

À PROPOS DES AUTEURS



Amine Zoughlami, M.D., MA

Le Dr Amine Zoughlami effectue sa cinquième année de résidence en gastro-entérologie pour adultes à l'Université McGill. Il a accompli sa formation médicale et en médecine interne à l'Université McGill, et a obtenu une maîtrise en bioéthique à l'Université de Montréal. Il va bientôt terminer un programme de bourse de recherche avancée sur les MII à l'Université Western et porte un intérêt à l'innovation médicale et à la mise en œuvre d'initiatives technologiques en gastro-entérologie.

Affiliation de l'auteur : *Division de gastro-entérologie et d'hépatologie, département de médecine, Centre de santé de l'Université McGill, Montréal (Québec), Canada*



Adel Arezki, M.D.

Le Dr Adel Arezki effectue sa quatrième année de résidence en urologie à l'Université McGill. Il a accompli sa formation médicale à l'Université McGill, durant laquelle ses intérêts de recherche se situaient à l'intersection de l'urologie et de la technologie. Ses travaux se concentrent sur l'intelligence artificielle, la modélisation des prédictions cliniques et les approches fondées sur les données pour améliorer la précision des diagnostics et les résultats des traitements en médecine.

Affiliation de l'auteur : *Division d'urologie, département de chirurgie, Centre de santé de l'Université McGill, Montréal (Québec), Canada*



Edgard Medawar, M.D.

Le Dr Edgard Medawar a suivi sa formation médicale à l'Université McGill. Il effectue actuellement sa troisième année de résidence en médecine interne à l'Université d'Ottawa et termine un doctorat en médecine expérimentale à l'Université de Montréal. L'année prochaine, il commencera une sous-spécialisation en gastro-entérologie pour adultes à l'Université de Montréal. Edgard porte un intérêt aux maladies inflammatoires de l'intestin et à l'entérectomie endoscopique.

Affiliations de l'auteur : *Département de médecine, Université d'Ottawa
Division de gastro-entérologie, Centre de Recherche du Centre Hospitalier de l'Université de Montréal, Montreal, Quebec*



Talat Bessissow, M.D., M. Sc., MDCM, FRCPC

Après avoir obtenu son diplôme de docteur en médecine à l'Université McGill (2005), Dr Talat Bessissow a suivi une formation de troisième cycle en médecine interne et en gastro-entérologie (2005-2010). En 2012, il s'est spécialisé dans les maladies inflammatoires de l'intestin et l'imagerie endoscopique de pointe à l'hôpital universitaire Gasthuisberg, à Louvain, en Belgique, sous la supervision de la professeure Séverine Vermeire. Il a également obtenu une maîtrise en médecine expérimentale et en épidémiologie à l'Université McGill en 2016. Depuis 2012, il est professeur agrégé à temps plein à la division de gastro-entérologie et médecin traitant au Centre de santé de l'Université McGill. Il est membre du Groupe de recherche MII de McGill et participe au programme sur l'intestin grêle de McGill. Ses recherches actuelles se concentrent sur le rôle et les résultats de la cicatrisation des muqueuses dans les maladies inflammatoires de l'intestin, ainsi que sur la détection précoce des lésions néoplasiques dans la colite ulcéreuse. Ses recherches lui ont permis de publier plus de 150 articles complets évalués par des pairs. Il est l'ancien président du Consortium canadien de recherche sur les MII. Il a également occupé le poste d'examineur pour plusieurs revues nationales et internationales.

Affiliation de l'auteur : Division de gastro-entérologie et d'hépatologie, département de médecine, Centre de santé de l'Université McGill, Montréal (Québec), Canada

Diagnostic par l'intelligence artificielle dans les maladies inflammatoires de l'intestin : applications et orientations futures

Amine Zoughlami, M.D., MA

Adel Arezki, M.D.

Edgard Medawar, M.D.

Talat Bessissow, M.D., M. Sc., MDCM, FRCPC

Principaux points à retenir

L'IA peut améliorer la précision, l'objectivité et la reproductibilité des évaluations des MII à travers de nombreux indices d'évaluation de la maladie.

Plusieurs modèles d'IA ont fait preuve de performances dont le niveau était celui d'un expert dans l'évaluation de l'activité des MII sur le plan endoscopique et histologique.

