

Artificial intelligence in inflammatory bowel disease: current status and opportunities

Ji Li, Jia-Ming Qian

Department of Gastroenterology, Chinese Academy of Medical Sciences & Peking Union Medical College Hospital, Beijing 100730, China.

In the past decade, artificial intelligence (AI) has been applied in several clinical studies to improve the medical care of patients with gastroenterological diseases, to accurately detect polyps and early cancer lesions, to facilitate the analysis of inflammatory lesions, to assess liver fibrosis, and to predict the response and clinical outcomes of medications.^[1,2]

Inflammatory bowel disease (IBD), comprising Crohn disease (CD) and ulcerative colitis (UC), is a prototypical complex disease that has been considered a significant healthcare burden in the healthcare system of Western countries.^[3] Considered as one major form of AI, machine learning (ML) is an application that performs repeated iterations of models progressively improving the performance of a specific task by adapting several methods including random forest (RF), support vector machine (SVM), longitudinal regression (LR), and elastic net regularized generalized linear model. ML has been used in several clinical studies comprising IBD patients to determine the differential diagnosis of IBD, to assess IBD, and to predict the response and clinical outcomes of medications used to treat IBD. However, ML-assisted diagnostic or predictive measures for the daily medical care of IBD patients has not been applied yet. Here, we briefly reviewed some representative studies and subsequently addressed the challenges and opportunities of AI in the medical care of IBD patients.

Based on the clinical, endoscopic, and histopathologic features, majority of CD and UC could be clearly differentiated. However, approximately 10% to 30% of patients could not be categorized into CD or UC, defined as IBD unclassified.^[4] Several biomarkers, including anti-*Saccharomyces cerevisiae* antibodies and antineutrophil cytoplasmic antibodies, T helper lymphocyte polarization in the lamina propria, and certain micro-RNAs (miRNAs) have been

indicated to have discriminating potential for the differential diagnosis of IBD, which are not yet well-adapted in clinical practice. Some studies attempted to construct a discriminating model using ML to improve the accuracy of the differential diagnosis of IBD based on the endoscopic and histopathologic findings, single nucleotide polymorphisms (SNPs), or microbiota. One retrospective study with a large sample size enrolled 20,076 CD and 15,307 UC patients from the International IBD Genetics Consortium genomic databases and constructed a combined model involving multiple SNPs using SVM, which had an area under the receiver operating characteristic curve (AUROC) of 0.864 for the differential diagnosis between CD and UC.^[5] Another study combined the endoscopic and histopathologic lesions in pediatric IBD (143 CD and 67 UC) patients to construct a model using SVM, which had an AUROC curve of 0.87.^[6] One study analyzed the gut microbiota in 20 CD and 19 UC patients and constructed a model based on the operational taxonomic units using RF, which had an AUROC of 0.72 for the differential diagnosis between CD and UC.^[7] However, all these studies were retrospective and did not involve a large sample size of IBD unclassified patients. Besides the above-mentioned studies, there were a few studies using ML to construct different models to discriminate UC or CD from the healthy controls including SNPs, miRNAs, or multi-omics with accuracy ranging from 78.9% to 92.8%.^[8,9]

Endoscopic procedures are significantly important to assess IBD. Some studies focused on the wireless capsule endoscopy in the evaluation of the gut ulcerative lesions using SVM or convolutional network among CD patients, and the accuracy of these algorithm models were all relatively high with rates ranging from 89.3% to 93.8%.^[10,11] However, these studies did not systemically evaluate the endoscopic disease status of CD patients. One study from Japan performed endoscopy to evaluate the active histopathologic lesions using SVM and obtained an algorithm with a high accuracy rate of

Access this article online

Quick Response Code:



Website:
www.cmj.org

DOI:
10.1097/CM9.0000000000000714

Correspondence to: Prof. Jia-Ming Qian, Department of Gastroenterology, Chinese Academy of Medical Sciences and Peking Union Medical College Hospital, Beijing 100730, China
E-Mail: qianjiaming1957@126.com

