

World Journal of *Gastroenterology*

World J Gastroenterol 2021 October 14; 27(38): 6348-6514



EDITORIAL

- 6348 Biomarkers for gastrointestinal adverse events related to thiopurine therapy
Zudeh G, Franca R, Stocco G, Decorti G

REVIEW

- 6357 Fully covered metal biliary stents: A review of the literature
Lam R, Muniraj T
- 6374 Intraoperative use of indocyanine green fluorescence imaging in rectal cancer surgery: The state of the art
Peltrini R, Podda M, Castiglioni S, Di Nuzzo MM, D'Ambra M, Lionetti R, Sodo M, Luglio G, Mucilli F, Di Saverio S, Bracale U, Corcione F

MINIREVIEWS

- 6387 Transcription factors specificity protein and nuclear receptor 4A1 in pancreatic cancer
Safe S, Shrestha R, Mohankumar K, Howard M, Hedrick E, Abdelrahim M
- 6399 Artificial intelligence for the early detection of colorectal cancer: A comprehensive review of its advantages and misconceptions
Viscaino M, Torres Bustos J, Muñoz P, Auat Cheein C, Cheein FA
- 6415 Faecal immunochemical test outside colorectal cancer screening?
Pin-Vieito N, Puga M, Fernández-de-Castro D, Cubiella J

ORIGINAL ARTICLE

Basic Study

- 6430 Fecal metabolomic profiles: A comparative study of patients with colorectal cancer vs adenomatous polyps
Nannini G, Meoni G, Tenori L, Ringressi MN, Taddei A, Niccolai E, Baldi S, Russo E, Luchinat C, Amedei A

Retrospective Cohort Study

- 6442 High total Joule heat increases the risk of post-endoscopic submucosal dissection electrocoagulation syndrome after colorectal endoscopic submucosal dissection
Ochi M, Kawagoe R, Kamoshida T, Hamano Y, Ohkawara H, Ohkawara A, Kakinoki N, Yamaguchi Y, Hirai S, Yanaka A, Tsuchiya K

Retrospective Study

- 6453 Effects of acute kidney injury on acute pancreatitis patients' survival rate in intensive care unit: A retrospective study
Shi N, Sun GD, Ji YY, Wang Y, Zhu YC, Xie WQ, Li NN, Han QY, Qi ZD, Huang R, Li M, Yang ZY, Zheng JB, Zhang X, Dai QQ, Hou GY, Liu YS, Wang HL, Gao Y

- 6465** Magnetic resonance imaging-radiomics evaluation of response to chemotherapy for synchronous liver metastasis of colorectal cancer

Ma YQ, Wen Y, Liang H, Zhong JG, Pang PP

Observational Study

- 6476** Deep learning *vs* conventional learning algorithms for clinical prediction in Crohn's disease: A proof-of-concept study

Con D, van Langenberg DR, Vasudevan A

- 6489** Serum soluble suppression of tumorigenicity 2 as a novel inflammatory marker predicts the severity of acute pancreatitis

Zhang Y, Cheng B, Wu ZW, Cui ZC, Song YD, Chen SY, Liu YN, Zhu CJ

CASE REPORT

- 6501** Monomorphic epitheliotropic intestinal T-cell lymphoma presenting as melena with long-term survival: A case report and review of literature

Ozaka S, Inoue K, Okajima T, Tasaki T, Arika S, Ono H, Ando T, Daa T, Murakami K

CORRECTION

- 6511** Correction to "Effect of probiotic *Lactobacillus plantarum* Dad-13 powder consumption on the gut microbiota and intestinal health of overweight adults". *World J Gastroenterol* 2021; 27(1): 107-128 [PMID: 33505154 DOI: 10.3748/wjg.v27.i1.107]

Rahayu ES

LETTER TO THE EDITOR

- 6513** Preservation of the superior rectal artery in laparoscopic colectomy for slow transit constipation: Is it really associated with better outcomes?

Parra RS, Feres O, Rocha JJR

ABOUT COVER

Editorial Board Member of *World Journal of Gastroenterology*, Veerapol Kukongviriyapan, PhD, Professor, Department of Pharmacology, Faculty of Medicine, Khon Kaen University, 123 Moo 16, Mittraphap Road, Muang District, Khon Kaen 40002, Thailand. veerapol@kku.ac.th

AIMS AND SCOPE

The primary aim of *World Journal of Gastroenterology* (WJG, *World J Gastroenterol*) is to provide scholars and readers from various fields of gastroenterology and hepatology with a platform to publish high-quality basic and clinical research articles and communicate their research findings online. WJG mainly publishes articles reporting research results and findings obtained in the field of gastroenterology and hepatology and covering a wide range of topics including gastroenterology, hepatology, gastrointestinal endoscopy, gastrointestinal surgery, gastrointestinal oncology, and pediatric gastroenterology.

INDEXING/ABSTRACTING

The WJG is now indexed in Current Contents®/Clinical Medicine, Science Citation Index Expanded (also known as SciSearch®), Journal Citation Reports®, Index Medicus, MEDLINE, PubMed, PubMed Central, and Scopus. The 2021 edition of Journal Citation Report® cites the 2020 impact factor (IF) for WJG as 5.742; Journal Citation Indicator: 0.79; IF without journal self cites: 5.590; 5-year IF: 5.044; Ranking: 28 among 92 journals in gastroenterology and hepatology; and Quartile category: Q2. The WJG's CiteScore for 2020 is 6.9 and Scopus CiteScore rank 2020: Gastroenterology is 19/136.