Le déploiement de modèles d'IA peut contribuer à uniformiser la qualité de l'évaluation des maladies dans les centres universitaires comme dans les centres communautaires.

Les prochaines étapes feront intervenir des modèles d'IA multimodaux. La mise au point de ces modèles et des systèmes unimodaux nécessitera des ensembles de données importants et diversifiés, ainsi qu'une gestion rigoureuse.

Introduction

La prise en charge des maladies inflammatoires de l'intestin (MII) repose sur des indices cliniques, endoscopiques et histologiques qui permettent d'évaluer l'activité de la maladie et d'orienter le traitement. Dans la pratique, les cliniciens utilisent de nombreux points de données pour formuler un plan de traitement. Les progrès accomplis dans l'intelligence artificielle (IA) offrent une occasion unique d'intégrer ces données pour approfondir notre compréhension ainsi que notre évaluation de la maladie.

Le rôle des indices d'activité de la maladie est au centre de la stratégie de traitement ciblé recommandée par le consensus STRIDE-II.¹ Toutefois, les indices couramment utilisés posent souvent des difficultés telles que la subjectivité, le peu de fiabilité entre examinateurs, et la granularité limitée dans l'évaluation de la gravité ou des différences phénotypiques. Les méthodes de l'IA peuvent répondre à plusieurs de ces problèmes. Ce bref aperçu narratif présente les concepts fondamentaux de l'IA que les cliniciens sont susceptibles de rencontrer à l'avenir, et aborde les principales applications couvrant les évaluations cliniques, endoscopiques, histologiques et multimodales de la maladie.

Intelligence artificielle – de quoi s'agit-il?

L'IA désigne l'informatisation de tâches qui sinon nécessiteraient la cognition humaine, telles que la reconnaissance des formes, la résolution de problèmes et la prise de décision. L'apprentissage automatique, un sous-ensemble de l'IA, fait référence à des modèles qui apprennent directement à partir de données au lieu d'être explicitement programmés pour le faire. L'apprentissage profond, un sous-ensemble de l'apprentissage automatique, utilise un grand nombre de couches de réseaux neuronaux pour apprendre des schémas complexes.

Classiquement, les modèles d'IA sont entraînés de manière supervisée ou non supervisée. Dans l'apprentissage supervisé, le modèle apprend à partir de données étiquetées. Par exemple, on montre à un modèle le cliché d'un ulcère qui serait étiqueté comme tel dans le contexte des MII. Dans l'apprentissage non supervisé, les modèles reconnaissent d'eux-mêmes les schémas dans les données. Les modèles sont entraînés sur un seul ensemble de données et testés sur un autre. La

généralisabilité fait référence à la capacité d'un modèle à maintenir ses performances lorsqu'il est appliqué à de nouvelles données. Les principaux obstacles à la performance des modèles sont le sur-apprentissage (lorsqu'un modèle apprend à partir de données de test, mais perd ses capacités en présence de nouvelles données) et le sous-apprentissage (lorsqu'un modèle n'est pas suffisamment exhaustif pour capturer les schémas, ce qui entraîne de mauvaises performances tant pour les ensembles de données d'apprentissage que les ensembles de données utilisés pour les tests). Le sur-apprentissage peut se produire dans des contextes où les données d'entraînement diffèrent radicalement des données de test, telles que des différences dans les modèles d'endoscopes, la qualité des images ou un mélange de cas de patients. Le sur-apprentissage peut être minimisé par des stratégies telles que l'utilisation de divers ensembles de données en plus de l'apprentissage fédéré, où les modèles sont entraînés localement et agrégés de manière centralisée.

Les réseaux neuronaux sont une classe d'algorithmes d'apprentissage automatique inspirés de la structure interconnectée des neurones dans le cerveau. Ils se composent de nombreuses couches, dont une couche d'entrée de données, une ou plusieurs couches de traitement et une couche de sortie de données. Au fur et à mesure que le réseau neuronal analyse les données, la force de la connexion entre les nœuds varie pour améliorer la qualité des données de sortie. Parmi les réseaux neuronaux, les réseaux neuronaux convolutifs sont plus couramment utilisés pour le traitement des images et des vidéos et sont largement appliqués dans les tâches endoscopiques telles que la détection des polypes. Le traitement du langage naturel fait également appel aux réseaux neuronaux pour faciliter la compréhension et la génération informatisées du langage humain. Une application du traitement du langage naturel est la mise au point de grands modèles de langage, qui sont entraînés sur de vastes ensembles de données pour prédire et générer un langage conversationnel, tel que ChatGPT (OpenAI, San Francisco, États-Unis).