Copyright © 2020 The Chinese Medical Association, produced by Wolters Kluwer, Inc. under the CC-BY-NC-ND license. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Chinese Medical Journal 2020;133(7)

Received: 04-01-2020 Edited by: Qiang Shi

91.0% in 187 UC patients and 22,853 images.^[12] Another study enrolled 952 UC patients and 30,322 colonoscopic images and constructed a model with convolutional network, with high AUROCs of 0.94–0.99 between Mayo 0–1 and Mayo 2–3 lesions.^[13] As discussed above, endoscopic studies using ML always involved a large number of images and had significant potential for endoscopic disease assessment.

Clearly determining the predictors and clinical outcomes of certain medications is significantly important to accurately determine the effective medications used to treat IBD. Waljee *et al*^[14] used RF methods to conduct several studies to construct different predictive models in IBD patients. They developed an algorithm involving age and laboratory tests to evaluate the disease remission in patients taking thiopurine, which had an AUROC curve of 0.79, greater than 0.49 based on 6-thioguanine nucleotide (6-TGN) levels, indicating the superiority of ML-based algorithm and the ineffectivity of 6-TGN test in predicting disease remission. Another study collected the data of the phase 3 clinical trial of vedolizumab in CD patients through the Clinical Study Data Request website and constructed a model using the routine laboratory tests (C-reactive protein and biochemical tests) 6 weeks after vedolizumab administration, with an AUROC of 0.75, suggesting the probability of using the laboratory data 6 weeks after vedolizumab administration to predict disease remission at week 52.^[15] Through the Veterans Health Administration Electronic Database, they also constructed an RF longitudinal algorithm model using the demographics, laboratory tests, and medications of 20,368 IBD patients, with a high AUROC of 0.85 in predicting hospitalization, with this model considered beneficial in categorizing patients according to their different risks for hospitalization and in administering individualized medications in each patient.^[16]

Following the increased clinical use of immunosuppressants and biological agents, the incidence of surgery for IBD decreased. However, it is still significantly important to determine the predictors for surgery. One study enrolled 239 CD patients and constructed some algorithms using the patients' clinical manifestations, radiologic findings, and laboratory tests using several ML methods, including RF, SVM, and artificial neural networks (ANNs). The RF model had the highest accuracy of 96.26%.^[17] An Indian team used RF to construct an algorithm used to predict colectomy in severe colitis patients with an accuracy of 77%.^[18]

Big data refers to sets of data whose scale and complexity require the use of dedicated analytical and statistical approaches. Currently, the most important data sources for medical big data include administrative databases, online clinical trials registries, electronic medical records, medical images, and omics data,^[3] which had been used in the abovementioned studies. However, these datasets were heterogeneous and retrospectively collected and lacked highly randomized controlled study design, limiting their value for future clinical application. Meanwhile, a gold standard method to establish the diagnosis of IBD does not exist yet, and poor consistency in the disease assessment based on the clinical manifestations, endoscopic scores, and histopathologic scores in CD patients, and many novel

biological agents for the treatment of IBD have resulted in significant challenges in constructing high-quality datasets. Considering the increasing prevalence of IBD, developing countries have significant opportunities to construct big datasets, with the consistent desire for high-quality databases with the following characteristics: databases with high volume, high speed for data collection, and a significant number of structured or unstructured variables and databases comprising crossover and longitudinal clinical information and novel biomarkers based on multi-omics.

Never causing foreseeable or unintentional harm is the milestone of the ethical issues of AI. Nowadays, majority of the AI studies in the field of IBD only have a variable accuracy of 72% to 96%. The misdiagnosis or misclassification based on the AI system is inevitable, possibly affecting the decision-making process and the patient's beneficence, preference, and right of informed consent. Minority bias, label bias, agency bias, and informed mistrust all might contribute to the bias of model development and deployment, which will affect the fairness and accuracy of the algorithm derived from AI.^[19] Sometimes, disparity is the major confounding bias to influence the accuracy and efficacy of the model. For example, there are significant differences in the genotypic aspects between the East and West, specifically the rarity of nucleotide-binding oligomerization domain-containing protein 2 (*NOD2*) mutation in Asian CD patients.^[20] When deploying the model, these disparities should be taken into consideration.

