standardise SB-CE reading, reporting and research. Indeed, in most studies evaluating the inter-observer agreement in the description of SB-CE, only the experts’ opinions are considered as the gold standard.

In an attempt to standardise the practice in SB-CE reading, our group first performed a Delphi consensus

to better define and describe vascular lesions found in SB-CE.¹² Similarly, we propose in the present study a consensus on the ND of U-I lesions seen in SB-CE in IBD patients. During the Delphi process, several important points were raised by the EG discussed subsequently by the CG. First, some U-I lesions are considered ‘nonspecific’, such as U-I lesions caused by the use of nonsteroidal anti-inflammatory drugs, by CD, by ischaemia or by other factors that cannot be easily or

categorically distinguished. Second, in SB-CE, as there is currently no available tool to precisely measure lesions, we could only use adjectives of size (e.g. 'diminutive') to describe the lesions instead of standardised units of measurement. Third, several types of U-I lesions are often found together in the same SB-CE frame, making it difficult to define their specific descriptions and illustrations. As an example, a consensus on the term 'hyperemia' could not be achieved within the two first rounds as this finding is nonspecific and often associated with other common U-I lesions. However, this issue was solved by delineating the regions of interest within the selected frames. Fourth, it is important to emphasise the difficulties in clearly distinguishing a superficial from a deep ulceration. Indeed, the evaluation of the depth of loss of tissue remains subjective and variable between experts, and a villous appearance contiguous to a mucosal break should be considered to appreciate and better characterise the entire lesion.

The present study has many strengths. First, the questionnaires were richly illustrated with high-quality, third-generation SB-CE images to better guide the experts. When required by the EG, still frames could be changed or delineated, and short video sequences could be specifically linked to the still images (e.g. when commenting on the term 'stenosis'). Such complementary information was provided when needed to increase the precision in the lesion to be described and to ensure a high reliability in the votes. Second, a substantial number of experts ($n = 27$) voted, they had a significant experience (mean 12.4 years) and performed a high volume (mean 136 cases per year) of SB-CE readings. Third, a 100% participation rate was obtained in all three rounds. Fourth, all the answers were anonymous, which encouraged open critique, while avoiding and preventing some experts from dominating in the process. Fifth, based on the EG's answers and comments, the members of the CG wrote out proposals and revisions, but they did not participate in the vote.

Some limitations should be acknowledged. First, almost all of the investigators were based in Europe. However, the 27 EG members were from 13 different countries, thus offering some meaningful diversity for such a study. Conversely, the five members of the CG were all affiliated with France, where the project was initiated, but they belonged to three different institutions. Second, the selection of the seven types of lesions to be studied was made by the CG based on their experience and on a review of the literature that had a special focus on the items of the CECDAI and the Lewis score.^{5,7} The EG was not specifically asked whether any other type of U-I lesion should be added to the list and voted for. For the seven above-mentioned different types of lesions, however, the EG

members could comment and ask for modifications. No EG expert requested adding an additional type of U-I to the list. Third, all images have been captured from the SB-CE Pillcam® SB3 system only. Therefore, it would be interesting to evaluate whether the ND described here are relevant with other SB-CE systems. Fourth, the current study represents a step forward in the standardisation of the description of CE lesions, and it does not assess the clinical relevance of such lesions according to specific clinical settings. Specifically, hyperemia and edema are subtle and debatable lesions. Their consensual descriptions by the experts were very short compared with that written for other lesions. These brief descriptions may indeed reflect the ideas of some IBD specialists that those findings may not be true manifestations or markers of CD, and that the validity of the Lewis Score is questionable for that reason.²¹ Fifth, the current study aimed to precisely describe lesions, but there is currently no available measurement tool in SB-CE, and no possibility to sample the lesions. Descriptive terms with adjectives (e.g. 'whitish/yellowish bottom', 'narrowing') rather than objective measurements (e.g. diameter, depth in mm ..) or pathology terms (e.g. fibrin bottom ..) were therefore used by both the CG and the EG. Timing was nevertheless available, and should have been quantified to precisely describe what the experts call 'withholding or delaying the passing of the video-capsule' when describing a stenosis. All in all, these descriptions may change when the SB-CE technology evolves and when readers call for even more precise and reproducible reports.

The proposed nomenclatures, atlas and descriptions of vascular and U-I lesions in SB-CE by our group pave the way to a study assessing their intra- and inter-observer reproducibility. We also plan to conduct survey on their clinical pertinence for the main indications of SB-CE (occult or overt gastrointestinal bleeding, and suspicion of CD).

Conclusion

An international group of SB-CE expert readers reached a consensus for the ND of the most frequent U-I lesions seen in CE. The current study has provided a strong asset for higher quality SB-CE reading and reporting in patients with CD. These names and descriptions will also be useful for medical education and for medical research purposes.

Declaration of conflicting interests

The authors disclose the following:

Romain Leenhardt is cofounder and shareholder of Augmented Endoscopy, and has given lectures for Abbvie.

Xavier Dray is cofounder and shareholder of Augmented

Endoscopy and has acted as a consultant for Alfasigma; Bouchara Recordati; Boston Scientific, Fujifilm, Medtronic, and Pentax.

Simon Panter has received support for a research fellow from Fuji and has acted as consultant for Medtronic.

Arnaud Bourreille has received consultant fees from Abbvie, Ferring, Janssen, Medtronic, MSD, Pfizer, Takeda, Tillotts, research grants from Abbvie, Takeda, Medtronic.

Uri Kopilov: has received consultant and speaker fees for Takeda, Abbvie, Janssen, Medtronic. Research support from Janssen, Medtronic, Takeda.

Rami Eliakim has received consultant fees from Abbvie, Takeda, Janssen and Medtronic; Research grant from Medtronic.

Funding

The Computer-Assisted Diagnosis for Capsule Endoscopy database has been made possible with the support of the Société Nationale Française de Gastroentérologie and with the support of MSD France.

Ethics approval

This Delphi consensus statement complies with the Declaration of Helsinki on research ethics.

Informed consent

De-identified images were extracted from the CAD-CAP database, a national multicenter database approved by the French Data Protection Authority.

ORCID iD

Enrique Pérez-Cuadrado Robles  <https://orcid.org/0000-0001-8254-7453>

References

- Iddan G, Meron G, Glukhovskiy A, et al. Wireless capsule endoscopy. *Nature* 2000; 405: 417.
- Pennazio M, Spada C, Eliakim R, et al. Small-bowel capsule endoscopy and device-assisted enteroscopy for diagnosis and treatment of small-bowel disorders: European Society of Gastrointestinal Endoscopy (ESGE) Clinical Guideline. *Endoscopy* 2015; 47: 352–376.
- Bourreille A, Ignjatovic A, Aabakken L, et al. Role of small-bowel endoscopy in the management of patients with inflammatory bowel disease: An international OMED- ECCO consensus. *Endoscopy* 2009; 41: 618–637.
- Enns RA, Hookey L, Armstrong D, et al. Clinical practice guidelines for the use of video capsule endoscopy. *Gastroenterology* 2017; 152: 497–514.
- Gal E, Geller A, Fraser G, et al. Assessment and validation of the new capsule endoscopy Crohn's disease activity index (CECDAI). *Dig Dis Sci* 2008; 53: 1933–1937.
- Niv Y, Ilani S, Levi Z, et al. Validation of the Capsule Endoscopy Crohn's Disease Activity Index (CECDAI or Niv score): A multicenter prospective study. *Endoscopy* 2012; 44:21–26.
- Gralnek IM, Defranchis R, Seidman E, et al. Development of a capsule endoscopy scoring index for small bowel mucosal inflammatory change. *Aliment Pharmacol Ther* 2008; 27: 146–154.
- Yablecovitch D, Lahat A, Neuman S, et al. The Lewis score or the capsule endoscopy Crohn's disease activity index: Which one is better for the assessment of small bowel inflammation in established Crohn's disease? *Ther Adv Gastroenterol* 2018; 11: 1756283X17747780.
- Rondonotti E, Soncini M, Girelli CM, et al. Can we improve the detection rate and interobserver agreement in capsule endoscopy? *Dig Liver Dis* 2012; 44: 1006–1011.
- Pezzoli A, Cannizzaro R, Pennazio M, et al. Interobserver agreement in describing video capsule endoscopy findings: A multicentre prospective study. *Dig Liver Dis* 2011; 43: 126–131.
- Mergener K, Ponchon T, Gralnek I, et al. Literature review and recommendations for clinical application of small-bowel capsule endoscopy, based on a panel discussion by international experts. *Endoscopy* 2007; 39: 895–909.
- Leenhardt R, Li C, Koulaouzidis A, et al. Nomenclature and semantic description of vascular lesions in small bowel capsule endoscopy: An international Delphi consensus statement. *Endosc Int Open* 2019; 7: E372–E379.
- Dalkey N and Helmer O. An experimental application of the DELPHI method to the use of experts. *Manag Sci* 1963; 9: 458–467.
- Hsu C-C and Sandford BA. The Delphi technique: Making sense of consensus. *Pract Assess Res Eval* 2007; 12: 1–8.
- Korman LY, Delvaux M, Gay G, et al. Capsule endoscopy structured terminology (CEST): Proposal of a standardized and structured terminology for reporting capsule endoscopy procedures. *Endoscopy* 2005; 37: 951–959.
- Aabakken L, Rembacken B, LeMoine O, et al. Minimal standard terminology for gastrointestinal endoscopy - MST 3.0. *Endoscopy* 2009; 41: 727–728.
- Delvaux M, Crespi M, Armengol-Miro JR, et al. Minimal standard terminology for digestive endoscopy: Results of prospective testing and validation in the GASTER project. *Endoscopy* 2000; 32: 345–355.
- Buisson A, Filippi J, Amiot A, et al. *Gastroenterology* 2015;148,S-445-6 (abstract).
- Leenhardt R, Li C, Le Mouel JP, et al. CAD-CAP: A 25000 images database serving the development of artificial intelligence for capsule endoscopy. *Endosc Int Open* (in press).
- Leenhardt R, Vasseur P, Li C, et al. A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. *Gastrointest Endosc* 2019; 89: 189–194.
- Omori T, Kambayashi H, Murasugi S, et al. Comparison of Lewis Score and Capsule Endoscopy Crohn's Disease Activity Index in patients with Crohn's disease. *Dig Dis Sci* 2019 (in pre

7.7.2. Imputabilité en vidéocapsule

Les différentes anomalies visualisées au cours d'une VCE-G peuvent avoir une imputabilité variable dépendant du contexte clinique. L'imputabilité attribuée à une anomalie détectée demeure subjective car dépendante de l'interprétation du lecteur de VCE. Un travail supplémentaire regroupant 15 experts a permis d'établir des abaques autour de la pertinence des lésions décrites précédemment en fonction du contexte clinique selon la classification de Saurin JC et al. (16,50). Ce travail apporte un peu plus à l'élaboration d'une vérité terrain nécessaire à l'apprentissage machine notamment dans la perspective d'une lecture assistée par l'IA qui selon le contexte pourrait stratifier l'imputabilité des anomalies détectées.



ELSEVIER

Available online at
ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com/en



ORIGINAL ARTICLE

A guide for assessing the clinical relevance of findings in small bowel capsule endoscopy: analysis of 8064 answers of international experts to an illustrated script questionnaire



R. Leenhardt^a, A. Koulaouzidis^b, D. McNamara^c, M. Keuchel^d,
 R. Sidhu^e, M.E. McAlindon^e, J.C. Saurin^f, R. Eliakim^g,
 I. Fernandez-Urien Sainz^h, J.N. Plevris^b, G. Rahmiⁱ,
 E. Rondonotti^j, B. Rosa^k, C. Spada^{l,m}, E. Tothⁿ, C. Houdeville^a,
 C. Li^{a,o}, M. Robaszekiewicz^p, P. Marteau^a, X. Dray^{a,*}

- ^a Sorbonne Université, Endoscopy Unit, Hôpital Saint-Antoine, APHP, Paris, France
^b The Royal Infirmary of Edinburgh, Centre For Liver & Digestive Disorders, Edinburgh, United Kingdom
^c TAGG Research Centre, Department of Clinical Medicine Tallaght Hospital, Trinity College Dublin, Ireland
^d Bethesda Krankenhaus Bergedorf, Klinik für Innere Medizin, Hamburg, Germany
^e Dept. of Gastroenterology, Royal Hallamshire Hospital, Sheffield, United Kingdom
^f Gastroenterology and Endoscopy Unit, Edouard Herriot Hospital, Lyon, France
^g Dept. of Gastroenterology, Sheba Medical Center, Ramat Gan, and Sackler School of Medicine, Tel-Aviv University, Tel-Aviv, Israel
^h Hospital de Navarra, Gastroenterology, Pamplona, Spain
ⁱ Université de Paris, Department of Gastroenterology and Digestive Endoscopy, Georges-Pompidou European Hospital, Paris, France
^j Gastroenterology Unit, Valduce Hospital, Como, Italy
^k Universidade do Minho, Hospital Senhora da Oliveira, Departamento de Gastroenterologia, Guimarães, Portugal
^l Digestive Endoscopy Unit and Gastroenterology, Fondazione Poliambulanza, Brescia, Italy
^m Digestive Endoscopy Unit, Università Cattolica del Sacro Cuore, Roma, Italy
ⁿ Department of Gastroenterology, Skåne University Hospital, Lund University, Malmö, Sweden
^o Drexel University, College of Arts & Sciences, Philadelphia, USA
^p La Cavale Blanche University Hospital, Endoscopy Unit, Brest, France

Available online 28 January 2021

* Corresponding author.
 E-mail address: xavier.dray@aphp.fr (X. Dray).

KEYWORDS

Capsule endoscopy;
Inflammatory bowel
disease;
Small bowel bleeding;
Small bowel
endoscopy

Abstract

Background and aim: Although recommended, the P-score used for assessing the pertinence / relevance of findings seen in small bowel (SB) capsule endoscopy (CE) is based on a low level of knowledge. The aim of this study was to evaluate the clinical relevance of the most frequent SBCE findings through an illustrated script questionnaire.

Materials and Methods: Sixteen types of SBCE findings were illustrated four times each in three different settings (occult and overt obscure gastrointestinal bleeding and suspected Crohn's

disease), and with a variable number ($n = 1/n = 2-5/n \geq 6$), thus providing a questionnaire with 192 scenarios and 576 illustrated questions. Fifteen international experts were asked to rate the finding's relevance for each question as very unlikely (-2) / unlikely (-1) / doubtful (0) / likely (+1) / very likely (+2). The median score (≤ -0.75 , between -0.75 and 0.75 , or ≥ 0.75) obtained for each scenario determined a low (P0), intermediate (P1) or high (P2) relevance, respectively.

Results: 8064 answers were analyzed. Participation and completion rates were 93% and 100%, respectively. In overt or occult OGIB, resultant P2 findings were 'typical angiectasia', 'deep ulceration', 'stenosis', and 'blood', whatever their numbers, and 'superficial ulcerations' when multiple. While in suspected CD, consensus P2 lesions were 'deep ulceration' and 'stenosis' whatever their numbers, and 'aphthoid erosions' and 'superficial ulcerations' when multiple.

Conclusion: This study establishes a guide for the evaluation of relevance of SBCE findings. It represents a step forward for SB-CE interpretation and is intended to be used as a tool for teaching and academic research.

© 2021 Published by Elsevier Masson SAS.

Introduction

Capsule endoscopy (CE) is of great help and recommended to investigate the small bowel (SB) in patients with obscure gastrointestinal bleeding (OGIB) or with suspected Crohn's disease, when gastroscopy and ileocolonoscopy are normal [1,2]. Diagnostic and prognostic evaluations are based on the reader's expertise as lesions found during the examination are usually not sampled by endoscopy or surgery for pathological assessment. Regarding diagnosis, the European Society for Gastrointestinal Endoscopy (ESGE) SB work-ing group has recently established consensual names and descriptions of the most frequent pathological findings seen in SBCE [3,4], in order to better standardize CE reading and teaching, and to guide research study protocols. Regarding prognosis, although misinterpretation of the clinical relevance of lesions can lead to inappropriate therapy [5], the medical literature is scarce. In the setting of OGIB, the ESGE suggests to use the Saurin's classification (P0-P1-P2, where P stands for "pertinence") [6] for the evaluation of the clinical relevance of the lesions [7]. According to this classification (the publication includes a short list of examples), P0 lesions have no clinical relevance, P1 lesions have an uncertain hemorrhagic potential, and P2 lesions are highly relevant sources of bleeding. Some clues are also given about the interpretation of inflammatory and ulcerative lesions in the Capsule Endoscopy Crohn's Disease Activity Index (CECDAI) [8] and in the Lewis score [9]. However, these scores aim to grade the severity of SBCE findings in patients with identified Crohn's disease, rather than to establish a diagnosis of Crohn's disease. Overall, the relevance of SBCE findings is based on a low level of knowledge. The aim of this study was to assess the clinical relevance of the most fre-

quent SBCE findings, by examining the answers of CE expert readers to an illustrated script questionnaire.