Ces concepts constituent un bref aperçu des principes fondamentaux de l'IA. Ensemble, ces méthodes sous-tendent les applications du domaine des MII présentées dans cet article.

Activité clinique de la maladie dans les MII – une partie seulement de la réponse

Les indices cliniques, tels que l'indice d'activité de la maladie de Crohn (CDAI), l'indice de Harvey-Bradshaw (HBI) et le score partiel de Mayo (pMS) sont largement utilisés pour évaluer l'activité de la maladie, mais chacun d'entre eux contient des éléments subjectifs. Le CDAI est sensible à la fiabilité entre examinateurs, au moins partiellement, car il repose sur une évaluation subjective dans le contexte d'éléments clés, tels que le « bien-être général »,² et il peut être fortement affecté par un biais de rappel. Bien que plus simples et plus faciles à utiliser, l'HBI et le pMS reposent également partiellement sur des éléments subjectifs. De plus, plusieurs éléments de ces indices peuvent être confondus avec des affections telles que le syndrome du côlon irritable, qui présente certaines similitudes avec une MII chez 7 à 25 % des patients.^{3,4} Ces limitations mettent en relief le rôle potentiel de l'IA pour compléter l'évaluation des symptômes grâce à l'intégration de données provenant de différentes sources et l'utilisation de mesures continues et objectives.

En dehors d'un rendez-vous à la clinique, les patients communiquent souvent avec leur médecin traitant par téléphone, par courriel ou par le biais d'un portail patient. L'IA peut être utilisée pour détecter une maladie active au cours de ces communications. Par exemple, une étude récente a appliqué le traitement du langage naturel à un forum en ligne sur les MII et a déterminé 20 marqueurs de substitution d'exacerbations cliniques à partir du langage des patients.⁵ Cette étude souligne le potentiel du traitement du langage naturel pour analyser d'autres sources de données provenant des patients, telles que les communications par le biais de messages, de courriels et du portail des patients. L'examen des interactions cliniques qui se déroulent entre-temps est également très chronophage. Les grands modèles de langage ont démontré leur capacité à extraire les résultats déclarés par les patients à partir des données cliniques relatives aux MII,⁶ et les systèmes d'examen des dossiers fondés sur l'IA peuvent reconnaître très précisément les manifestations extra-intestinales dans les notes cliniques relatives aux MII.⁷ Des systèmes similaires peuvent être utilisés pour diminuer le temps que passent les cliniciens à examiner les dossiers, et mettre efficacement en évidence les changements pertinents survenus entre les visites.

Toutefois, une mise en garde importante concerne la qualité des données d'entrée : le transfert de notes, les dossiers incomplets ou l'absence de quantification contribuent tous à une classification erronée et à un manque de précision.