Considering the AI model's generic characteristics and unavailability or transparency in calculating progress, it is always challenging to conduct external validation for the AI model. Moreover, it is also relatively difficult to determine why an AI model has several errors when it fails in clinical practice. Hence, the AI model should develop its explainability, which is the ability to explain what happens when the model makes a decision.^[21] However, more powerful ML models always involved more parameters, which might be not easily explainable. Hence, when making decisions using the AI assistant system in future clinical practice, the healthcare providers should be well-informed or should be knowledgeable about the system.

One of the most important impulses for the development of AI is the evident progression of computer science and bioinformatics, which is essential to adequately manage and integrate big data. Several methods have been developed with varying levels of performance. Among the abovementioned studies, several of them were conducted based on several ML methods (RF, ANN, SVM, and LR) to determine the best algorithm model.^[17,22] Statisticians should get involved in the study design as early as possible, especially in clinical studies with multi-center, large sample size, novel interventions, who will definitely give valuable suggestions for the data collection, data analysis with appropriate methods and interpretation of the results.

Following the application and progress of AI in the field of economics and some medical fields, AI would definitely contribute to the improvement of medical care of IBD patients. On the road of application of AI assistant medications in IBD, prospective randomized controlled

studies with large sample sizes and high-quality databases are always required to determine its effectiveness and safety. Meanwhile, it is essentially important to cope with the ethical concerns and facilitate collaborations among clinicians, statisticians, and bioinformaticians.

Funding

This work was supported by grants from the Fundamental Research Funds for the Central Universities (No. 3332018012) and the Funding for Elite Training in the Dongcheng District, Beijing (No. 2018-34).

Conflicts of interest

None.