Materials and methods**Study design**

The study design was based on a series of illustrated script questions on the clinical relevance of CE findings in the SB, with various types and numbers of findings, in various clinical settings.

The questionnaire was prepared by a core group of 3 physicians (RL, XD, PhM), assisted by one premed student (CL) and one GI fellow (CH). Sixteen types of SB findings were selected. The list included five vascular lesions (as listed and defined by a previous consensus [3]): 'typical angiectasia', 'red spot/dot', 'erythematous patch', 'phlebectasia', 'diminutive angiectasia'); seven inflammatory and ulcerative lesions (as listed and defined by another previous consensus [4]): 'aphthoid erosion', 'superficial ulceration', 'deep ulceration', 'edema', 'denudation', 'hyperemia', 'stenosis'); two lymphatic lesions ('chylous cysts', 'lymphangiectasia'); 'blood'; and normal frames (used as controls). Each question displayed a third-generation SBCE (SB3, Medtronic, MN, USA) still frame associated with a written diagnosis (Fig. 1). Four different illustrations were chosen for each of the 16 types of findings. All illustrated findings included for voting were labeled with a precise diagnosis. Tumors, polyps, and coeliac disease were not addressed in this study.

For each illustrated finding, the experts had to provide their interpretation in 3 circumstances (this is later referred to as number category): if the finding identified in the SBCE

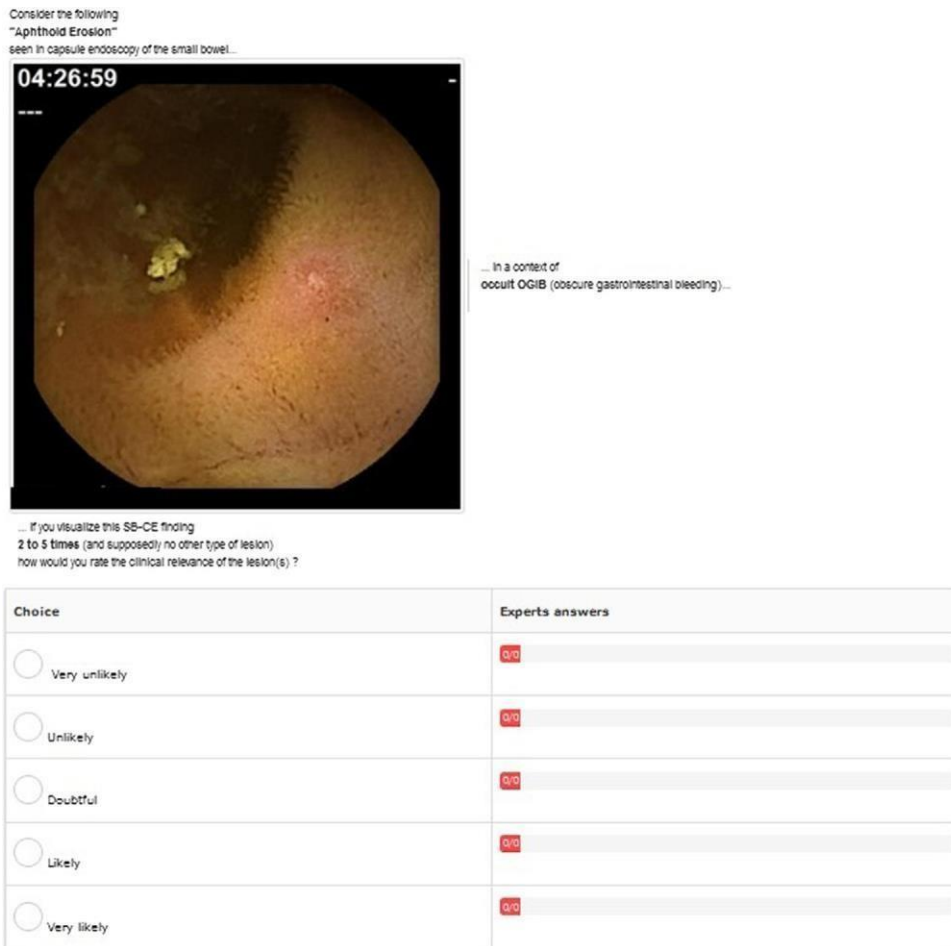


Figure 1 Screenshot of the internet-based questionnaire, displaying one of the 576 labelled and illustrated questions, with one type of finding (amongst the 16 different types), one category of numbers (amongst the 3 different categories), and one clinical setting (amongst the 3 different ones).

Table 1 Numerical scale for voting process.

Vote	Numerical value
Very unlikely	-2
Unlikely	-1
Doubtful	0
Likely	1
Very likely	2

was observed only once (a single lesion), if the finding was observed 2–5 times, and if it was found 6 or more times. Experts were asked to give their opinion on the relevance (for diagnosis and/or for start or change of a therapy) of this type and number of finding(s) in the three following clinical settings: overt OGIB / occult OGIB / suspected Crohn’s disease. For each specific question, experts were asked to rate the clinical relevance of the labeled and illustrated finding, taking into account its type and number category, and the clinical setting, as follows: very unlikely (-2) / unlikely (-1) / doubtful (0) / likely (+1) / very likely (+2) (Table 1 and Fig. 2). Overall, through this questionnaire, each individual expert had to answer a set of four different labeled

and illustrated questions (named “scenario”) for one type of finding (amongst the 16 different types), one number category (amongst the 3 different categories), and one clinical setting (amongst the 3 different ones), which makes a total of 192 scenarios, with a total of 576 illustrated questions per expert. This approach (four questions per scenario) was chosen first because the answers of the experts to the very same questions may not be perfectly reproducible, and second because the expert’s opinion may vary with the different illustrations of a specific type of finding. The questionnaire was internet-based using an e-learning electronic platform. The 576 questions were shuffled. The core group checked that the four questions within a scenario were never consecutive. Experts were allowed to log in and out at any time, until they completed all 576 questions.

Expert group

According to the methodology of script questionnaires, the core group decided to include 10–20 experts in the study [10,11]. An expert in the field of SBCE was defined as a board-certified gastroenterologist with current clinical, teaching and research activities in SBCE. A minimum 5-

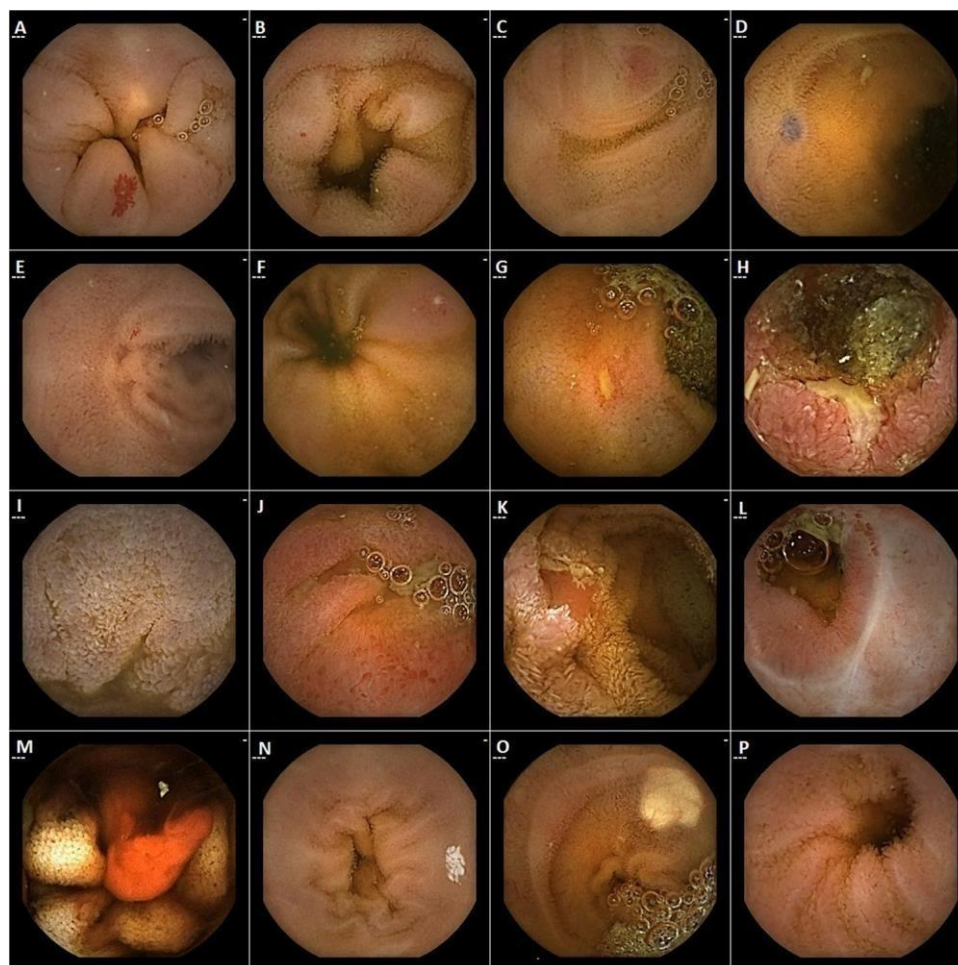


Figure 2 Examples of illustrations (amongst four per category) of the 16 different types of findings assessed for their clinical relevance on the internet-based questionnaire. (a) typical angiectasia; (b) red dot/spot; (c) erythematous patch; (d) phlebectasia; (e) diminutive angiectasia; (f) aphthoid erosion; (g) superficial ulceration; (h) deep ulceration; (i) edema; (j) hyperemia; (k) denudation; (l) stenosis; (m) blood; (n) lymphangiectasia; (o) chylous cyst; (p) normal mucosa.(3,4).

year experience in SBCE reading, with annual reading of at least 50 video recordings a year, teaching activities, and publications related to SBCE in international peer-reviewed journals, were required to be considered an expert. Fifteen international experts fulfilling these criteria were invited.

Analysis

All answers were reviewed and analyzed. A mean score was first calculated per expert and per scenario. The median score (with first Q1 and third Q3 quartiles) of the 14 mean scores per scenario obtained from the 14 experts were then calculated. The clinical relevance of findings was based on these median scores per scenario. Keys for interpretation of the median scores are given in Table 2. Findings with a

median score below or equal to -0.75 were considered to have a low clinical relevance (P0). Findings with a median score between -0.75 and $+0.75$ were considered to have an intermediate/doubtful clinical relevance (P1), whereas those with a mean score over $+0.75$ were considered to have a high clinical relevance (P2). Other quantitative data were reported as mean \pm standard deviation (SD).

Table 2 Clinical interpretation of votes, according to median score.

Median score	Interpretation
$-2.00 \leq \text{median score} \leq -0.75$	Low clinical relevance (P0)
$-0.75 < \text{median score} < +0.75$	Intermediate/doubtful clinical relevance (P1)
$+0.75 \geq \text{median score} \geq +2.00$	High clinical relevance (P2)

Table 3 Interpretation of the votes in the setting of overt obscure gastrointestinal bleeding. Each box displays the median (Q1;Q3) score of the mean scores obtained from the 14 experts to 4 different illustrated questions (56 answers per “scenario”). P0, P1 or P2 interpretations (and the color of the box) are determined according to Table 2.

Type of finding	Found once	Found 2 to 5 times	Found 6 times or more
<i>Vascular lesions</i>			
Typical angiectasia	+1.00 (+0.50 ; +1.44)	+1.50 (+0.81 ; +1.75)	+1.62 (+1.06 ; +2.00)
Red spot /dot	-2.00 (-2.00 ; -1.44)	-1.75 (-2.00 ; -1.00)	-1.62 (-2.00 ; -0.75)
Erythematous patch	-2.00 (-2.00 ; -1.44)	-1.87 (-2.00 ; -1.00)	-1.87 (-2.00 ; -1.00)
Phlebectasia	-1.87 (-2.00 ; -1.12)	-1.50 (-1.75 ; -0.32)	-1.50 (-2.00 ; -0.81)
Diminutive angiectasia	-2.00 (-2.00 ; -1.00)	-1.00 (-1.25 ; -0.75)	-0.87 (-1.25 ; -0.06)
<i>Ulcerative and inflammatory lesions</i>			
Aphthoid erosion	-1.62 (-2.00 ; -0.81)	-0.50 (-0.87 ; +0.19)	-0.37 (-0.94 ; -0.06)
Superficial ulceration	-0.50 (-1.44 ; +0.37)	+0.75 (+0.25 ; +1.00)	+1.00 (+0.31 ; +1.19)
Deep ulceration	+1.62 (+1.00 ; +1.94)	+2.00 (+1.37 ; +2.00)	+2.00 (+2.00 ; +2.00)
Edema	-1.62 (-1.94 ; -1.00)	-0.75 (-1.00 ; -0.50)	-0.75 (-1.00 ; -0.25)
Hyperemia	-2.00 (-2.00 ; -1.25)	-1.00 (-1.81 ; -0.75)	-1.00 (-1.81 ; -0.75)
Denudation	-1.00 (-1.44 ; -0.81)	-0.62 (-1.00 ; -0.25)	-0.25 (-0.69 ; +0.25)
Stenosis	+1.12 (+0.56 ; +1.25)	+1.12 (+0.81 ; +1.94)	+1.25 (+0.81 ; +2.00)
<i>Other</i>			
Normal	-2.00 (-2.00 ; -2.00)	-2.00 (-2.00 ; -2.00)	-2.00 (-2.00 ; -2.00)
Lymphangiectasia	-2.00 (-2.00 ; -1.25)	-2.00 (-2.00 ; -2.00)	-2.00 (-2.00 ; -2.00)
Chylous cyst	-2.00 (-2.00 ; -2.00)	-2.00 (-2.00 ; -1.75)	-2.00 (-2.00 ; -2.00)
Blood	+2.00 (+2.00 ; +2.00)	+2.00 (+2.00 ; +2.00)	+2.00 (+2.00 ; +2.00)

Results

Fourteen experts participated to the study, and one declined the invitation (participation rate 93.3%). The 14 participating experts were based in France (n = 2), Germany (n = 1), Ireland (n = 1), Israel (n = 1), Italy (n = 2), Portugal (n = 1), Spain (n=1), Sweden (n=1), and United Kingdom (n=4).

The mean age of the members of the experts' group was 47.6 ± 9.7 years. The mean CE reading experience of these experts was 15.5 ± 2.9 years year, with a mean declared number of 200 ± 126 readings annually. All were active SBCE readers, and were trainers for SBCE reading. All were authors of publications related to SBCE in international peer-reviewed journals. The 14 participating experts com-

Table 4 Interpretation of the votes, in the setting of occult obscure gastrointestinal bleeding. Each box displays the median (Q1;Q3) score of the mean scores obtained from the 14 experts to 4 different illustrated questions (56 answers per “scenario”). P0, P1 or P2 interpretations (and the color of the box) are determined according to [Table 2](#).