L'utilisation nouvelle de capteurs de santé portables est un terrain enthousiasmant pour l'évaluation clinique de l'activité de la maladie. Dans une étude menée auprès d'une cohorte de 309 patients équipés de dispositifs portables grand public, les données physiologiques, notamment la fréquence cardiaque (FC), la FC au repos, la variabilité de la FC et la saturation en oxygène, ont été associées à des enquêtes sur les symptômes quotidiens et à des marqueurs biochimiques.⁸ Les modèles d'apprentissage automatique ont eu la capacité de prédire les exacerbations (définies comme étant des symptômes accompagnés de signes biochimiques tels que le taux de calprotectine fécale, le taux de protéine C-réactive et la vitesse de sédimentation des érythrocytes) jusqu'à 49 jours avant leur survenue. Le recueil continu de données par le biais de dispositifs portables peut permettre l'identification rapide des patients exposés à un risque d'exacerbations, et par conséquent la réalisation de tests plus précoces et d'évaluations proactives. Quoique les données physiologiques non invasives provenant des dispositifs portables offrent des renseignements précieux grâce à l'apprentissage automatique, les données biochimiques acquises par les dispositifs portables sont encore plus prometteuses. Les progrès futurs de la technologie des dispositifs portables pourraient permettre une détection en temps réel des données biochimiques. Un dispositif portable non invasif reposant sur la transpiration a été récemment mis au point et permet de mesurer les taux de calprotectine, d'interleukine-6 et de protéine C-réactive dans la sueur.⁹ Dans le cadre de l'étude, le capteur a été en mesure de faire la distinction entre les patients atteints de colite ulcéreuse (CU) endoscopiquement active et les patients chez qui la CU était inactive, sur la base des taux de calprotectine dans la sueur. Les mesures de chaque marqueur reposant sur la transpiration ont également montré des corrélations modérées à fortes avec les taux sériques correspondants. Bien que la validation longitudinale soit en attente, cette validation de principe laisse présager un avenir prometteur où les signes de l'inflammation détectés en temps réel pourront faciliter un tri rapide des patients,

ainsi que des évaluations et des modifications de traitement en temps utile.

Ces innovations offrent la possibilité de faire passer le suivi des maladies en temps réel d'un modèle périodique, à des moments donnés, à un modèle proactif dans lequel les changements et les discussions peuvent avoir lieu avant la survenue d'une modification notable de l'état clinique.

Évaluation endoscopique dans les MII – comment faire mieux?

L'évaluation endoscopique de l'activité de la maladie dans les MII repose largement sur le score endoscopique simplifié de la MC (SES-CD) et le score endoscopique de Mayo (MES). Le score SES-CD évalue les ulcères, les régions affectées et les sténoses dans tous les segments et s'est avéré très fiable entre les évaluateurs centralisés dans les études cliniques.¹⁰ Toutefois, sa généralisabilité et son adoption dans la pratique communautaire restent incertaines. Par contre, le score MES offre une approche plus simple, mais peut manquer de précision, car il repose sur des seuils subjectifs, tels que la distinction entre fragilité légère et fragilité de la muqueuse. Les modèles d'évaluation de l'activité endoscopique assistés par l'IA peuvent fournir un indice systématique et reproductible de l'activité endoscopique de la maladie.

Lors d'une étude menée par Gottlieb et al., 795 vidéos endoscopiques provenant d'une étude de phase II sur le mirikizumab dans la CU ont été évaluées de manière centralisée par un seul examinateur à l'aide du score MES et de l'indice endoscopique de gravité de la colite ulcéreuse (UCEIS), puis elles ont été analysées au moyen d'un modèle d'apprentissage profond qui a montré une forte concordance entre les deux indices.¹¹ De même, Fan et al. ont entraîné un modèle de système de notation IA pour la CU à l'aide d'images fixes, et l'ont testé sur 20 vidéos endoscopiques de toute la longueur de l'intestin qu'ils ont divisé en cinq segments.¹² Le modèle a permis d'obtenir une concordance dans 83 % des segments présentant une maladie active et dans 100 % des segments où la maladie était inactive, et a généré des cartes du côlon colorées, un outil graphique intuitif représentant la gravité de la maladie. Ces résultats suggèrent que les algorithmes d'apprentissage profond peuvent détecter et distinguer les maladies actives à un niveau qui est celui d'un expert.

La granularité demeure une difficulté dans la notation de la CU. Dans une étude menée en 2023 par Kim et al. qui portait sur une cohorte de 492 patients atteints de CU ayant démontré des améliorations endoscopiques du score MES de 1 à 0, les évaluations endoscopiques de l'activité de la maladie réalisées par des gastro-entérologues ont été comparées à celles d'un algorithme d'apprentissage profond.¹³ Les résultats montrent que le modèle a surpassé le consensus d'un groupe de gastro-entérologues, et a fourni des résultats plus précis avec une capacité supérieure à distinguer les scores MES 0 et 1. L'algorithme a notamment maintenu son niveau de performance sur un ensemble de données externes. Ces résultats sont à l'appui de l'utilisation de l'IA comme un outil d'appoint pour améliorer la notation dans la CU et permettre une discrimination subtile près des seuils cliniques.