References

1. Le Berre C, Sandborn WJ, Aridhi S, Devignes MD, Fournier L, Smail-Tabbone M, *et al.* Application of artificial intelligence to gastroenterology and hepatology. *Gastroenterology* 2020;158:76–94. doi: 10.1053/j.gastro.2019.08.058.
2. El Hajjar A, Rey JF, Jay N, Natoli G, Peyrin-Biroulet L. Artificial intelligence in gastrointestinal endoscopy: general overview. *Chin Med J* 2020;133:326–324. doi: 10.1097/CM9.0000000000000623.
3. Olivera P, Danese S, Jay N, Natoli G, Peyrin-Biroulet L. Big data in IBD: a look into the future. *Nat Rev Gastroenterol Hepatol* 2019;16:312–321. doi: 10.1038/s41575-019-0102-5.
4. Burakoff R. Indeterminate colitis: clinical spectrum of disease. *J Clin Gastroenterol* 2004;38:S41–43. doi: 10.1097/01.mcg.0000123991.13937.7e.
5. Wei Z, Wang W, Bradfield J, Li J, Cardinale C, Frackelton E, *et al.* Large sample size, wide variant spectrum, and advanced machine-learning technique boost risk prediction for inflammatory bowel disease. *Am J Hum Genet* 2013;92:1008–1012. doi: 10.1016/j.ajhg.2013.05.002.
6. Mossotto E, Ashton JJ, Coelho T, Beattie RM, MacArthur BD, Ennis S. Classification of paediatric inflammatory bowel disease using machine learning. *Sci Rep* 2017;7:2427. doi: 10.1038/s41598-017-02606-2.
7. Forbes JD, Chen CY, Knox NC, Marrie RA, El-Gabalawy H, de Kievit T, *et al.* A comparative study of the gut microbiota in immune-mediated inflammatory diseases—does a common dysbiosis exist? *Microbiome* 2018;6:221. doi: 10.1186/s40168-018-0603-4.
8. Douglas GM, Hansen R, Jones CMA, Dunn KA, Comeau AM, Bielawski JP, *et al.* Multi-omics differentially classify disease state and treatment outcome in pediatric Crohn's disease. *Microbiome* 2018;6:13. doi: 10.1186/s40168-018-0398-3.
9. Duttagupta R, DiRienzo S, Jiang R, Bowers J, Gollub J, Kao J, *et al.* Genome-wide maps of circulating miRNA biomarkers for ulcerative colitis. *PLoS One* 2012;7:e31241. doi: 10.1371/journal.pone.0031241.
10. Charisis VS, Hadjileontiadis LJ. Potential of hybrid adaptive filtering in inflammatory lesion detection from capsule endoscopy images. *World J Gastroenterol* 2016;22:8641–8657. doi: 10.3748/wjg.v22.i39.8641.
11. 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;89:357–363. doi: 10.1016/j.gie.2018.10.027.
12. Maeda Y, Kudo SE, Mori Y, Misawa M, Ogata N, Sasanuma S, *et al.* Fully automated diagnostic system with artificial intelligence using endocytoscopy to identify the presence of histologic inflammation associated with ulcerative colitis (with video). *Gastrointest Endosc* 2019;89:408–415. doi: 10.1016/j.gie.2018.09.024.
13. Ozawa T, Ishihara S, Fujishiro M, Saito H, Kumagai Y, Shichijo S, *et al.* Novel computer-assisted diagnosis system for endoscopic disease activity in patients with ulcerative colitis. *Gastrointest Endosc* 2019;89:416–421. doi: 10.1016/j.gie.2018.10.020.
14. Waljee AK, Sauder K, Patel A, Segar S, Liu B, Zhang Y, *et al.* Machine learning algorithms for objective remission and clinical outcomes with thiopurines. *J Crohns Colitis* 2017;11:801–810. doi: 10.1093/ecco-jcc/jjx014.
15. Waljee AK, Liu B, Sauder K, Zhu J, Govani SM, Stidham RW, *et al.* Predicting corticosteroid-free biologic remission with vedolizumab in Crohn's disease. *Inflamm Bowel Dis* 2018;24:1185–1192. doi: 10.1093/ibd/izy031.
16. Waljee AK, Lipson R, Wiitala WL, Zhang Y, Liu B, Zhu J, *et al.* Predicting hospitalization and outpatient corticosteroid use in inflammatory bowel disease patients using machine learning. *Inflamm Bowel Dis* 2017;24:45–53. doi: 10.1093/ibd/izx007.
17. Dong Y, Xu L, Fan Y, Xiang P, Gao X, Chen Y, *et al.* A novel surgical predictive model for Chinese Crohn's disease patients. *Medicine* 2019;98:e17510. doi: 10.1097/MD.00000000000017510.
18. Jain S, Kedia S, Sethi T, Bopanna S, Yadav DP, Goyal S, *et al.* Predictors of long-term outcomes in patients with acute severe colitis: a northern Indian cohort study. *J Gastroenterol Hepatol* 2018;33:615–622. doi: 10.1111/jgh.13921.
19. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. *Ann Intern Med* 2018;169:866–872. doi: 10.7326/M18-1990.
20. Guo C, Wu K. Risk genes of inflammatory bowel disease in Asia: what are the most important pathways affected? *Dig Dis* 2016;34:5–11. doi: 10.1159/000442917.
21. Geis JR, Brady AP, Wu CC, Spencer J, Ranschaert E, Jaremko JL, *et al.* Ethics of artificial intelligence in radiology: summary of the joint European and North American multisociety statement. *Radiology* 2019;293:436–440. doi: 10.1148/radiol.2019191586.
22. Isakov O, Dotan I, Ben-Shachar S. Machine learning-based gene prioritization identifies novel candidate risk genes for inflammatory bowel disease. *Inflamm Bowel Dis* 2017;23:1516–1523. doi: 10.1097/MIB.0000000000001222.

How to cite this article: Li J, Qian JM. Artificial intelligence in inflammatory bowel disease: current status and opportunities. *Chin Med J* 2020;133:757–759. doi: 10.1097/CM9.0000000000000714