Type of finding	Found once	Found 2 to 5 times	Found 6 times or more
Vascular lesions			
Typical angiectasia	+0.87 (+0.50 ; +1.00)	+1.75 (+1.25 ; +2.00)	+2.00 (+1.25 ; +2.00)
Red spot /dot	-2.00 (-2.00 ; -1.37)	-2.00 (-1.75 ; -0.44)	-1.62 (-1.94 ; -0.31)
Erythematous patch	-1.75 (-2.00 ; -1.31)	-1.37 (-2.00 ; -0.44)	-1.00 (-2.00 ; -0.12)
Phlebectasia	-2.00 (-2.00 ; -1.12)	-1.50 (-2.00 ; -0.75)	-1.25 (-2.00 ; -0.81)
Diminutive angiectasia	-1.37 (-1.75 ; -1.00)	-0.62 (-1.44 ; -0.06)	-0.12 (-0.94 ; 0.00)
Ulcerative and inflammatory lesions			
Aphthoid erosion	-1.00 (-1.75 ; -0.44)	+0.12 (0.00 ; +0.87)	+0.62 (+0.25 ; +1.00)
Superficial ulceration	0.00 (-0.04 ; + 0.94)	+1.00 (+1.00 ; + 1.44)	+1.25 (+1.00 ; + 1.75)
Deep ulceration	+1.87 (+1.06 ; +2.00)	+2.00 (+1.81 ; +2.00)	+2.00 (+1.62 ; +2.00)
Edema	-1.00 (-1.56 ; -0.75)	-0.25 (-0.44 ; +0.19)	-0.25 (-0.94 ; +0.19)
Hyperemia	-1.37 (- 1.75 ; -1.00)	-0.50 (-0.94 ; -0.25)	-0.27 (-0.94 ; -0.06)
Denudation	-1.00 (-1.50 ; -0.56)	+0.00 (-0.19 ; +0.19)	+0.25 (0.00 ; +0.75)
Stenosis	+1.37 (+1.00 ; +1.69)	+1.75 (+1.06 ; +2.00)	+2.00 (+1.06 ; +2.00)
Other			
Normal	-2.00 (-2.00 ; -2.00)	-2.00 (-2.00 ; -2.00)	-2.00 (-2.00 ; -2.00)
Lymphangiectasia	-2.00 (-2.00 ; -2.00)	-2.00 (-2.00 ; -2.00)	-2.00 (-2.00 ; -1.56)
Chylous cyst	-1.81 (-2.00 ; -2.00)	-2.00 (-2.00 ; -2.00)	-1.81 (-2.00 ; -2.00)
Blood	+2.00 (+2.00 ; +2.00)	+2.00 (+2.00 ; +2.00)	+2.00 (+2.00 ; +2.00)

pleted all 576 questions of the questionnaire (completion rate 100%): 8064 answers were therefore available for analysis. Normal frames served as controls, and had median score

of -2.00 , whatever the numbers of findings and the clinical settings. Lymphatic findings (*‘lymphangiectasia’* and *‘chylous cyst’*) were also consistently rated of low pertinence

(P0), with median score of -2.00 in all scenarios.

Overt obscure gastrointestinal bleeding

Any finding of *‘typical angiectasia’*, *‘deep ulceration’*, *‘stenosis’*, *‘blood’*, and multiple findings of *‘superficial ulcerations’*, were considered highly relevant (P2) in the setting of overt OGIB ([Table 3](#)). Findings of isolated *‘superficial ulceration’* and of multiple *‘aphthoid erosion’* and *‘denudation’* were of intermediate/doubtful pertinence (P1). Other

Table 5 Interpretation of the votes, in the setting of suspected Crohn's disease. Each box displays the median (Q1;Q3) score of the mean scores obtained from the 14 experts to 4 different illustrated questions (56 answers per "scenario"). P0, P1 or P2 interpretations (and the color of the box) are determined according to [Table 2](#).

Type of finding	Found once	Found 2 to 5 times	Found 6 times or more
Vascular lesions			
Typical angiectasia	-2.00 (-2.00 ; -1.75)	-1.75 (-2.00 ; -1.50)	-1.87 (-2.00 ; -1.12)
Red spot /dot	-2.00 (-2.00 ; -2.00)	-2.00 (-2.00 ; -2.00)	-2.00 (-2.00 ; -1.81)
Erythematous patch	-2.00 (-2.00 ; -1.37)	-1.12 (-2.00 ; -0.62)	-1.12 (-2.00 ; -0.50)
Phlebectasia	-2.00 (-2.00 ; -2.00)	-2.00 (-2.00 ; -1.81)	-2.00 (-2.00 ; -1.75)
Diminutive angiectasia	-2.00 (-2.00 ; -2.00)	-2.00 (-2.00 ; -2.00)	-2.00 (-2.00 ; -1.81)
Ulcerative and inflammatory lesions			
Aphthoid erosion	-0.87 (-1.19 ; -0.06)	+0.87 (+0.00 ; +1.00)	1.00 (+0.00 ; +1.25)
Superficial ulceration	-0.12 (-0.50 ; +0.44)	+1.00 (+1.00 ; +1.25)	+1.50 (+1.06 ; +1.94)
Deep ulceration	+1.37 (+1.00 ; +2.00)	+2.00 (+2.00 ; +2.00)	+2.00 (+2.00 ; +2.00)
Edema	-0.25 (-0.69 ; 0.00)	+0.25 (0.00 ; +0.44)	+0.50 (+0.06 ; +0.50)
Hyperemia	-1.37 (-1.94 ; -1.00)	-0.75 (-1.31 ; 0.00)	-0.62 (-1.25 ; 0.00)
Denudation	-1.12 (-1.50 ; -0.81)	0.00 (-0.69 ; +0.62)	+0.25 (-0.37 ; +0.50)
Stenosis	+1.00 (+1.00 ; +1.00)	+2.00 (+1.81 ; +2.00)	+2.00 (+1.75 ; +2.00)
Other			
Normal	-2.00 (-2.00 ; -2.00)	-2.00 (-2.00 ; -1.81)	-2.00 (-2.00 ; -2.00)
Lymphangiectasia	-2.00 (-2.00 ; -2.00)	-2.00 (-2.00 ; -2.00)	-2.00 (-2.00 ; -2.00)
Chylous cyst	-2.00 (-2.00 ; -2.00)	-2.00 (-2.00 ; -2.00)	-2.00 (-2.00 ; -1.81)
Blood	0.00 (-0.75 ; +0.81)	-0.12 (-0.94 ; +0.75)	+0.12 (-0.44 ; +0.69)

scenarios had low relevance median scores (P0) in this setting.

Occult obscure gastrointestinal bleeding

Details regarding the results in this setting are given in [Table 4](#). Overall, 'typical angiectasia', 'deep ulceration',

'stenosis' and 'blood' were categorized as highly relevant (P2), whatever their numbers. 'Superficial ulceration' was considered of high clinical relevance (P2) when seen at multiple sites, but of intermediate/doubtful relevance (P1) when found only once. Findings of 'diminutive angiectasia', 'aphthoid erosion', 'edema', 'hyperemia' and 'denudation', were of intermediate/doubtful relevance (P1) when found at multiple sites, and of low relevance (P0) when

only found once. Other scenarios had low relevance median scores (P0) in this setting.

Suspected Crohn's disease

Any finding of '*deep ulceration*' or '*stenosis*', and multiple findings of '*superficial ulceration*' and '*apthoid erosion*', were considered highly relevant (P2) in this setting (Table 5). An isolated finding of '*superficial ulceration*' was deemed of intermediate/doubtful relevance (P1) and a finding of '*apthoid erosion*' was deemed of low relevance (P0). Findings of '*edema*', '*hyperemia*' and '*denudation*' were of low to intermediate/doubtful relevance, according to their number category. A finding of '*blood*' was of intermediate/doubtful relevance (P1), with large interquartile ranges (IQR). All vascular lesions ('*typical angiectasia*', '*diminutive angiectasia*', '*red spot/dot*', '*erythematous patch*', '*phlebectasia*') were considered of low clinical relevance (P0).

Discussion

The present study provides a stronger basis for the definition of three different levels of clinical relevance (P0, P1, and P2), as proposed by Saurin et al. [6], regarding various numbers/ranges of 16 types of findings, in the three most frequent clinical settings for which a SBCE examination is recommended. These results are based on the analysis of 8064 answers of 14 international experts to an illustrated script questionnaire.

Overall, this study suggests that findings of '*red spots/dot*', '*erythematous patch*', '*phlebectasia*', '*lymphangiectasia*' and '*chylous cysts*' are of low clinical relevance (P0), and may be considered as normal variants in most cases, as recently proposed by other authors [12]. Conversely, findings of '*deep ulceration*' or '*stenosis*' are highly relevant (P2) in most cases. The interpretation of other findings is more subtle, and may vary according to the clinical setting and to the number of similar findings in one recording.

The results of the experts to the script questionnaire in the settings of overt and occult OGIB are given with details in Tables 3 and 4, and can be discussed in a more global manner. Findings of '*typical angiectasias*' were clinically relevant to the experts in all scenarios. In the original proposal by Saurin et al. [6], '*angiectasias*' were listed as findings with a high hemorrhagic potential (P2). In the 1-year follow-up of the patients of this series, 61% of those with P2 lesions were treated, compared to 23% of patients with P1 or P0 lesions [13]. These results are consistent with those of a prospective, multicenter study, where patients undergoing an endoscopic ablation of '*typical angiectasia*' (P2) had a significantly lower risk of bleeding recurrence than patients with intermediate/doubtful pertinence (P1) SB vascular lesions [14]. "*Diminutive angiectasias*" were deemed to be of intermediate/doubtful clinical relevance (P1) when multiple (and P0 when singular) in scenarios of occult OGIB. Any other of the listed vascular lesions were of low relevance (P0), whatever their numbers and the clinical setting. All other vascular lesions were categorized P0 lesions, therefore deemed unlikely to be responsible for OGIB according to the experts. Overall, inflammatory and

ulcerative lesions in the setting of OGIB, whether overt or patent, were considered of interest by the experts in terms of clinical relevance, with higher median scores with deeper tissue loss (from '*denudation*' to '*deep ulceration*'), and when found at multiple sites. The presence of '*blood*' was always deemed of high relevance (P2, with a median score of +2.0 in all cases) in the setting of OGIB, and this can be seen as a positive control for our study. In some studies focusing on SBCE examination for OGIB, a specific 'P3' score for '*blood*' [14] was extended from the original score by Saurin et al. that had only three categories (P0, P1, and P2) [6].

In the setting of suspected Crohn's disease, a global interpretation of the results is possible as well. Unsurprisingly, none of the listed vascular lesions was deemed of interest (all had a median score below -1.0, with a P0 interpretation) whatever their type and number. Conversely, all inflammatory or ulcerative lesions were relevant to the diagnosis of Crohn's disease, with a step-wise increase in median score with higher numbers of findings, and with deeper tissue loss (from '*denudation*' to '*apthoid erosion*', and then to '*superficial*' and '*deep*' ulcerations). Still, stigmata of inflammation and ulceration are not specific signs of Crohn's disease, and the interpretation regarding their clinical relevance was often intermediate/doubtful (P1), as illustrated by the lower median scores of more subtle lesions ('*hyperemia*', '*denudation*', '*edema*', and isolated '*apthoid erosion*') compared to '*ulcerations*' (whether superficial or deep). Findings of '*stenosis*' were probably considered more specific of Crohn's disease, whether isolated or multiple, and were therefore quite consistently categorized as highly relevant findings (P2) by experts. Overall, our global reading of the results is consistent with both Lewis score [5] and CECDAI [4], where '*stenosis*'/'*stricture*' is a specific item, and where the extent/width of tissue inflammation or tissue loss increases both scores. Similarly, in the original classification by Saurin et al., a "*large ulceration*" was considered highly relevant (P2) in term of bleeding potential [6]. Interestingly, the interpretation of the presence of '*blood*' was inconsistent for experts in cases of suspected Crohn's disease, with a median score equal or close to 0.0, but with an IQR over 1.0.

Up to now, any evidence on the relevance of lesions found in SBCE was weak. Most SBCE reports, courses and outcome measurements for clinical studies relied on the judgment, insight, opinion and experience of the physician/teacher/investigator. A script questionnaire is considered one of the most reliable and valid tools for the assessment of judgment and clinical reasoning [15–17]. A script questionnaire was more appropriate to build up a guide than a Delphi process that aims to draw a consensus. This study has several strengths regarding this methodology. First, a large number of well-selected experts was involved, as proposed in the standard methodology for script questionnaires [10,11,18]. Second, the questionnaires were illustrated with high-quality, third-generation SBCE images, with vascular and ulcerative / inflammatory findings being named and described according to two international Delphi consensus [3,4]. Third, the high (576) number of questions proposed to the experts, in a random (non-consecutive) order, provided a large overview of their clinical thinking regarding various types and numbers of findings, in different clinical settings, named "scenarios". Fourth, the internet-

based questionnaire allowed high participation (93%) and completion (100%) rates, and independent answers from the experts. Fifth, normal images were also interpreted, thus providing a control group of images, where median pertinence scores were consistently at the lowest level (-2.00). Similarly, 'blood' was a positive control in the setting of OGIB, with scores consistently at the highest level (+2.00). These results in both control groups strengthen the internal validity of the study.

The study has also some limitations. First, the list of 16 SB findings included for voting was established by a core group of non-voting investigators, based on their experience and on previous Delphi consensus [3,4] which could have been extended. Noticeably, the core group decided not to include obvious relevant lesions to the experts (mainly protruding lesions, such as polyps, tumors, varices, and submucosal lesions). Second, the level of evidence employed remains low, although higher than that of a single expert's opinion (as seen today in reports, in courses, and in study protocols). Longitudinal studies could bring higher levels of evidence, but their feasibility is challenging [13]. Last, the clinical settings / scenarios were restricted to three basic indications, but SBCE examination actually covers a wider range of clinical situations in routine practice.

The results presented here aim to better standardize SBCE reading, reporting, teaching and research. Indeed, in a previous study by Saurin et al., a tandem reading of first-generation SBCE recordings showed a 60.4% interobserver concordance rate regarding the pertinence of 225 findings in 60 patients with OGIB [6]. Concordance rates were 59.1%, 51.2% and 75.9% for P0, P1 and P2 lesions, respectively. The CEDAI and the Lewis scores are validated to grade the severity of SBCE findings in patients with known Crohn's disease [8,9], but they have not been evaluated as diagnostic tools. We therefore believe that our results bring a higher level of knowledge regarding the clinical relevance of SB findings in the main indications for CE. Still, many median and IQR scores in our guide reflect some uncertainty from the experts regarding the relevance of subtle findings seen in SBCE. Indeed, most P1 findings should be considered with caution (not overinterpreting and not overlooking. . .), as well as some P0

and P2 findings where the IQR is broad and/or overlapping -0.75 and +0.75 thresholds. Moreover, when talking about relevance, it is important to integrate as much clinical information as possible. For example, OGIB was divided in the questionnaire as overt or occult, but the interpretation of SBCE may vary a lot, depending on other clinical factors. Active on-going vs. past bleeding, the age, gender, comorbidities, treatments (antiplatelet agents, anticoagulants, non-steroidal anti-inflammatory drugs) of patients and subtle findings (*Helicobacter pylori* gastritis, diverticulosis) as well as quality indicators (completeness and cleanliness of bidirectional endoscopy for instance) can all affect interpretation. Similarly, findings deemed herein of intermediate/doubtful relevance SBCE in suspected Crohn's disease may be interpreted differently with the availability of additional data such as symptoms, treatments (most importantly intake of non-steroidal anti-inflammatory drugs), laboratory work-up, imaging, and pathology reports.

Conclusion

SBCE has become a routine endoscopy examination. Efforts have been made to standardize CE training courses, with an international core curriculum having been created [19]. The assessment of SBCE reading competencies is still to be defined and classified in many countries [20,21]. To date, terminology, description of CE findings [3,4], and standards for quality reporting [22] for SBCE have been published, all of which aid in developing a CE reporting consensus. The ESGE recommends grading the level of relevance of SBCE using scores [7]. We believe that this study will help standardizing SBCE interpretation and reporting by physicians, and possibly by automated, artificial intelligence-based systems (patent pending). We believe that the new standards (names, descriptions, and levels of clinical relevance) proposed by our group, based on surveys designed by a core group and addressed to international experts, will also help to better define outcome measurements for research study protocols.

Disclosures

Romain Leenhardt is cofounder and shareholder of Augmented Endoscopy, and has given lectures for Abbvie

Xavier Dray is cofounder and shareholder of Augmented Endoscopy and has acted as a consultant for Alfasigma; Bouchara Recordati; Boston Scientific, Fujifilm, Medtronic, and Pentax

Rami Eliakim has received consultant fees from Abbvie, Takeda, Janssen and Medtronic; Research grant from Medtronic.