Dans la MC, le score SES-CD repose sur des évaluations sujettes à la partialité, telles que la taille de l'ulcère et la surface affectée. Une variabilité marquée entre les examinateurs peut être notée dans des sous-scores spécifiques, même chez des spécialistes de la gastro-entérologie.¹⁴ Un modèle d'IA entraîné pour évaluer l'ulcération dans la MC a présenté une forte corrélation avec le score SES-CD total, une corrélation modérée, mais significative, avec les taux de calprotectine fécale et, fait important, a montré une capacité à détecter une rémission clinique supérieure à celle du SES-CD.¹⁵ Pour l'évaluation de l'intestin grêle, les applications de l'IA aux techniques de vidéo-capsule endoscopique ont rapidement progressé et ont surpassé les gastro-entérologues dans la détection des saignements ainsi que dans les délais de lecture des images.¹⁶ Dans les MII, la détection assistée par ordinateur des érosions et des ulcères a atteint une sensibilité et une spécificité supérieure à 90 %, ¹⁷ avec une excellente discrimination entre les ulcères superficiels et les ulcères graves.¹⁸ Plus récemment, un score généré par l'IA pour évaluer la gravité de la MC dans l'intestin grêle a montré une forte corrélation avec le score de Lewis.¹⁹

Malgré les excellents résultats, l'hétérogénéité demeure. Une méta-analyse récente a révélé une variabilité marquée de la précision de l'IA pour évaluer la cicatrisation de la muqueuse dans la CU à travers les ensembles de données,²⁰ soulignant ainsi le besoin d'un entraînement normalisé des algorithmes et d'une validation externe plus poussée.

Évaluation histologique dans les MII – l'IA comme grand égaliseur?

La rémission histologique est de plus en plus reconnue comme un objectif thérapeutique potentiel dans les MII, en particulier dans la CU. Toutefois, l'évaluation histologique représente beaucoup de travail et nécessite une expertise de sous-spécialité, ce qui limite son adoption généralisée. Najdawi et al. ont entraîné une série de réseaux neuronaux convolutifs pour reconnaître les tissus et les cellules et générer un ensemble de caractéristiques interprétables par l'homme pour prédire les résultats, notamment la densité cellulaire et les zones de tissus affectées.²¹ Dans cet ensemble, 13 caractéristiques sélectionnées par consensus ont permis aux experts de prédire très exactement les résultats qui montraient une forte corrélation avec l'indice histologique de Nancy, et d'obtenir une précision de 97 % dans la détection de la rémission histologique. Notamment, la concordance du modèle avec les pathologistes gastro-intestinaux était comparable au consensus entre les pathologistes, ce qui indique une performance dont le niveau est celui d'un expert.

L'évaluation histologique assistée par l'IA peut également prédire les résultats cliniques. En utilisant l'indice de rémission histologique PICaSSO (PHRI), un modèle d'IA a été capable de prédire la rechute clinique, avec des performances similaires à celles de l'évaluation d'un spécialiste en pathologie, les résultats générés par l'IA étant obtenus en seulement 9,8 secondes.²²

Ces résultats illustrent comment l'IA peut démocratiser l'expertise histologique, en particulier dans les milieux communautaires où le domaine de la pathologie dédié à la gastro-entérologie peut être limité.

L'IA dans les MII – rassembler le tout

La prise de décision dans les MII est par nature multimodale, et l'IA est utile pour interpréter des signaux hétérogènes. Chen et al. ont mis au point un outil d'aide à la décision clinique qui utilise uniquement les formules sanguines complètes pour prédire de manière non invasive l'étendue et la gravité de l'inflammation colique. L'aire sous la courbe des caractéristiques de fonctionnement du récepteur atteint 0,81 lors de la distinction entre une colite étendue et une proctite sur des ensembles de données de validation externes.²³ Des points de données supplémentaires peuvent

être intégrés pour évaluer la gravité de la maladie avec plus de certitude et de granularité. L'intégration de plusieurs paramètres d'activité de la maladie en un seul indice a été reconnue comme potentiellement utile, surtout pour augmenter la sensibilité à la réponse thérapeutique dans les études comportant des échantillons de petite taille.²⁴ L'intégration de données multimodales avec l'apprentissage automatique a également été appliquée aux profils d'expression génique pour prédire la réponse clinique aux traitements innovants,²⁵ ou aux modèles intégrant les antécédents cliniques et les données biochimiques pour prédire le risque chirurgical lié à la MC à un an.²⁶ Ces efforts soulignent le potentiel de l'IA dans l'optimisation du choix du traitement et du pronostic.

Les données d'une étude de phase II sur le mirikizumab dans la CU ont été utilisées dans le cadre d'une fusion de modèles d'IA qui combinait les entrées de données endoscopiques et histologiques. Ce modèle fusionné a surpassé les modèles individuels d'IA unimodale dans la prédiction de la rémission histologique.²⁷ Cette étude fournit une validation de principe importante à l'appui de l'utilisation de l'IA pour intégrer de nombreuses données sur l'activité de la maladie de façon à mieux prédire la cicatrisation. Les recherches futures devraient explorer l'application de la fusion de modèles pour prédire les résultats cliniques et endoscopiques.

Et maintenant, où allons-nous?

Le rôle croissant de l'IA dans les MII est extrêmement prometteur, mais sa mise en œuvre exigera collaboration et attention. Premièrement, l'utilisation d'ensembles de données diversifiés et multicentriques est une priorité pour se protéger contre le sur-apprentissage et améliorer la généralisabilité. *Gastronet-5M*, un ensemble de données endoscopiques accessible au public, compilé à partir de huit centres néerlandais utilisant différents systèmes d'endoscopie (Fuji, Olympus, Pentax), illustre la manière dont des ensembles de données d'entraînement diversifiés peuvent améliorer les performances des modèles dans une variété de tâches liées à l'endoscopie.²⁸

Deuxièmement, l'IA devrait augmenter, mais pas remplacer, le jugement clinique. Des données récentes ont montré une diminution des taux de détection des adénomes lors des coloscopies standard effectuées à la suite de coloscopies assistées par l'IA, ce qui semble indiquer un risque

de dépendance excessive vis-à-vis de l'IA.²⁹ Le maintien des compétences et de l'autonomie des cliniciens restera essentiel.

Troisièmement, la collaboration interdisciplinaire sera essentielle, car la recherche sur les MII reconnaît de plus en plus la valeur de l'évaluation transmurale et explore le potentiel des marqueurs moléculaires et génétiques. La mise en œuvre des outils d'IA devrait également inclure les centres communautaires, où l'expertise et le nombre de patients dans les services des MII peuvent être limités, ce qui contribuera à la normalisation des traitements.

Enfin, l'intégration de l'IA dans les soins des MII représente un changement de paradigme. S'ils sont mis en œuvre de manière responsable, ces outils apporteront un niveau d'objectivité et de reproductibilité indispensable à l'évaluation des maladies. La prochaine étape sera la validation prospective, à travers de vastes ensembles de données multicentriques. L'IA a le potentiel d'aider les gastro-entérologues à dispenser des soins plus précoces, plus précis et, surtout, équitables pour tous les patients atteints de MII.

Coordonnées

Talat Bessissow, M.D., M. Sc., MDCM, FRCPC
Courriel : talat.bessissow@mcgill.ca

Divulgence de renseignements financiers

A.Z. : Aucune déclaration.

A.A. : Aucune déclaration.

E.W. : Aucune déclaration.

T.B. : Aucune déclaration.