John N Plevris has received educational support from Dr Falk, Tillots pharmaceuticals, Jinshan, and research support from Aquillant (Fujifilm).

Deirdre McNamara has received research support/grants from Abbvie and Medtronic and consultant fees from Takeda and MSD.

Acknowledgements

This study has been made possible with the support of San-do (a Novartis company) and with the help of the Société Française d'Endoscopie Digestive (SFED) and its full-service eLearning provider (Theia, Grenoble, France).

References

- [1] Iddan G, Meron G, Glukhovskiy A, Swain P. Wireless capsule endoscopy. *Nature* 2000;405(6785):417.
- [2] Pennazio M, Spada C, Eliakim R, Keuchel M, May A, Mulder CJ, et al. Small-bowel capsule endoscopy and device-assisted enteroscopy for diagnosis and treatment of small-bowel disorders: European Society of Gastrointestinal Endoscopy (ESGE) Clinical Guideline. *Endoscopy* 2015;47(4):352–76.
- [3] Leenhardt R, Li C, Koulaouzidis A, Cavallaro F, Cholet F, Eliakim R, et al. Nomenclature and semantic description of vascular lesions in small bowel capsule endoscopy: an international Delphi consensus statement. *Endosc Int Open* 2019;7(3):E372–9.
- [4] Leenhardt R, Buisson A, Bourreille A, Marteau P, Koulaouzidis A, Li C, et al. Nomenclature and semantic descriptions of ulcer-

- ative and inflammatory lesions seen in Crohn's disease in small bowel capsule endoscopy: an international Delphi consensus statement. *United Eur Gastroenterol J* 2020;8(1):99–107 [Epub ahead of print].
- [5] Saurin JC, Pioche M. Why should we systematically specify the clinical relevance of images observed at capsule endoscopy? *Endosc Int Open* 2014;2(2):E88–89.
- [6] Saurin J-C, Delvaux M, Gaudin J-L, Fassler I, Villarejo J, Vahedi K, et al. Diagnostic value of endoscopic capsule in patients with obscure digestive bleeding: blinded comparison with video push-enteroscopy. *Endoscopy* 2003;35(7):576–84.
- [7] Spada C, McNamara D, Despott EJ, Adler S, Cash BD, Fernández-Urién I, et al. Performance measures for small-bowel endoscopy: a european society of gastrointestinal endoscopy (ESGE) quality improvement initiative. *United Eur Gastroenterol J* 2019;7(5):614–41.
- [8] Gal E, Geller A, Fraser G, Levi Z, Niv Y. Assessment and validation of the new capsule endoscopy Crohn's disease activity index (CECDAI). *Dig Dis Sci* 2008;53(7):1933–7.
- [9] Gralnek IM, Defranchis R, Seidman E, Leighton JA, Legnani P, Lewis BS. Development of a capsule endoscopy scoring index for small bowel mucosal inflammatory change. *Aliment Pharmacol Ther* 2008;27(2):146–54.
- [10] Gagnon R, Charlin B, Coletti M, Sauvé E, van der Vleuten C. Assessment in the context of uncertainty: how many members are needed on the panel of reference of a script concordance test? *Med Educ* 2005;39(3):284–91.
- [11] Lubarsky S, Dory V, Duggan P, Gagnon R, Charlin B. Script concordance testing: from theory to practice: AMEE guide no. 75. *Med Teach* 2013;35(3):184–93.
- [12] Ding Z, Shi H, Zhang H, Meng L, Fan M, Han C, et al. Gastroenterologist-level identification of small bowel diseases and normal variants by capsule endoscopy using a deep-learning model. *Gastroenterology* 2019;157(4):1044–54.
- [13] Saurin JC, Delvaux M, Vahedi K, Gaudin JL, Villarejo J, Florent C, et al. Clinical impact of capsule endoscopy compared to push enteroscopy: 1-year follow-up study. *Endoscopy* 2005;37(4):318–23.
- [14]
- Rahmi G, Samaha E, Vahedi K, Delvaux M, Gay G, Lamou- liatte H, et al. Long-term follow-up of patients undergoing capsule and double-balloon enteroscopy for identification and treatment of small-bowel vascular lesions: a prospective, multicenter study. *Endoscopy* 2014;46(7):591–7.
- [15] Karila L, Francois H, Monnet X, Noel N, Roupret M, Gajdos V, et al. [The script concordance test: a multimodal teaching tool]. *Rev Med Interne* 2018;39(7):566–73.
- [16] Lubarsky S, Charlin B, Cook DA, Chalk C, van der Vleuten CPM. Script concordance testing: a review of published validity evidence. *Med Educ* 2011;45(4):329–38.
- [17] Nazim SM, Talati JJ, Pinjani S, Biyabani SR, Ather MH, Norcini JJ. Assessing clinical reasoning skills using Script Concordance Test (SCT) and extended matching questions (EMQs): a pilot for urology trainees. *J Adv Med Educ Prof* 2019;7(1):7–13.
- [18] Lubarsky S, Gagnon R, Charlin B. Scoring the script concordance test: not a black and white issue. *Med Educ* 2013;47(12):1159–61.
- [19] Fernandez-Urien I, Panter S, Carretero C, Davison C, Dray X, Fedorov E, et al. International core curriculum for capsule endoscopy training courses. *Endosc Int Open* 2017;5(6):E526–38.
- [20] Rajan E, Iyer PG, Oxentenko AS, Pardi DS, Alexander JA, Baron TH, et al. Training in small-bowel capsule endoscopy: assessing and defining competency. *Gastrointest Endosc* 2013;78(4):617–22.
- [21] Rajan E, Martinez M, Gorospe E, Al Bawardy B, Dobashi A, Mara KC, et al. Prospective multicenter study to evaluate capsule endoscopy competency using a validated assessment tool. *Gastrointest Endosc* 2020;91(5):1140–5.
- [22] Korman LY, Delvaux M, Gay G, Hagenmuller F, Keuchel M, Friedman S, et al. Capsule endoscopy structured terminology (CEST): proposal of a standardized and structured terminology for reporting capsule endoscopy procedures. *Endoscopy* 2005;37(10):951–9.

7.8. Détection des lésions vasculaires

Les AGD sont les lésions vasculaires les plus fréquemment rencontrées de l'IG (51). Le développement d'algorithmes automatiques permettant la détection de ces lésions représente un axe essentiel de développement dans l'implémentation de l'IA en VCE. Afin d'approcher cette problématique, un apprentissage machine a été effectué à l'échelle de l'image sur des données annotées de haute qualité issues de la banque d'images CAD CAP (38) permettant le développement d'algorithmes de reconnaissance automatique. Ce travail a fait l'objet de plusieurs communications orales en congrès (52), d'une publication scientifique (53) assortie d'un éditorial (54), et d'un prix d'innovation de l'APHP (55). L'outil développé permet de détecter mais également de segmenter la lésion au sein de l'image avec d'excellentes performances. Dans la lignée de ce travail, notre équipe a démontré les performances de cet algorithme pour la détection universelle des angiodysplasies issues d'images provenant d'autres systèmes de capsule (56).

A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy

Romain Leenhardt, MD,¹ Pauline Vasseur, MSc,² Cynthia Li, MSc,^{1,3} Jean Christophe Saurin, MD, PhD,⁴ Gabriel Rahmi, MD, PhD,⁵ Franck Cholet, MD, PhD,⁶ Aymeric Becq, MD,¹ Philippe Marteau, MD, PhD,¹ Aymeric Histace, PhD,^{*,2} Xavier Dray, MD, PhD^{1,2,*}, The CAD-CAP Database Working Group^y

Paris, Cergy Pontoise, Lyon, Brest, France; Philadelphia, Pennsylvania, USA

Background and Aims: GI angiectasia (GIA) is the most common small-bowel (SB) vascular lesion, with an inherent risk of bleeding. SB capsule endoscopy (SB-CE) is the currently accepted diagnostic procedure. The aim of this study was to develop a computer-assisted diagnosis tool for the detection of GIA.

Methods: Deidentified SB-CE still frames featuring annotated typical GIA and normal control still frames were selected from a database. A semantic segmentation images approach associated with a convolutional neural network (CNN) was used for deep-feature extractions and classification. Two datasets of still frames were created and used for machine learning and for algorithm testing.

Results: The GIA detection algorithm yielded a sensitivity of 100%, a specificity of 96%, a positive predictive value of 96%, and a negative predictive value of 100%. Reproducibility was optimal. The reading process for an entire SB-CE video would take 39 minutes.

Conclusions: The developed CNN-based algorithm had high diagnostic performances, allowing detection of GIA in SB-CE still frames. This study paves the way for future automated CNN-based SB-CE reading softwares.

GI angiectasia (GIA) is an acquired vascular malformation defined as a clearly demarcated, bright-red, flat lesion, consisting of tortuous and clustered capillary dilatations, within the mucosal layer.¹ GIA is the most common small-bowel (SB) vascular lesion, and it is associated with a risk of GI bleeding.² SB capsule endoscopy

(SB-CE) is the currently accepted diagnostic method. Its diagnostic yield is approximately 60% in patients with obscure GI bleeding.³ Any given SB-CE video has a mean number of 50,000 still frames. Thus, analysis of the procedure is a tedious process, and GIA can be missed. Computer-aided diagnosis (CAD) is an

Abbreviations: CAD, computer-assisted diagnosis; CAD-CAP, Computer-Assisted Diagnosis for Capsule Endoscopy; CNN, convolutional neural network; GIA, GI angiectasia; SB, small bowel; SB-CE, small-bowel capsule endoscopy.

DISCLOSURE: J. Saurin is a consultant for Capsvision, Medtronic, and Intromedic. G. Rahmi is a consultant for Medtronic. X. Dray is a consultant for Boston Scientific, Fujifilm, Pentax, and Medtronic. S. Huvelin is a consultant for Medtronic. All other authors disclosed no financial relationships relevant to this publication.

*Drs Histace and Dray are co-last authors.

^yComputer-Assisted Diagnosis for Capsule Endoscopy Database Working Group: Sylvie Sacher-Huvelin: MD, Farida Mesli: MD, Chloé Leandri: MD, Isabelle Nion-Larmurier: MD, Stéphane Lecleire: MD, Romain Gerard: MD, Clotilde Duburque: MD, Geoffroy Vanbiervliet: MD, Xavier Amiot: MD, Jean Philippe Le Mouel: MD, Michel Delvaux: MD, Pierre Jacob: MSc, Camille Simon-Shane: MSc, Olivier Romain: PhD.

Copyright © 2018 by the American Society for Gastrointestinal Endoscopy 0016-5107/\$36.00
<https://doi.org/10.1016/j.gie.2018.06.036>

Received April 7, 2018. Accepted June 29, 2018.

Current affiliations: Sorbonne University, Department of Hepato-Gastroenterology, APHP, Saint Antoine Hospital, Paris (1), ETIS, Université de Cergy-Pontoise, ENSEA, CNRS, Cergy-Pontoise Cedex, France (2), Drexel University, College of Arts & Sciences, Philadelphia, Pennsylvania, USA (3), Department of Endoscopy and Gastroenterology, Pavillon L, Hôpital Edouard Herriot, Lyon (4), Georges Pompidou European Hospital, APHP, Department of Gastroenterology and Endoscopy, Paris (5), Digestive Endoscopy Unit, University Hospital, Brest, France (6).

Reprint requests: Romain Leenhardt, MD, Sorbonne University, Department of Hepato-Gastroenterology, Paris F-75012, France.

If you would like to chat with an author of this publication, you may contact Dr Leenhardt at romainleni@gmail.com.

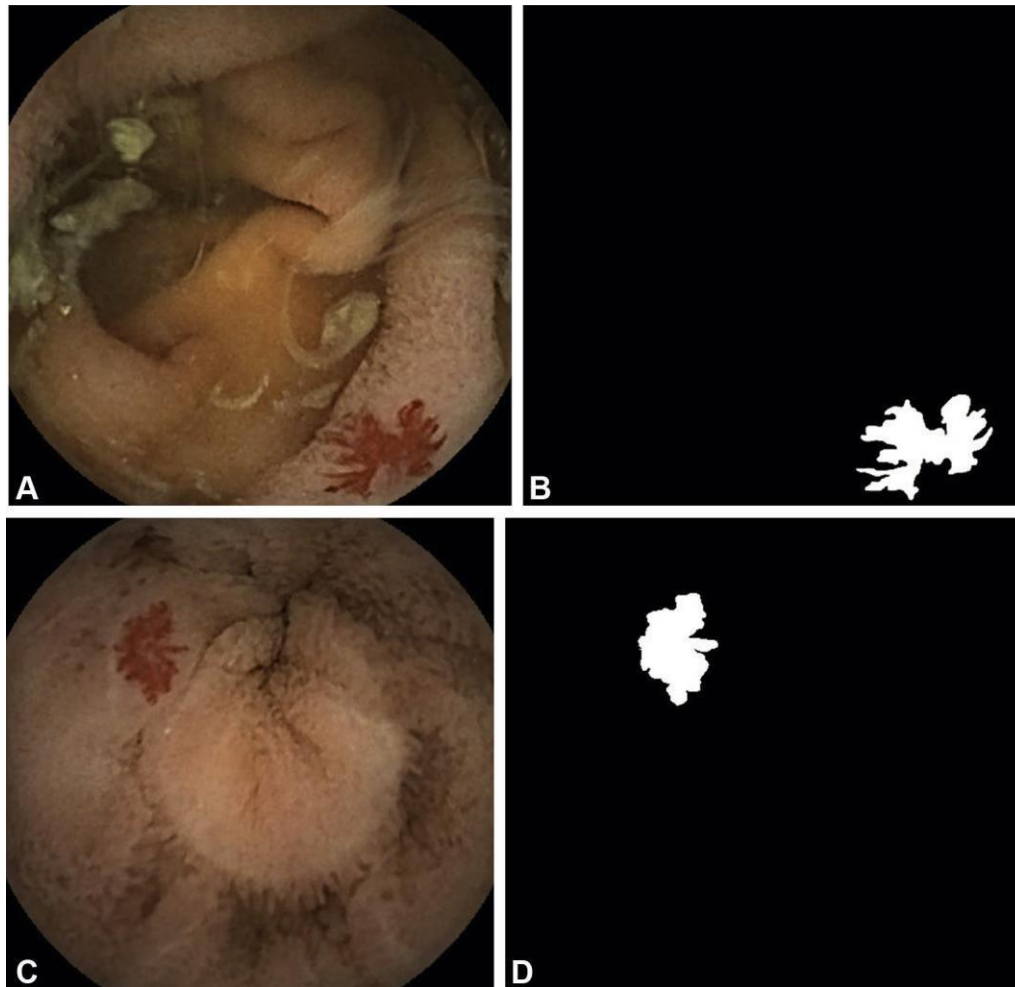


Figure 1. Annotations of small-bowel still frames featuring GI angiectasia. A and C, Native deidentified still frame showing GI angiectasia. B and D, Annotation of the GI angiectasia within the original still frame.

interdisciplinary technology combining elements of artificial intelligence, computer vision, and pathology image processing. CAD is increasingly used for medical practice and research purposes. A CAD algorithm would be a valuable asset with regard to the time-consuming task of reviewing SB-CE, and data have been published on the matter.^{4,5} The aim of this study was to develop a CAD algorithm for the purpose of GIA detection in the setting of SB-CE and to evaluate its diagnostic performances on still images.

MATERIAL AND METHODS

Computer-assisted diagnosis for a capsule endoscopy database

SB-CE still frames were extracted from the Computer-Assisted Diagnosis for Capsule Endoscopy database (CAD-CAP). CAD-CAP is a French national multicenter database endorsed by the Société Française d'Endoscopie Digestive and is approved by the French Data Protection Authority, in which still frames (associated with

short adjacent video sequences) collected from 4166 deidentified, third-generation, SB-CE videos (PillCam SB3 system, Medtronic, Minneapolis, Minn, USA) routinely recorded in 13 French centers can be found. This database comprises 6360 still frames, with a pathology finding from 1341 SB-CEs.

GIA still frames datasets

After initial training, pathology findings of the 6360 still frames were manually annotated by pre-med students under the supervision of SB-CE expert readers, by using Adobe Photoshop CS6 (San Jose, Calif, USA) and GIMP softwares (Orinda, Calif, USA) with a Wacom pen tablet (Kazo, Japan) connected to a laptop (Fig. 1). Among the 6360 still frames with pathology findings, a total of 2946 still frames with a vascular lesion were extracted. Then, still frames with typical non-hemorrhagic GIA were selected by the expert readers. Typical GIAs were defined as clearly demarcated, bright-red, flat lesions.¹ These still frames belonged to 208 SB-CEs (126 men and 82 women, with a mean [\pm standard deviation] age of 72 years \pm 11.7

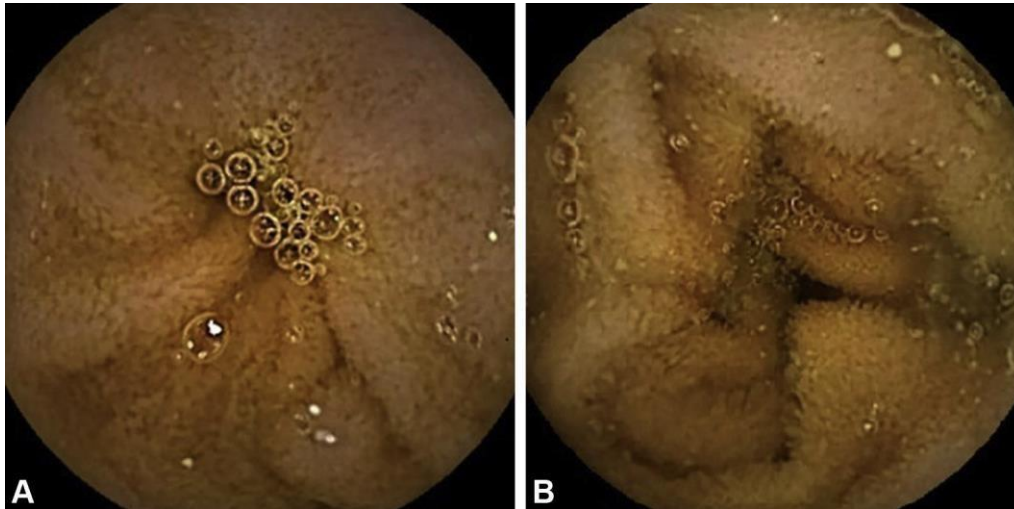


Figure 2. Normal small-bowel video capsule endoscopy still frames.

years) performed in the setting of obscure GI bleeding (62.5% for anemia and 37.5% for external bleeding). To avoid overfitting, successive frames of the same lesions were excluded from the database. Two datasets were created by using these still frames. The first dataset (A1) was used for the purpose of machine training and learning. The second dataset (A2) was used for the testing phase.

Control still frames datasets

Among the normal SB-CEs of the database, 20,000 normal still frames were extracted from 200 capsule endoscopy copies (Fig. 2). The number of still frames kept for analysis matched that of the ones with a GIA. Two datasets were created. The first dataset (N1) was used for the purpose of machine training and learning. The second dataset (N2) was used for the testing phase.

Development of the detection algorithm

The basic morphologic features of a typical GIA image (size, shape, color, pattern, and contour) were discussed between the clinical investigators and computed-vision specialists. Taking these features into account, a hand-crafted colorimetric segmentation approach of SB-CE still frames was initially designed, based on the red over green ratio and on the variation of the image saturation. However, the primary results of this type of computed analysis were unsatisfactory. Algorithms based on machine learning approaches were then tested to improve these initial results. Convolutional neural network (CNN)-based approaches appeared most promising. CNNs have been applied widely to a variety of pattern recognition problems in the past few years, including in the setting of SB bleeding.⁶ Basically, a CNN is a computerized structure that allows large data classification. When dealing with images, CNN mimics the human eye-brain

interface.⁷ CNNs essentially work in 2 steps. The first step, named *deep feature extraction*, extracts local features by training the network on a given dataset of images for which pathology findings were delimited previously. During the second step, CNNs classify new images based on features extracted and learned from the first step. Overall, the approach proposed here was to create a CNN-based semantic segmentation algorithm in order to detect and precisely localize GIAs within the frame.

Analysis of still frame datasets

All still frames were validated by a group of capsule endoscopy expert readers, thus building the ground truth reference CAD-CAP database. The first phase of this analysis (training and learning phase) aimed to achieve an adequate distinction between normal and GIA still frames by using the N1 dataset as a control for A1. The second phase (testing phase) aimed to evaluate the performances of the algorithm by using the N2 dataset as a control for A2 (Fig. 3).

Outcomes

The primary endpoint of the study was the sensitivity of the algorithm. Specificity and positive and negative predictive values were the secondary endpoints. The intraobserver kappa correlation coefficient and the time of analysis also were calculated.

RESULTS

Among the 2946 still frames with vascular lesions, 600 with typical non-hemorrhagic GIA were identified and included to create 2 datasets of 300 still frames each (A1 and A2). Next, 600 normal still frames were selected randomly to create both control datasets (N1 and N2) (Fig. 3).

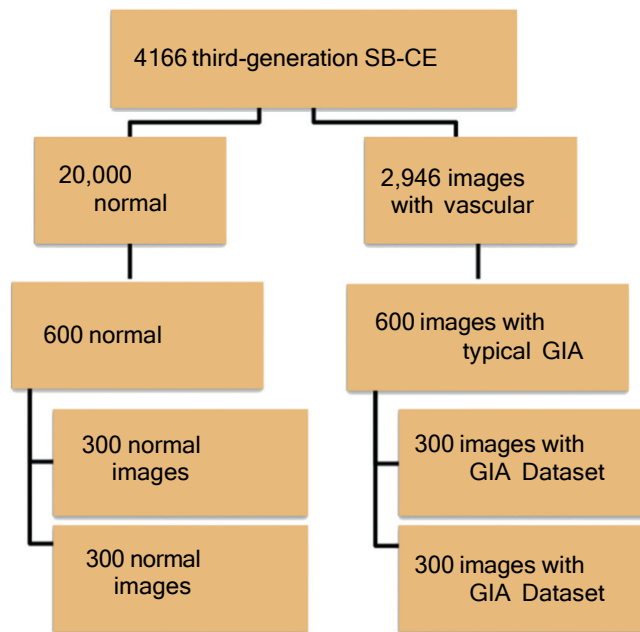


Figure 3. Study flow chart. *SB-CE*, small-bowel capsule endoscopy; *GIA*, GI angiectasia; *N1*, *N2*, The first dataset (N1) was used for the purpose of machine training and learning. The second dataset (N2) was used for the testing phase; *A1*, *A2*, The first dataset (A1) was used for the purpose of machine training and learning. The second dataset (A2) was used for the testing phase.

The algorithm, associating a segmentation approach with a CNN-based approach for deep feature extraction, yielded a sensitivity for GIA detection of 100.0% (95% confidence interval [CI], 100%-100%). Further, a specificity of 96.0% (95% CI, 93.78%-98.22%), a positive predictive value of 96.15% (95% CI, 93.97%-98.33%), and a negative predictive value of 100.0% (95% CI, 100%-100%) were calculated (Table 1, Fig. 4). The reproducibility of these measures was optimal ($\kappa = 1.0$). When we used a dedicated graphical processor unit (approximate cost 1200V) with an adequate programming language (CUDA, Nvidia, Santa Clara, Calif, USA), the reading process for a still image was 46.8 ± 1.8 ms/image, giving an estimated 2340 ± 90 seconds (39 ± 1.5 minutes) for a full-length SB video with a mean number of 50,000 frames.

DISCUSSION

The proposed CNN-based segmentation algorithm achieved excellent diagnostic accuracy for GIA detection in SB-CE still frames. Moreover, this detection method is fully automated and perfectly reproducible. Furthermore, we can extrapolate that the automated reading of a full-length SB-CE would take less than an hour. Some strengths of this study should be emphasized. First, we built a tight collaboration between clinical investigators and experts in the field of computer vision to guide the segmentation approach. Second, we constructed a solid ground truth

TABLE 1. Diagnostic performances of a convolutional neural network-based algorithm associated with a semantic segmentation images approach, for detection of GIA

	Images with GIA	Normal images	Total
Positive test	300	12	312
Negative test	0	288	288
Total	300	300	
Sensitivity, % CI 95%	Specificity % (CI 95%)	PPV % (CI 95%)	NPV % (CI 95%)
100 (100-100)	96.0 (93.78-99.22)	96.15 (93.97-99.3)	100 (100-100)

GIA, GI angiectasia; *CI*, confidence interval; *PPV*, positive predictive value; *NPV*, negative predictive value.

as a reference, based on a recent international and consensual definition of GIA.¹ Further, a significant number of manually annotated GIA still frames extracted from third-generation SB-CEs of a large, national, multicenter database were used. Third, this study was focused on the recognition of typical GIA only. Future studies may address the issue of subtle vascular lesions automated identification, like red spots or erythematous patches. However, it is likely that these latter lesions have a low risk of GI bleeding, with little clinical pertinence. Fourth, we performed a strong validation of our results by using 2 separate datasets for machine learning and for validation of the algorithm.

Several limitations must be discussed as well. First, our analysis was conducted on still frames, as opposed to full-length videos. The main drawback of this approach is that generalization of the use of this algorithm to full-length videos cannot be made. Second, all still frames come from the SB-CE PillCam SB3 system. Therefore, the algorithm performances could differ with other SB-CE systems through different inner characteristics of image processing. Third, all analyses regarding pathology still frames were done with clean images. Future studies at the video level will have to face the issue of SB preparation quality, possibly generating false positive results.

SB-CE reading by physicians is tedious (mean number of 50,000 SB frames per video) and time-consuming (30-40 minutes per read in most studies).⁸ This entails an inherent risk of missed lesions during the reading process by physicians. Several computerized tools have therefore been developed to ease the burden of reading SB-CEs, by facilitating and accelerating this process.⁹ Some computerized systems have shown promising results, but only a few are readily available in commercialized SB-CE reading software. For instance, the Suspected Blood Indicator algorithm (Medtronic) has been shown to have an excellent sensitivity (99%) for the detection of active GIB. However, its performances for detection of potentially bleeding lesions is limited (sensitivity of 55%, specificity of 58%).¹⁰ The Quick-View algorithm (Medtronic) selects 10% of the most relevant images of an SB-CE and reduces the reading time to 11.6

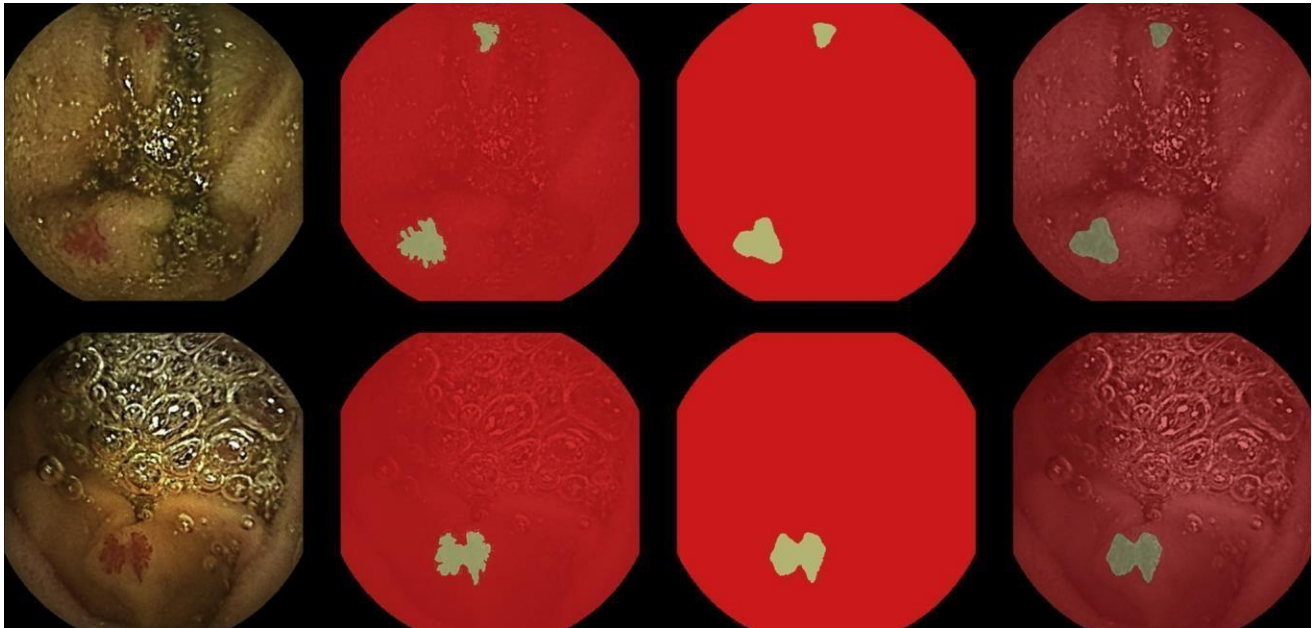


Figure 4. Process for detection of GI angiectasia. From left to right: original still frame, annotated image, convolutional neural network (CNN) approach, CNN approach on original still frame.

minutes. However, the risk of missed lesions is significant, and results are discordant (28%) when compared with human reading.¹¹ Further studies are warranted before the Quick-View algorithm is to be used in clinical practice. The ExpressView algorithm (Capsvision, Saratoga, Calif) reduces by 50% the reading time of SB-CEs, with a sensitivity 78% for the detection of relevant lesions.¹² Although these algorithms are promising, improvements are necessary. Based on our study, we believe that CNNs make up the most suitable approach for addressing this issue. CNNs are inspired by the human neuronal synapse system and are increasingly used in the field of medical imaging. In the setting of SB-CEs, CNNs have yet to be fully evaluated. Pan et al⁶ developed a CNN by using color texture features to recognize and mark bleeding regions, with a sensitivity of 93% and a specificity of 96%. Recently, Noya et al⁴ published the results of the performances of a CNN algorithm used for the purpose of GIA detection. This algorithm yielded a sensitivity of 90% and a specificity of 97%. The low number of patients and the inclusion of second-generation SB-CEs limited this study. Nonetheless, the results of this study concur with that of ours, thus providing external validity to our study. Further studies are needed to develop other algorithms for the detection of active bleeding and clots, inflammatory lesions, and polyps in the field of SB-CE. Then only, prospective studies will be necessary to evaluate the clinical relevance of cloud-based, advanced computer-aided diagnosis tools for SB-CE readings. Still, we believe that such tools will only supplement, but not replace, the readings of capsule endoscopy by fully trained gastroenterologists.

In conclusion, our study lays the foundation of a new generation of highly accurate, automated, SB-CE reading software and calls for future prospective evaluations of CNNs used on full-length SB-CEs.

ACKNOWLEDGMENT

The Computer-Assisted Diagnosis for Capsule Endoscopy database is endorsed by the Société Française d'Endoscopie Digestive. It has been made possible with the support of the Société Nationale Française de Gastro-entérologie and Mayoly Spindler (MSD).

REFERENCES

1. Leenhardt R, Ly C, Koulaouzidis T, et al. Terminology and description of vascular lesions in small bowel capsule endoscopy: an international Delphi consensus statement [abstract]. *Gastrointest Endosc* 2018;87: AB149-50.
2. Becq A, Rahmi G, Perrod G, et al. Hemorrhagic angiodysplasia of the digestive tract: pathogenesis, diagnosis, and management. *Gastroint-est Endosc* 2017;86:792-806.
3. Pennazio M, Spada C, Eliakim R, et al. Small-bowel capsule endoscopy and device-assisted enteroscopy for diagnosis and treatment of small-bowel disorders: European Society of Gastrointestinal Endoscopy (ESGE) Clinical Guideline. *Endoscopy* 2015;47:352-76.
4. Noya F, Alvarez-Gonzalez MA, Benitez R. Automated angiodysplasia detection from wireless capsule endoscopy. *Conf Proc IEEE Eng Med Biol Soc* 2017:3158-61.
5. Iakovidis DK, Koulaouzidis A. Software for enhanced video capsule endoscopy: challenges for essential progress. *Nat Rev Gastroenterol Hepatol* 2015;12:172-86.
6. Pan G, Yan G, Qiu X, et al. Bleeding detection in wireless capsule endoscopy based on probabilistic neural network. *J Med Syst* 2011;35:1477-84.