Références

1. Turner D, Ricciuto A, Lewis A, D'Amico F, Dhaliwal J, Griffiths AM, 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(5):1570–1583. doi:10.1053/j.gastro.2020.12.031
2. Sandborn WJ, Feagan BG, Hanauer SB, Lochs H, Lofberg R, Modigliani R, et al. A review of activity indices and efficacy endpoints for clinical trials of medical therapy in adults with Crohn's disease. *Gastroenterology*. 2002;122(2):512–530. doi:10.1053/gast.2002.31072
3. Burdine LK, Rakowsky S, Grossberg L, Rabinowitz L, Center BIR, Cheifetz AS, et al. Irritable bowel syndrome/inflammatory bowel disease overlap: less common than we think. *Gastro Hep Adv*. 2024;3(8):1135–1137. Published 2024 Aug 10. doi:10.1016/j.gastha.2024.08.005
4. Fairbrass KM, Costantino SJ, Gracie DJ, Ford AC. Prevalence of irritable bowel syndrome-type symptoms in patients with inflammatory bowel disease in remission: a systematic review and meta-analysis. *Lancet Gastroenterol Hepatol*. 2020;5(12):1053–1062. doi:10.1016/S2468-1253(20)30300-9
5. Ricci L, Toussaint Y, Becker J, Najjar H, Renier A, Choukour M, et al. Web-based and machine learning approaches for identification of patient-reported outcomes in inflammatory bowel disease. *Dig Liver Dis*. 2022;54(4):483–489. doi:10.1016/j.dld.2021.09.005
6. Patel PV, Davis C, Ralbovsky A, Tinoco D, Williams CYK, Slatter S, et al. Large language models outperform traditional natural language processing methods in extracting patient-reported outcomes in inflammatory bowel disease. *Gastro Hep Adv*. 2025;4(2):100563. Published 2024 Oct 10. doi:10.1016/j.gastha.2024.10.003
7. Stidham RW, Yu D, Zhao X, Bishu S, Rice M, Bourque C, et al. Identifying the Presence, Activity, and Status of Extraintestinal Manifestations of Inflammatory Bowel Disease Using Natural Language Processing of Clinical Notes. *Inflamm Bowel Dis*. 2023;29(4):503–510. doi:10.1093/ibd/izac109
8. Hirten RP, Danieleto M, Sanchez-Mayor M, Whang JK, Lee KW, Landell K, et al. Physiological data collected from wearable devices identify and predict inflammatory bowel disease flares. *Gastroenterology*. 2025;168(5):939–951.e5. doi:10.1053/j.gastro.2024.12.024
9. Shahub S, Kumar RM, Lin KC, Banga I, Choi NK, Garcia NM, et al. Continuous monitoring of CRP, IL-6, and calprotectin in inflammatory bowel disease using a perspiration-based wearable device. *Inflamm Bowel Dis*. 2025;31(3):647–654. doi:10.1093/ibd/izae054

10. Khanna R, Zou G, D'Haens G, Rutgeerts P, McDonald JW, Daperno M, et al. Reliability among central readers in the evaluation of endoscopic findings from patients with Crohn's disease. *Gut*. 2016;65(7):1119–1125. doi:10.1136/gutjnl-2014-308973
11. Gottlieb K, Requa J, Karnes W, Chandra Gudivada R, Shen J, Rael E, et al. Central reading of ulcerative colitis clinical trial videos using neural networks. *Gastroenterology*. 2021;160(3):710–719.e2. doi:10.1053/j.gastro.2020.10.024
12. Fan Y, Mu R, Xu H, Xie C, Zhang Y, Liu L, et al. Novel deep learning-based computer-aided diagnosis system for predicting inflammatory activity in ulcerative colitis. *Gastrointest Endosc*. 2023;97(2):335–346. doi:10.1016/j.gie.2022.08.015
13. Kim JE, Choi YH, Lee YC, Seong G, Song JH, Kim TJ, et al. Deep learning model for distinguishing Mayo endoscopic subscore 0 and 1 in patients with ulcerative colitis. *Sci Rep*. 2023;13(1):11351. Published 2023 Jul 13. doi:10.1038/s41598-023-38206-6
14. Khanna R, Feagan BG, Zou G, Stitt LW, McDonald JWD, Bressler B, et al. Reliability and responsiveness of clinical and endoscopic outcome measures in Crohn's disease. *Inflamm Bowel Dis*. 2025;31(3):706–715. doi:10.1093/ibd/izae089
15. Cai L, Wittrup E, Minoccheri C, Hiatt TK, Rice MD, Bishu S, et al. Artificial intelligence for quantifying endoscopic mucosal ulceration in Crohn's disease. *Clin Gastroenterol Hepatol*. Published online August 18, 2025. doi:10.1016/j.cgh.2025.05.026
16. Spada C, Piccirelli S, Hassan C, Ferrari C, Toth E, Gonzalez-Suarez B, et al. AI-assisted capsule endoscopy reading in suspected small bowel bleeding: a multicentre prospective study. *Lancet Digit Health*. 2024;6(5):e345–e353. doi:10.1016/S2589-7500(24)00048-7
17. Fan S, Xu L, Fan Y, Wei K, Li L. Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images. *Phys Med Biol*. 2018;63(16):165001. Published 2018 Aug 10. doi:10.1088/1361-6560/aad51c
18. Barash Y, Azaria L, Soffer S, Margalit Yehuda R, Shlomi O, Ben-Horin 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(1):187–192. doi:10.1016/j.gie.2020.05.066
19. Cardoso P, Mascarenhas M, Afonso J, Ribeiro T, Mendes F, Martins M, et al. Deep learning and minimally invasive inflammatory activity assessment: a proof-of-concept study for development and score correlation of a panendoscopy convolutional network. *Therap Adv Gastroenterol*. 2024;17:17562848241251569. Published 2024 May 27. doi:10.1177/17562848241251569
20. Rimondi A, Gottlieb K, Despott EJ, Iacucci M, Murino A, Tontini GE. Can artificial intelligence replace endoscopists when assessing mucosal healing in ulcerative colitis? A systematic review and diagnostic test accuracy meta-analysis. *Dig Liver Dis*. 2024;56(7):1164–1172. doi:10.1016/j.dld.2023.11.005
21. Najdawi F, Sucipto K, Mistry P, Hennek S, Jayson CKB, Lin M, et al. Artificial intelligence enables quantitative assessment of ulcerative colitis histology. *Mod Pathol*. 2023;36(6):100124. doi:10.1016/j.modpat.2023.100124
22. Iacucci M, Parigi TL, Del Amor R, Meseguer P, Mandelli G, Bozzola A, et al. Artificial intelligence enabled histological prediction of remission or activity and clinical outcomes in ulcerative colitis. *Gastroenterology*. 2023;164(7):1180–1188.e2. doi:10.1053/j.gastro.2023.02.031
23. Chen H, Lin X, Pan X, Xu H, Zhang X, Liang G, et al. Development and validation of a blood routine-based extent and severity clinical decision support tool for ulcerative colitis. *Sci Rep*. 2023;13(1):21368. Published 2023 Dec 4. doi:10.1038/s41598-023-48569-5
24. Jairath V, Jeyarajah J, Zou G, Parker CE, Olson A, Khanna R, et al. A composite disease activity index for early drug development in ulcerative colitis: development and validation of the UC-100 score. *Lancet Gastroenterol Hepatol*. 2019;4(1):63–70. doi:10.1016/S2468-1253(18)30306-6
25. He M, Li C, Tang W, Kang Y, Zuo Y, Wang Y. Machine learning gene expression predicting model for ustekinumab response in patients with Crohn's disease. *Immun Inflamm Dis*. 2021;9(4):1529–1540. doi:10.1002/iid3.506
26. Stidham RW, Liu Y, Enchakalody B, Van T, Krishnamurthy V, Su GL, et al. the use of readily available longitudinal data to predict the likelihood of surgery in Crohn disease. *Inflamm Bowel Dis*. 2021;27(8):1328–1334. doi:10.1093/ibd/izab035
27. Iacucci M, Santacroce G, Meseguer P, Dieguez A, Del Amor R, Kolawole BB, et al. Endo-Histo foundational fusion model: a novel artificial intelligence for assessing histologic remission and response to therapy in ulcerative colitis clinical trial. *J Crohns Colitis*. 2025;19(7):jjaf108. doi:10.1093/ecco-jcc/jjaf108
28. Jong MR, Boers TGW, Fockens KN, Jukema JB, Kusters CHJ, Jaspers TJM, et al. GastroNet-5M: a multicenter dataset for developing foundation models in gastrointestinal endoscopy. *Gastroenterology*. Published online July 30, 2025. doi:10.1053/j.gastro.2025.07.030
29. Budzyn K, Romanczyk M, Kitala D, Kolodziej P, Bugajski M, Adami HO, et al. Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study. *Lancet Gastroenterol Hepatol*. 2025;10(10):896–903. doi:10.1016/S2468-1253(25)00133-5