7. Chen P-J, Lin M-C, Lai M-J, et al. Accurate classification of diminutive colorectal polyps using computer-aided analysis. *Gastroenterology* 2018;154:568-75.
8. McAlindon ME, Ching H-L, Yung D, et al. Capsule endoscopy of the small bowel. *Ann Transl Med* 2016;4:369.
9. Koulaouzidis A, Iakovidis DK, Karargyris A, et al. Optimizing lesion detection in small-bowel capsule endoscopy: from present problems to future solutions. *Expert Rev Gastroenterol Hepatol* 2015;9:217-35.
10. Yung DE, Sykes C, Koulaouzidis A. The validity of suspected blood indicator software in capsule endoscopy: a systematic review and meta-analysis. *Expert Rev Gastroenterol Hepatol* 2017;11:43-51.
11. Saurin J-C, Lapalus MG, Cholet F, et al. Can we shorten the small-bowel capsule reading time with the "Quick-view" image detection system? *Dig Liver Dis* 2012;44:477-81.
12. Saurin JC, Bramli S, Heyries L, et al. Express view reading efficiently reduces reading time for small bowel capsule endoscopy: a prospective multicentric study. *Endosc Int Open* 2018;6:E616-21.

7.9. Détection multiclasse

Afin d'obtenir un outil permettant une détection des différents types de lésions, notre travail d'apprentissage machine s'est étendu aux lésions vasculaires d'imputabilité intermédiaire dans les saignements (points rouges, patchs érythémateux, angiectasies diminutives), aux lésions inflammatoires de la MC (érosion/ulcération muqueuse) ainsi qu'aux lésions protrusives essentiellement représentées par les polypes. Nous avons mené une étude à l'échelle de la vidéo afin de valider les performances de l'algorithme développé. Une évaluation rétrospective sur 149 vidéos a été effectuée en comparant la lecture conventionnelle humaine à la lecture assistée par l'algorithme d'IA par trois experts en VCE. Les VCE incluses étaient réalisées dans un contexte de SSIG en excluant les hémorragies actives de l'IG. Le critère de jugement principal était la détection de lésions P1 et/ou P2 à l'échelle du patient. Le temps d'analyse était également étudié. Les sensibilité et spécificité après évaluation consensuelle étaient de 86 % et 100 % respectivement. Les valeurs prédictives positive et négative étaient de 100 % et 93 % respectivement. La précision diagnostique de l'algorithme était mesurée à 95.3 %. Le nombre moyen d'image à lire était réduit de 10000 à 150 images pour un temps de lecture réduit de 30 à 3 minutes en moyenne. Ces données n'ont pas encore été publiées mais ont été communiquées en congrès (57). Des travaux futurs prospectifs de validation externe sont nécessaires pour l'évaluation des performances de cet outil dans la détection des lésions inflammatoires, des polypes et des saignements actifs et font l'objet de demandes de financement.

7.10. La problématique de la détection du sang

La détection d'une hémorragie digestive de l'IG est une problématique complexe de recherche du fait de la présentation variable de l'aspect du sang. En effet, un saignement de l'IG peut se présenter sous la forme de petites traces de sang rouge présentes sur quelques images uniquement, de volumineux caillots de sang rouge sur plusieurs dizaines ou centaines d'images ou encore de sang noir traduisant un saignement plus ancien et digéré au sein de l'IG qui dans ce cas de figure peut être présent sur des milliers d'images (Figure 16). Cet axe de recherche est d'autant plus complexe que ces données sont rares : seule une minorité des VCE-G en contexte de SSIG révèlent une hémorragie active, tous aspects confondus. La méthodologie à adopter est par conséquent complexe et variable en fonction de l'objectif de recherche.

Une approche initiale utilisant des algorithmes colorimétriques n'a pas permis d'obtenir des résultats suffisants notamment dans la détection du sang noir. L'utilisation de CNN semble plus prometteuse mais nécessite d'élargir notre banque d'images. Nous travaillons à une banque de données plus riche (quantitativement et qualitativement) pour parfaire la détection des différents types de saignement.

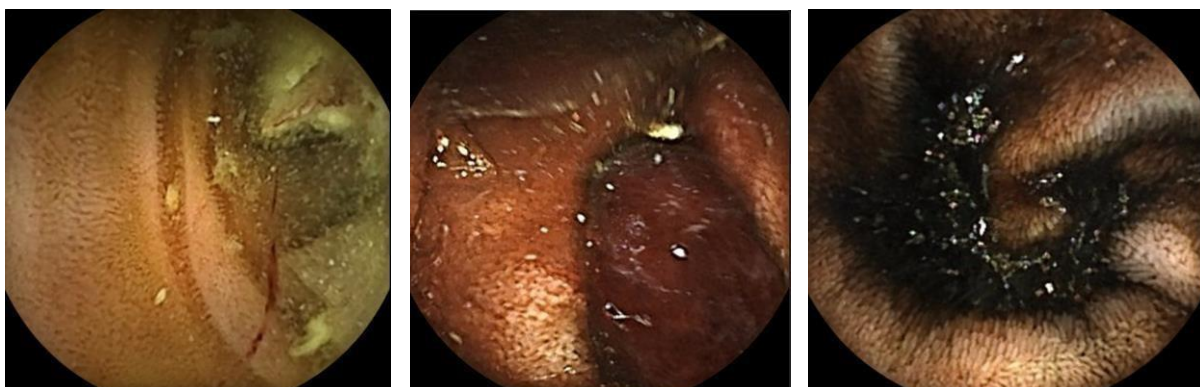


Figure 16 : Illustration d'images de saignement, avec de gauche à droite : trace de sang, caillots de sang, sang noir digéré

8. Valorisation industrielle

Les travaux de recherche présentés dans ce manuscrit sont le fruit d'une collaboration étroite entre le laboratoire de recherche ETIS (Equipes Traitement de l'Information et Systèmes), un centre commun au CNRS, à l'université de Cergy Pontoise et à l'École nationale supérieure de l'électronique et de ses applications (ENSEA) et l'équipe du Centre d'Endoscopie Digestive de l'hôpital Saint Antoine (Sorbonne Université, APHP, Paris, France), par l'entremise de l'Office de transfert de technologie et de partenariat industriel (OTTPI) de l'APHP. Une première phase clé du développement a été rendue possible par le soutien financier de la SATT IdInnov (devenue Erganeo). A partir de 2016, les travaux se sont focalisés autour du développement d'outils d'aide à la détection en VCE. Les travaux scientifiques présentés dans ce manuscrit ont été initialement valorisés par des publications scientifiques et présentations en congrès internationaux mais également par le prix APInnov 2018 de l'interne innovant de l'APHP. Depuis, deux brevets ont été déposés pour l'évaluation de la pertinence des lésions selon le contexte clinique et pour l'évaluation automatique de la qualité de la visualisation de l'IG.

Ce projet innovant a progressivement gagné en notoriété et a abouti à un projet ambitieux de valorisation industrielle afin que ces outils développés puissent servir notre quotidien de

médecin et répondre aux besoins actuels en VCE. L'année 2019 a donc amené la fusion des différents travaux en une unique solution d'IA dédiée à la lecture des VCE et à la création d'une startup nommée Augmented Endoscopy. Ma contribution au sein de cette société fait l'objet d'une convention de « conseil scientifique » avec mes tutelles.

Une partie du travail de cette thèse s'est donc également consacrée au développement de cette société : création d'une image de marque, développement d'un site internet et d'une plateforme SAAS (*Software As a Service*). En 2020, la solution Axaro.AI est créée et offre la première plateforme sécurisée basée sur le cloud permettant la lecture des VCE grâce aux outils de l'IA. Cette plateforme réunit toutes les briques algorithmiques développées antérieurement en un outil permettant la lecture accélérée des VCE assistée d'une aide à la détection d'anomalies (au sein de la vidéo, et au sein de chaque image d'intérêt, avec l'indication d'un pourcentage de confiance) et à l'évaluation de la qualité de la préparation.



Figure 17 : Solution Axaro.AI

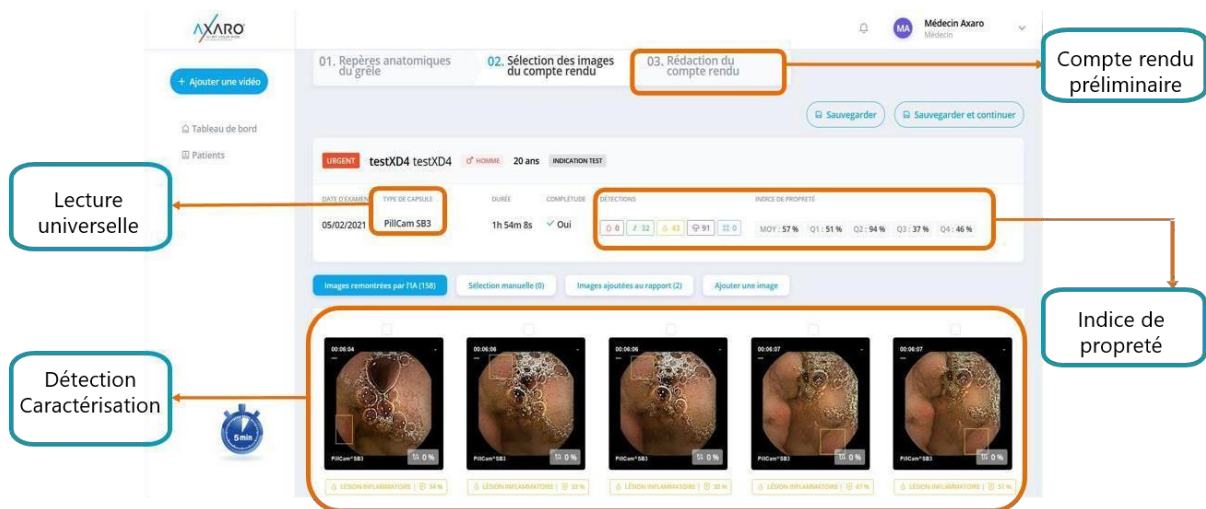


Figure 18 : présentation de la plateforme Axaro.AI

Ce projet est lauréat du concours d'innovation i-Lab 2022 financé par la Bpi France (banque publique d'investissement). Les perspectives de ce projet sont nombreuses avec tout d'abord l'amélioration des performances diagnostiques des algorithmes grâce à l'apprentissage actif lié à l'enrichissement de la base de données (protocole NOCE, démarré le 10 février 2023, dont j'assume la responsabilité scientifique). Certains partenariats industriels sont en cours de discussion pour valoriser cette solution d'IA et l'intégrer dans notre pratique clinique. Cet outil a également vocation à devenir à terme une plateforme pour faciliter les échanges de données de santé entre praticiens.

9. Discussion

Ce travail de thèse applicative se positionne à la frontière d'un monde médicotechnique riche et complexe. Il est avant tout le fruit d'une collaboration étroite s'articulant autour d'une expertise pluridisciplinaire associant la médecine clinique à l'analyse d'image assistée par ordinateur (*computer vision*). L'émergence des méthodes d'apprentissage profond a considérablement accéléré l'implémentation de l'IA en médecine. Cette dernière décennie a vu une augmentation exponentielle du nombre de publications scientifiques dans ce domaine avec des applications cliniques concrètes et pertinentes. A titre d'exemple, en 2018, la *Food and Drug Administration* (FDA) a approuvé aux Etats Unis l'utilisation d'un système de reconnaissance automatique pour le diagnostic de la rétinopathie diabétique (58). Notre Histoire nous démontre que les différentes révolutions technologiques

passées ont pu provoquer des bouleversements dans la façon de concevoir nos métiers. La révolution numérique à laquelle nous assistons promet d'en être un nouvel exemple.

Les champs d'application de l'IA en VCE sont vastes et nécessitent d'être encadrés dans leur développement et leur utilisation en pratique clinique. La rentabilité diagnostique des lésions de l'IG représente un critère de qualité d'une VCE mais demeure variable selon l'indication de l'examen. De plus, il est démontré que cette rentabilité diagnostique est influencée par l'expertise du lecteur, variant de l'expert (>500 VCE lues par an) au stagiaire (59). L'implémentation de l'IA en VCE est donc hautement souhaitable pour palier à cette variation inhérente à la lecture conventionnelle humaine.

Les premières démonstrations scientifiques des performances de l'IA en VCE se sont concentrées sur le développement de solutions permettant souvent la détection d'un seul type d'anomalie (lésions vasculaires, inflammatoires, lésions protrusives) (53,60,61). Or en pratique clinique, la nature des lésions est souvent peu prévisible et il apparaît indispensable de développer un outil permettant une détection exhaustive des différents types de lésions. De plus, ces travaux précurseurs étaient réalisés à l'échelle d'images fixes et non de vidéos entières ce qui limite l'extrapolation des résultats. Des travaux plus récents ont démontré les performances d'algorithmes permettant une détection multiclasse des lésions (62–64). Certains auteurs ont même proposé des algorithmes permettant d'évaluer la pertinence clinique des lésions détectées selon la classification de Saurin (65). Certaines firmes proposent déjà leur solution d'IA, souvent intégrée au logiciel de lecture avec d'excellentes performances diagnostiques associées à une réduction du temps de lecture (66,67). Néanmoins, les données disponibles de la littérature médicale paraissent encore insuffisantes pour permettre de s'affranchir complètement d'une lecture conventionnelle.

Un des principaux avantages attendus de l'implémentation de l'IA en VCE réside dans la réduction du temps de lecture des examens. A ce sujet, la littérature médicale apporte une réponse claire en démontrant la supériorité nette en terme de rapidité de la lecture associée aux algorithmes d'IA par rapport à la lecture conventionnelle (62,66,67). Pour autant, cet indicateur largement utilisé comme critère de jugement dans la majorité des études disponibles, manque de précision. En effet, l'interprétation d'une VCE associe différentes étapes successives: la lecture de l'enregistrement, suivie d'une interprétation des lésions

détectées, suivie de l'évaluation de la pertinence des lésions selon le contexte clinique, suivie de la rédaction d'un compte rendu d'examen. Toutes ces étapes distinctes demandent du temps qui n'est pas pris en compte dans la méthodologie des études disponibles comparant l'IA à la lecture conventionnelle. Les futures études évaluant le gain de temps obtenu lié à l'utilisation des outils d'IA pourraient à l'avenir prendre en compte ces différents critères afin de réellement apprécier l'intérêt de l'IA.

Un des critères de qualité établi par l'ESGE de la réalisation d'une VCE est un taux de visualisation adéquate de l'IG supérieur à 95 % (26). Pour autant, cet indicateur présente une importante variabilité inter et intra observateur (41) et demeure subjectif ce qui limite son utilisation et son évaluation. Un intérêt réel et prometteur de l'introduction de l'IA en VCE repose sur l'évaluation standardisée et reproductible de la qualité de la visualisation de la muqueuse intestinale. Cette perspective ouvre la voie à un vaste champ d'évaluation possible sur la comparaison des modalités de préparation entourant la procédure et le rendement diagnostique associé. De plus, cet outil pourrait aider à déterminer précisément un seuil de propreté de l'IG en dessous duquel on recommanderait de répéter la VCE au risque de manquer des lésions.

Les différents travaux présentés dans ce manuscrit apportent des résultats à la fois cliniques et fondamentaux. Les études cliniques visant à préciser la description et le lexique des lésions de l'IG forment le socle d'un apprentissage machine de qualité grâce à la constitution d'une base de données homogènes et robustes. Concernant la détection d'anomalies de l'IG, bien que les résultats obtenus pour la détection des lésions vasculaires soient performants (53), cette étude a été réalisée à l'échelle de l'image et une validation externe prospective à l'échelle de la vidéo est nécessaire pour confirmer les résultats. L'algorithme développé pour l'évaluation de la qualité de la visualisation de l'IG est une des premières études publiées à approcher cette problématique à l'échelle vidéo en utilisant des algorithmes basés sur les CNN (42). Ce travail répond à la démarche qualité entreprise par l'ESGE de mettre au point un logiciel d'évaluation de la préparation de l'IG (26).

Plusieurs limites existent concernant les différentes solutions d'IA décrites à ce jour. Tout d'abord, la majorité des études disponibles sont rétrospectives. De plus, la plupart des équipes développant des algorithmes d'IA effectuent leur apprentissage machine à partir

d'images fixes sélectionnées et parfois délimitées générant un risque de sur-apprentissage. Enfin, il existe peu d'étude disponible de validation externe de ces outils réduisant l'extrapolation de leur résultat. La firme Chinoise Ankon® a mené en 2022 un essai prospectif européen en cours de publication sur 137 patients présentant une suspicion de saignement digestif occulte. Leur solution d'IA comparée à la lecture conventionnelle humaine obtenait une sensibilité 93,3% Vs 79,0% ($p=0,005$) en réduisant le temps de lecture de 33'42 minutes à 3'50 minutes (68).

L'implémentation de l'IA en endoscopie digestive a déjà bouleversé nos pratiques et génère indéniablement des changements majeurs. En effet, l'IA exacerbe cette crainte de la déqualification des professionnels de santé induite par le fait de cesser d'accomplir une tâche et donc la perte de l'expertise en perspective. Pour autant, cette IA est la promesse d'outils fiables, infatigables, dotés de capacités d'analyse et d'une rapidité d'exécution dépassant largement celles du cerveau humain. Il est donc primordial d'accompagner ces développements de la conception à l'industrialisation des produits pour en maîtriser le fonctionnement.

Les outils d'IA offrent des perspectives très vastes de recherche et de développement dans le domaine de la VCE et certaines approches restent encore peu étudiées. Si l'on se positionne dans la perspective d'une capsule panentérique explorant l'intégralité du tube digestif (de la bouche à l'anus) analysées par les outils de l'IA, la question de la reconnaissance automatique des repères anatomiques apparaît primordiale. L'identification des repères anatomiques est d'autant plus déterminante qu'elle permet de calculer les indices du temps de transit basés sur le temps de progression de la capsule, indispensables dans la planification d'éventuels examens endoscopiques complémentaires. L'IA de demain devrait être capable de distinguer précisément les différentes parties du tube digestif en identifiant les zones de transition entre les différents segments.

Un autre champ exploratoire pourrait s'attacher à développer un outil d'évaluation automatique de la taille des lésions rencontrées dans l'IG. En effet, cette donnée reste problématique de façon générale en endoscopie digestive car elle demeure souvent subjective et dépendante de l'interprétation de l'expert humain. Or, de nombreuses recommandations de pratique clinique et de surveillance sont basées sur la taille des lésions

d'intérêts. Le développement d'outils d'évaluation automatique pourrait représenter un axe intéressant de développement en VCE dans les années futures.

Le champ d'application de la VCE s'est progressivement étendu au reste du tube digestif pour s'intéresser à la région oeso-gastrique avec l'avènement de la VCE à guidage magnétique dotée d'IA (69). Par ailleurs, dans certaines indications, le dépistage du cancer du côlon peut être réalisé par une VCE colique dédiée pour laquelle les outils d'IA disponibles restent pour le moment moins aboutis que ceux dédiés à l'IG (70). La perspective d'une capsule panentérique dotée d'outils d'IA est donc réjouissante pour les années à venir. Le développement d'algorithmes de détection des lésions spécifiques à chaque organe paraît nécessaire. Ces briques algorithmiques réunies vont être à l'origine d'un nombre de plus en plus important d'images sélectionnées par l'IA à analyser par l'expert humain. Un axe de développement en cours vise à étudier les redondances et les similitudes entre les images de façon à ne sélectionner que les images les plus pertinentes dans le but de réduire le nombre total d'images à analyser. Enfin certaines équipes travaillent également au développement d'une IA embarquée au sein de la capsule de façon à accélérer le rendu des diagnostics effectués (71).

De nombreuses questions demeurent sur la façon dont ces nouveaux outils doivent être utilisés en pratique clinique. Est-il préférable de bénéficier des images d'intérêts sélectionnées par l'IA avant, pendant ou après une lecture conventionnelle ? Peut-on se passer de la lecture conventionnelle ? Des questions complexes de responsabilité émergent face à ces nouveaux outils. Qui sera responsable en cas d'erreur diagnostique ou de lésions manquées ? Les autorités de régulation se penchent activement sur ces questions. Les applications de l'IA excellent dans la modélisation de relations complexes entre des variables données mais le cheminement rationnel pour aboutir au résultat reste parfois obscure. En effet, ces outils présentent une limite en terme d'explicabilité des systèmes (phénomène de « boîte noire ») qui constitue un véritable défi scientifique et sociétal. Cette part d'inconnue complexifie l'adoption de l'IA en endoscopie mais plus spécifiquement en VCE. Le flou juridique entourant la responsabilité de la lecture d'une VCE associée aux outils de l'IA expose les médecins et leurs institutions à des freins au développement, à l'implémentation et donc à l'amélioration de ces outils. De même, le degré de confiance accordé à ces nouvelles technologies va déterminer leur acceptabilité.

L'IA perceptive est déjà omniprésente dans notre quotidien et nous fascine tous les jours au travers de ses applications qui semblent illimitées. Elle s'oppose à l'IA décisionnelle, encore peu visible dans notre société, mais dont l'impact sera nettement plus important et source de peurs au travers du concept de singularité technologique.

Au-delà du concept de déqualification liée à l'utilisation de l'IA, la question du remplacement de l'homme par la machine se pose également. D'un point de vue général, il semble irréaliste que les patients acceptent qu'une machine, dépourvue d'empathie, procède seule à l'ensemble du cycle de soin. Le caractère profondément humain du soin restera à la charge du médecin et des soignants. Pour autant, les conséquences de l'informatisation et de la digitalisation de la médecine ont un bilan jugé mitigé par certains auteurs (72). Ils dénoncent une dégradation de la qualité de vie des médecins au travail associée à l'altération de la relation médecin/malades liée au temps important passé sur les dossiers informatiques plutôt qu'auprès des patients eux-mêmes. L'espoir formulé par l'arrivée de l'IA offre une perspective de redressement de cette tendance en libérant le praticien des tâches répétitives à faible valeur pour se recentrer sur un temps plus clinique, relationnel et décisionnel.

10. Conclusion

Les différents travaux présentés dans ce manuscrit apportent des perspectives cliniques et fondamentales en VCE-G. Les fondements cliniques autour de la terminologie en VCE et la création de la base de données constituent le socle d'un apprentissage machine robuste. Les algorithmes développés dans les travaux présentés offrent des solutions performantes qui répondent à un besoin réel autour de la lecture des VCE. La valorisation industrielle de cette solution effectuée en parallèle de ce travail ouvre des perspectives de recherche et de collaboration prometteuses.

L'intelligence artificielle fait d'ores et déjà partie intégrante de notre société et du monde médical. Le système de soin se voit bouleversé par l'avènement de ces nouveaux outils et leur intégration en pratique clinique. Cette révolution technologique doit être accompagnée pour en maîtriser le contenu et maximiser son potentiel. L'IA est une opportunité pour le corps

médical de prodiguer des soins plus sûrs et de façon plus efficiente. L'IA est également la promesse d'un gain de temps espéré au profit de la relation médecin/patient qui elle, ne pourra pas reposer sur la machine.

Références

1. Haute Autorité de Santé. Cancer colorectal : modalités de dépistage et de prévention chez les sujets à risque élevé et très élevé. Saint-Denis La Plaine: HAS; 2017.
2. Iddan G, Meron G, Glukhovskiy A, Swain P. Wireless capsule endoscopy. *Nature*. 25 mai 2000;405(6785):417.
3. Albert JG, Nachtigall F, Wiedbrauck F, Dollinger MM, Gittinger FS, Hollerbach S, et al. Minimizing procedural cost in diagnosing small bowel bleeding: comparison of a strategy based on initial capsule endoscopy versus initial double-balloon enteroscopy. *Eur J Gastroenterol Hepatol*. juin 2010;22(6):679- 88.
4. Nahon S, Hagège H, Latrive JP, Rosa I, Nalet B, Bour B, et al. Epidemiological and prognostic factors involved in upper gastrointestinal bleeding: results of a French prospective multicenter study. *Endoscopy*. nov 2012;44(11):998- 1008.
5. Chait MM. Lower gastrointestinal bleeding in the elderly. *World J Gastrointest Endosc*. 16 mai 2010;2(5):147- 54.
6. Longstreth GF. Epidemiology and outcome of patients hospitalized with acute lower gastrointestinal hemorrhage: a population-based study. *Am J Gastroenterol*. mars 1997;92(3):419- 24.
7. Kornbluth A, Colombel JF, Leighton JA, Loftus E, ICCE. ICCE consensus for inflammatory bowel disease. *Endoscopy*. oct 2005;37(10):1051- 4.
8. Maunoury V, Savoye G, Bourreille A, Bouhnik Y, Jarry M, Sacher-Huvelin S, et al. Value of wireless capsule endoscopy in patients with indeterminate colitis (inflammatory bowel disease type unclassified). *Inflamm Bowel Dis*. févr 2007;13(2):152- 5.
9. Johanssen S, Boivin M, Lochs H, Voderholzer W. The yield of wireless capsule endoscopy in the detection of neuroendocrine tumors in comparison with CT enteroclysis. *Gastrointest Endosc*. avr 2006;63(4):660- 5.
10. Joyce AM, Burns DL, Marcello PW, Tronic B, Scholz FJ. Capsule endoscopy findings in celiac disease associated enteropathy-type intestinal T-cell lymphoma. *Endoscopy*. juin 2005;37(6):594- 6.
11. Cortegoso Valdivia P, Skonieczna-Żydecka K, Elosua A, Sciberras M, Piccirelli S, Rullan M, et al. Indications, Detection, Completion and Retention Rates of Capsule Endoscopy in Two Decades of Use: A Systematic Review and Meta-Analysis. *Diagn Basel Switz*. 28 avr 2022;12(5):1105.
12. Pennazio M, Spada C, Eliakim R, Keuchel M, May A, Mulder CJ, et al. Small-bowel capsule endoscopy and device-assisted enteroscopy for diagnosis and treatment of small-

bowel disorders: European Society of Gastrointestinal Endoscopy (ESGE) Clinical Guideline. *Endoscopy*. avr 2015;47(4):352- 76.

13. Rahmi G, Samaha E, Vahedi K, Delvaux M, Gay G, Lamouliatte H, et al. Long-term follow-up of patients undergoing capsule and double-balloon enteroscopy for identification and treatment of small-bowel vascular lesions: a prospective, multicenter study. *Endoscopy*. juill 2014;46(7):591- 7.

14. Pennazio M, Rondonotti E, Despott EJ, Dray X, Keuchel M, Moreels T, et al. Small-bowel capsule endoscopy and device-assisted enteroscopy for diagnosis and treatment of small-bowel disorders: European Society of Gastrointestinal Endoscopy (ESGE) Guideline - Update 2022. *Endoscopy*. janv 2023;55(1):58- 95.

15. de Leusse A, Vahedi K, Edery J, Tiah D, Fery-Lemonnier E, Cellier C, et al. Capsule endoscopy or push enteroscopy for first-line exploration of obscure gastrointestinal bleeding? *Gastroenterology*. mars 2007;132(3):855- 62; quiz 1164- 5.

16. Saurin JC, Delvaux M, Gaudin JL, Fassler I, Villarejo J, Vahedi K, et al. Diagnostic value of endoscopic capsule in patients with obscure digestive bleeding: blinded comparison with video push-enteroscopy. *Endoscopy*. juill 2003;35(7):576- 84.

17. Pennazio M, Santucci R, Rondonotti E, Abbiati C, Beccari G, Rossini FP, et al. Outcome of patients with obscure gastrointestinal bleeding after capsule endoscopy: report of 100 consecutive cases. *Gastroenterology*. mars 2004;126(3):643- 53.

18. Liao Z, Gao R, Xu C, Li ZS. Indications and detection, completion, and retention rates of small-bowel capsule endoscopy: a systematic review. *Gastrointest Endosc*. févr 2010;71(2):280- 6.

19. Saurin JC, Delvaux M, Vahedi K, Gaudin JL, Villarejo J, Florent C, et al. Clinical impact of capsule endoscopy compared to push enteroscopy: 1-year follow-up study. *Endoscopy*. avr 2005;37(4):318- 23.

20. Lee HS, Lim YJ, Shim KN, Moon CM, Song HJ, Kim JO, et al. Diagnostic Value of Small Bowel Capsule Endoscopy in Isolated Ileitis: A CAPENTRY Study. *Dig Dis Sci*. janv 2017;62(1):180- 7.

21. González-Suárez B, Rodriguez S, Ricart E, Ordás I, Rimola J, Díaz-González Á, et al. Comparison of Capsule Endoscopy and Magnetic Resonance Enterography for the Assessment of Small Bowel Lesions in Crohn's Disease. *Inflamm Bowel Dis*. 19 mars 2018;24(4):775- 80.

22. Rondonotti E, Spada C, Adler S, May A, Despott EJ, Koulaouzidis A, et al. Small-bowel capsule endoscopy and device-assisted enteroscopy for diagnosis and treatment of small-bowel disorders: European Society of Gastrointestinal Endoscopy (ESGE) Technical Review. *Endoscopy*. 2018;50(4):423- 46.

23. ASGE Technology Committee, Wang A, Banerjee S, Barth BA, Bhat YM, Chauhan S, et al. Wireless capsule endoscopy. *Gastrointest Endosc*. déc 2013;78(6):805- 15.

24. Rahmi G, Cholet F, Gaudric M, Filippi J, Duburque C, Bramli S, et al. Effect of Different Modalities of Purgative Preparation on the Diagnostic Yield of Small Bowel Capsule for the Exploration of Suspected Small Bowel Bleeding: A Multicenter Randomized Controlled Trial. *Am J Gastroenterol*. 1 févr 2022;117(2):327- 35.
25. Yung DE, Plevris JN, Leenhardt R, Dray X, Koulaouzidis A, ESGE Small Bowel Research Working Group. Poor Quality of Small Bowel Capsule Endoscopy Images Has a Significant Negative Effect in the Diagnosis of Small Bowel Malignancy. *Clin Exp Gastroenterol*. 2020;13:475- 84.
26. Spada C, McNamara D, Despott EJ, Adler S, Cash BD, Fernández-Urién I, et al. Performance measures for small-bowel endoscopy: A European Society of Gastrointestinal Endoscopy (ESGE) Quality Improvement Initiative. *United Eur Gastroenterol J*. juin 2019;7(5):614- 41.
27. Dehaene S Le Cun, Yann, Girardon J. *La plus belle histoire de l'intelligence: des origines aux neurones artificiels : vers une nouvelle étape de l'évolution*. 2018.
28. Esteva A, Chou K, Yeung S, Naik N, Madani A, Mottaghi A, et al. Deep learning-enabled medical computer vision. *NPJ Digit Med*. 8 janv 2021;4(1):5.
29. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. *N Engl J Med*. 04 2019;380(14):1347- 58.
30. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 02 2017;542(7639):115- 8.
31. Wang P, Berzin TM, Glissen Brown JR, Bharadwaj S, Becq A, Xiao X, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut*. 27 févr 2019;
32. Beg S, Card T, Sidhu R, Wronska E, Ragunath K, UK capsule endoscopy users' group. The impact of reader fatigue on the accuracy of capsule endoscopy interpretation. *Dig Liver Dis Off J Ital Soc Gastroenterol Ital Assoc Study Liver*. août 2021;53(8):1028- 33.
33. Yung DE, Sykes C, Koulaouzidis A. The validity of suspected blood indicator software in capsule endoscopy: a systematic review and meta-analysis. *Expert Rev Gastroenterol Hepatol*. janv 2017;11(1):43- 51.
34. Saurin JC, Lapalus MG, Cholet F, D'Halluin PN, Filoche B, Gaudric M, et al. Can we shorten the small-bowel capsule reading time with the « Quick-view » image detection system? *Dig Liver Dis Off J Ital Soc Gastroenterol Ital Assoc Study Liver*. juin 2012;44(6):477- 81.
35. Leenhardt R, Fernandez-Urien Sainz I, Rondonotti E, Toth E, Van de Bruaene C, Baltes P, et al. PEACE: Perception and Expectations toward Artificial Intelligence in Capsule Endoscopy. *J Clin Med*. 6 déc 2021;10(23):5708.
36. Messmann H, Bisschops R, Antonelli G, Libânio D, Sinonquel P, Abdelrahim M, et al. Expected value of artificial intelligence in gastrointestinal endoscopy: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. *Endoscopy*. 21 oct 2022;

37. Leenhardt R, Koulaouzidis A, Histace A, Baatrup G, Beg S, Bourreille A, et al. Key research questions for implementation of artificial intelligence in capsule endoscopy. *Ther Adv Gastroenterol*. 2022;15:17562848221132684.
38. Leenhardt R, Li C, Le Mouel JP, Rahmi G, Saurin JC, Cholet F, et al. CAD-CAP: a 25,000-image database serving the development of artificial intelligence for capsule endoscopy. *Endosc Int Open*. mars 2020;8(3):E415- 20.
39. Rosa B, Margalit-Yehuda R, Gatt K, Sciberras M, Girelli C, Saurin JC, et al. Scoring systems in clinical small-bowel capsule endoscopy: all you need to know! *Endosc Int Open*. juin 2021;9(6):E802- 23.
40. Brotz C, Nandi N, Conn M, Daskalakis C, DiMarino M, Infantolino A, et al. A validation study of 3 grading systems to evaluate small-bowel cleansing for wireless capsule endoscopy: a quantitative index, a qualitative evaluation, and an overall adequacy assessment. *Gastrointest Endosc*. févr 2009;69(2):262- 70, 270.e1.
41. Dray X, Houist G, Le Mouel JP, Saurin JC, Vanbiervliet G, Leandri C, et al. Prospective evaluation of third-generation small bowel capsule endoscopy videos by independent readers demonstrates poor reproducibility of cleanliness classifications. *Clin Res Hepatol Gastroenterol*. 16 mars 2021;45(6):101612.
42. Leenhardt R, Souchaud M, Houist G, Le Mouel JP, Saurin JC, Cholet F, et al. A neural network-based algorithm for assessing the cleanliness of small bowel during capsule endoscopy. *Endoscopy*. 2 nov 2020;
43. Korman LY, Delvaux M, Gay G, Hagenmuller F, Keuchel M, Friedman S, et al. Capsule endoscopy structured terminology (CEST): proposal of a standardized and structured terminology for reporting capsule endoscopy procedures. *Endoscopy*. oct 2005;37(10):951-9.
44. Cortegoso Valdivia P, Deding U, Bjørsum-Meyer T, Baatrup G, Fernández-Urién I, Dray X, et al. Inter/Intra-Observer Agreement in Video-Capsule Endoscopy: Are We Getting It All Wrong? A Systematic Review and Meta-Analysis. *Diagn Basel Switz*. 2 oct 2022;12(10):2400.
45. Leenhardt R, Li C, Koulaouzidis A, Cavallaro F, Cholet F, Eliakim R, et al. Nomenclature and semantic description of vascular lesions in small bowel capsule endoscopy: an international Delphi consensus statement. *Endosc Int Open*. mars 2019;7(3):E372- 9.
46. Leenhardt R, Buisson A, Bourreille A, Marteau P, Koulaouzidis A, Li C, et al. Nomenclature and semantic descriptions of ulcerative and inflammatory lesions seen in Crohn's disease in small bowel capsule endoscopy: An international Delphi consensus statement. *United Eur Gastroenterol J*. 23 déc 2019;205064061989586.
47. Baltes P, Dray X, Riccioni ME, Pérez-Cuadrado-Robles E, Fedorov E, Wiedbrauck F, et al. Small-bowel capsule endoscopy in patients with Meckel's diverticulum: clinical features, diagnostic workup, and findings. A European multicenter I-CARE study. *Gastrointest Endosc*. 23 déc 2022;S0016-5107(22)02224-6.
48. Elli L, Marinoni B, Sidhu R, Bojarski C, Branchi F, Tontini GE, et al. Nomenclature and Definition of Atrophic Lesions in Small Bowel Capsule Endoscopy: A Delphi Consensus

Statement of the International Capsule endoscopy REsearch (I-CARE) Group. *Diagn BaselSwitz*. 13 juill 2022;12(7):1704.

49. Sciberras M, Conti K, Elli L, Scaramella L, Riccioni ME, Marmo C, et al. Score reproducibility and reliability in differentiating small bowel subepithelial masses from innocent bulges. *Dig Liver Dis Off J Ital Soc Gastroenterol Ital Assoc Study Liver*. oct 2022;54(10):1403-9.

50. Leenhardt R, Koulaouzidis A, McNamara D, Keuchel M, Sidhu R, McAlindon ME, et al. A guide for assessing the clinical relevance of findings in small bowel capsule endoscopy: analysis of 8064 answers of international experts to an illustrated script questionnaire. *Clin Res Hepatol Gastroenterol*. 28 janv 2021;45(6):101637.

51. Becq A, Rahmi G, Perrod G, Cellier C. Hemorrhagic angiodysplasia of the digestive tract: pathogenesis, diagnosis, and management. *Gastrointest Endosc*. 2017;86(5):792-806.

52. Leenhardt et al. A highly sensitive and highly specific convolutional neural network-based algorithm for automated diagnosis of angiodysplasia in small bowel capsule endoscopy. *DDW 2018 ASGE Abstracts*.

53. Leenhardt R, Vasseur P, Li C, Saurin JC, Rahmi G, Cholet F, et al. A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. *Gastrointest Endosc*. 2019;89(1):189-94.

54. Byrne MF, Donnellan F. Artificial intelligence and capsule endoscopy: Is the truly « smart » capsule nearly here? *Gastrointest Endosc*. 2019;89(1):195-7.

55. Trophée de l'Interne Innovant: Les Rencontres de Transfert de Technologies de l'AP-HP 5 juin 2018.

56. Houdeville C, Souchaud M, Leenhardt R, Beaumont H, Benamouzig R, McAlindon M, et al. A multisystem-compatible deep learning-based algorithm for detection and characterization of angiectasias in small-bowel capsule endoscopy. A proof-of-concept study. *Dig Liver Dis Off J Ital Soc Gastroenterol Ital Assoc Study Liver*. déc 2021;53(12):1627-31.

57. Leenhardt et al. Intelligence artificielle : actualités 2022 grêle et côlon. Communication orale, Vidéodigest nov 2022.

58. FDA OKs AI Device to Detect Diabetic Retinopathy in Primary Care [Internet]. *Medscape*. [cité 8 févr 2019]. Disponible sur: <http://www.medscape.com/viewarticle/895061>

59. Sidhu R, Chetcuti Zammit S, Baltés P, Carretero C, Despott EJ, Murino A, et al. Curriculum for small-bowel capsule endoscopy and device-assisted enteroscopy training in Europe: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. *Endoscopy*. août 2020;52(8):669-86.

60. Aoki T, Yamada A, Aoyama K, Saito H, Tsuboi A, Nakada A, et al. Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network. *Gastrointest Endosc*. 2019;89(2):357-363.e2.

61. Saito H, Aoki T, Aoyama K, Kato Y, Tsuboi A, Yamada A, et al. Automatic detection and classification of protruding lesions in wireless capsule endoscopy images based on a deep convolutional neural network. *Gastrointest Endosc.* 18 févr 2020;
62. Ding Z, Shi H, Zhang H, Meng L, Fan M, Han C, et al. Gastroenterologist-level Identification of Small Bowel Diseases and Normal Variants by Capsule Endoscopy Using a Deep-learning Model. *Gastroenterology.* 25 juin 2019;
63. Aoki T, Yamada A, Kato Y, Saito H, Tsuboi A, Nakada A, et al. Automatic detection of various abnormalities in capsule endoscopy videos by a deep learning-based system: a multicenter study. *Gastrointest Endosc.* janv 2021;93(1):165-173.e1.
64. Ding Z, Shi H, Zhang H, Zhang H, Tian S, Zhang K, et al. Artificial intelligence-based diagnosis of abnormalities in small-bowel capsule endoscopy. *Endoscopy.* janv 2023;55(1):44- 51.
65. Mascarenhas Saraiva MJ, Afonso J, Ribeiro T, Ferreira J, Cardoso H, Andrade AP, et al. Deep learning and capsule endoscopy: automatic identification and differentiation of small bowel lesions with distinct haemorrhagic potential using a convolutional neural network. *BMJ Open Gastroenterol.* sept 2021;8(1):e000753.
66. Xie X, Xiao YF, Zhao XY, Li JJ, Yang QQ, Peng X, et al. Development and Validation of an Artificial Intelligence Model for Small Bowel Capsule Endoscopy Video Review. *JAMA Netw Open.* 1 juill 2022;5(7):e2221992.
67. Piccirelli S, Mussetto A, Bellumat A, Cannizzaro R, Pennazio M, Pezzoli A, et al. New Generation Express View: An Artificial Intelligence Software Effectively Reduces Capsule Endoscopy Reading Times. *Diagn Basel Switz.* 22 juill 2022;12(8):1783.
68. Uzan J, et al. Actualités en capsule endoscopique, Communication orale, Vidéodigest 2022.
69. Pan J, Xia J, Jiang B, Zhang H, Zhang H, Li ZS, et al. Real-time identification of gastric lesions and anatomical landmarks by artificial intelligence during magnetically controlled capsule endoscopy. *Endoscopy.* nov 2022;54(11):E622- 3.
70. Moen S, Vuik FER, Kuipers EJ, Spaander MCW. Artificial Intelligence in Colon Capsule Endoscopy-A Systematic Review. *Diagn Basel Switz.* 17 août 2022;12(8):1994.
71. Sahafi A, Wang Y, Rasmussen CLM, Bollen P, Baatrup G, Blanes-Vidal V, et al. Edge artificial intelligence wireless video capsule endoscopy. *Sci Rep.* 12 août 2022;12:13723.
72. Wachter R. New York, NY: McGraw-Hill; 2015. ISBN: 9780071849463.

Contributions scientifiques

Yung DE, Plevris JN, Leenhardt R, Dray X, Koulaouzidis A; ESGE Small Bowel Research Working Group. Poor Quality of Small Bowel Capsule Endoscopy Images Has a Significant Negative Effect in the Diagnosis of Small Bowel Malignancy. *Clin Exp Gastroenterol*. 2020 Oct 21;13:475-484. doi: 10.2147/CEG.S260906. PMID: 33116745; PMCID: PMC7586059.

Dray X, Houist G, Le Mouel JP, Saurin JC, Vanbiervliet G, Leandri C, Rahmi G, Duburque C, Kirchgessner J, Leenhardt R, Cholet F, Quénéhervé L. Prospective evaluation of third-generation small bowel capsule endoscopy videos by independent readers demonstrates poor reproducibility of cleanliness classifications. *Clin Res Hepatol Gastroenterol*. 2021 Nov;45(6):101612. doi: 10.1016/j.clinre.2020.101612. Epub 2021 Mar 16. PMID: 33740610.

Houdeville C, Leenhardt R, Souchaud M, Velut G, Carbonell N, Nion-Larmurier I, Nuzzo A, Histace A, Marteau P, Dray X. Evaluation by a Machine Learning System of Two Preparations for Small Bowel Capsule Endoscopy: The BUBS (Burst Unpleasant Bubbles with Simethicone) Study. *J Clin Med*. 2022 May 17;11(10):2822. doi: 10.3390/jcm11102822. PMID: 35628947; PMCID: PMC9146856.

Elli L, Marinoni B, Sidhu R, Bojarski C, Branchi F, Tontini GE, Chetcuti Zammit S, Khater S, Eliakim R, Rondonotti E, Saurin JC, Bruno M, Buchkremer J, Cadoni S, Cavallaro F, Dray X, Ellul P, Urien IF, Keuchel M, Kopylov U, Koulaouzidis A, Leenhardt R, Baltes P, Beaumont H, Marmo C, McNamara D, Mussetto A, Nemeth A, Cuadrado Robles EP, Perrod G, Rahmi G, Riccioni ME, Robertson A, Spada C, Toth E, Triantafyllou K, Wurm Johansson G, Rimondi A; International CAPsule endoscopy REsearch (I-CARE) Group. Nomenclature and Definition of Atrophic Lesions in Small Bowel Capsule Endoscopy: A Delphi Consensus Statement of the International CAPsule endoscopy REsearch (I-CARE) Group. *Diagnostics (Basel)*. 2022 Jul 13;12(7):1704. doi: 10.3390/diagnostics12071704. PMID: 35885608; PMCID: PMC9325291.

Sciberras M, Conti K, Elli L, Scaramella L, Riccioni ME, Marmo C, Cadoni S, McAlindon M, Sidhu R, O'hara F, McNamara D, Rondonotti E, Piccirelli S, Spada C, Bruno M, Keuchel M, Baltes P, Calleja N, Valdivia PC, de'Angelis GL, Margalit-Yehuda R, Koulaouzidis A, Dray X, Ellul P; International CAPsule endoscopy REsearch (I-CARE) Group. Score reproducibility and reliability in differentiating small bowel subepithelial masses from innocent bulges. *Dig Liver Dis*. 2022 Oct;54(10):1403-1409. doi: 10.1016/j.dld.2022.06.027. Epub 2022 Aug 5. PMID: 35934647.

Baltes P, Dray X, Riccioni ME, Pérez-Cuadrado-Robles E, Fedorov E, Wiedbrauck F, Chetcuti

Zammit S, Cadoni S, Bruno M, Rondonotti E, Johansson GW, Mussetto A, Beaumont H, Perrod G, McNamara D, Plevris J, Spada C, Pinho R, Rosa B, Hervas N, Leenhardt R, Marmo C, Esteban-Delgado P, Ivanova E, Keuchel M; of the International Capsule Endoscopy Research (I-CARE) Meckel study group. Small-bowel capsule endoscopy in patients with Meckel's diverticulum: clinical features, diagnostic workup, and findings. A European multicenter I-CARE study. *Gastrointest Endosc.* 2022 Dec 23:S0016-5107(22)02224-6. doi: 10.1016/j.gie.2022.12.014. Epub ahead of print. PMID: 36572128.

Communications orales

Nomenclature and Semantic Description of Vascular Lesions in Small Bowel Capsule Endoscopy: an International Delphi Consensus Statement, UEGW 2017 Barcelone

Un algorithme hautement sensible et hautement spécifique pour la détection des angiodysplasies de l'intestin grêle en VCE, JFHOD 2018 Paris

A Neural Network Algorithm for Detection of Gastrointestinal Angiodysplasia During SB-CE, DDW 2018, Washington

Intelligence Artificielle : actualités 2022 grêle et côlon, Vidéo-digest, nov 2022

Posters

Romain Leenhardt, Marc Souchaud, Guy Houist, J.L. Lemouel, J.C. Saurin, Franck Cholet, Gabriel Rahmi, Céline Leandri, Aymeric Histace, Xavier Dray

A Highly Sensitive Neural Network-Based algorithm For Assessing the Cleanliness of Small Bowel During Capsule Endoscopy

UEG Week, Oct 2020, Virtual, France

Romain Leenhardt, Marc Souchaud, Aymeric Becq, Aymeric Histace, Xavier Dray

Artificial Intelligence Alleviates The Tedious Task of Reading Small Bowel Capsule Endoscopy Recordings

European Society of Gastrointestinal Endoscopy Days (ESGE Days), Apr 2020, Dublin, Ireland. pp.S235-S236, <10.1055/s-0040-1704737>

DOI : 10.1055/s-0040-1704737

Brevets

Marc Souchaud, Aymeric Histace, Romain Leenhardt, Xavier Dray

Device and method for producing a digital video classifier

France, Patent n° : EP3977909A1. 2022

Xavier Dray, Romain Leenhardt, Aymeric Histace

Methods and devices for calculating a level of "clinical relevance" for abnormal small bowel findings captured by capsule endoscopy video

United States, Patent n°: US20220039639A1. 2022

Ouvrages

Xavier Dray, Aymeric Histace, Romain Leenhardt

L'IA appliquée à la vidéocapsule endoscopique de l'intestin grêle : état de l'art et perspectives

E. Coron, G. Vanbiervliet. Intelligence artificielle et maladie digestive, Le Grand Métier, pp.33-45, 2021