RESPONSIBLE EDITORS FOR THIS ISSUE

Production Editor: Ji-Hong Lin; Production Department Director: Yun-Jie Ma; Editorial Office Director: Ze-Mao Gong.

NAME OF JOURNAL

World Journal of Gastroenterology

ISSN

ISSN 1007-9327 (print) ISSN 2219-2840 (online)

LAUNCH DATE

October 1, 1995

FREQUENCY

Weekly

EDITORS-IN-CHIEF

Andrzej S Tarnawski, Subrata Ghosh

EDITORIAL BOARD MEMBERS

<http://www.wjgnet.com/1007-9327/editorialboard.htm>

PUBLICATION DATE

October 14, 2021

COPYRIGHT

© 2021 Baishideng Publishing Group Inc

INSTRUCTIONS TO AUTHORS

<https://www.wjgnet.com/bpg/gerinfo/204>

GUIDELINES FOR ETHICS DOCUMENTS

<https://www.wjgnet.com/bpg/GerInfo/287>

GUIDELINES FOR NON-NATIVE SPEAKERS OF ENGLISH

<https://www.wjgnet.com/bpg/gerinfo/240>

PUBLICATION ETHICS

<https://www.wjgnet.com/bpg/GerInfo/288>

PUBLICATION MISCONDUCT

<https://www.wjgnet.com/bpg/gerinfo/208>

ARTICLE PROCESSING CHARGE

<https://www.wjgnet.com/bpg/gerinfo/242>

STEPS FOR SUBMITTING MANUSCRIPTS

<https://www.wjgnet.com/bpg/GerInfo/239>

ONLINE SUBMISSION

<https://www.f6publishing.com>



Observational Study

Deep learning vs conventional learning algorithms for clinical prediction in Crohn's disease: A proof-of-concept study

Danny Con, Daniel R van Langenberg, Abhinav Vasudevan

ORCID number: Danny Con 0000-0002-4983-6103; Daniel R van Langenberg 0000-0003-3662-6307; Abhinav Vasudevan 0000-0001-5026-9014.

Author contributions: Con D contributed conceptualization, data collection, statistical analysis, data interpretation, manuscript drafting; van Langenberg DR contributed conceptualization, data interpretation, reviewing of manuscript critically for important intellectual content; Vasudevan A contributed conceptualization, data collection, data interpretation, reviewing of manuscript critically for important intellectual content; all authors approved the final version of the manuscript.

Institutional review board

statement: This study was reviewed and approved by the Eastern Health Office of Research & Ethics (approval number: LR 61/2015).

Informed consent statement:

Patients were not required to give informed consent to the study because the analysis used anonymous clinical data that were obtained retrospectively.

Conflict-of-interest statement: Con D has no relevant conflicts of interest to declare. AV has received

Danny Con, Daniel R van Langenberg, Abhinav Vasudevan, Department of Gastroenterology and Hepatology, Eastern Health, Box Hill 3128, Victoria, Australia

Daniel R van Langenberg, Abhinav Vasudevan, Faculty of Medicine, Nursing and Health Sciences, Monash University, Box Hill 3128, Victoria, Australia

Corresponding author: Danny Con, MD, Doctor, Statistician, Department of Gastroenterology and Hepatology, Eastern Health, 8 Arnold Street, Box Hill 3128, Victoria, Australia.
dannycon302@gmail.com

Abstract

BACKGROUND

Traditional methods of developing predictive models in inflammatory bowel diseases (IBD) rely on using statistical regression approaches to deriving clinical scores such as the Crohn's disease (CD) activity index. However, traditional approaches are unable to take advantage of more complex data structures such as repeated measurements. Deep learning methods have the potential ability to automatically find and learn complex, hidden relationships between predictive markers and outcomes, but their application to clinical prediction in CD and IBD has not been explored previously.

AIM

To determine and compare the utility of deep learning with conventional algorithms in predicting response to anti-tumor necrosis factor (anti-TNF) therapy in CD.

METHODS

This was a retrospective single-center cohort study of all CD patients who commenced anti-TNF therapy (either adalimumab or infliximab) from January 1, 2010 to December 31, 2015. Remission was defined as a C-reactive protein (CRP) < 5 mg/L at 12 mo after anti-TNF commencement. Three supervised learning algorithms were compared: (1) A conventional statistical learning algorithm using multivariable logistic regression on baseline data only; (2) A deep learning algorithm using a feed-forward artificial neural network on baseline data only; and (3) A deep learning algorithm using a recurrent neural network on repeated data. Predictive performance was assessed using area under the receiver operator characteristic curve (AUC) after 10× repeated 5-fold cross-validation.

financial support to attend educational meetings from Ferring. van Langenberg DR has served as a speaker and/or received travel support from Takeda, Ferring and Shire. He has consultancy agreements with Abbvie, Janssen and Pfizer. He received research funding grants for investigator-driven studies from Ferring, Shire and AbbVie.

Data sharing statement: No additional data are available.

STROBE statement: The authors have read the STROBE Statement-checklist of items, and the manuscript was prepared and revised according to the STROBE Statement-checklist of items.

Open-Access: This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

Manuscript source: Invited manuscript

Specialty type: Gastroenterology and hepatology

Country/Territory of origin: Australia

Peer-review report's scientific quality classification

Grade A (Excellent): 0
Grade B (Very good): B
Grade C (Good): C
Grade D (Fair): 0
Grade E (Poor): 0

Received: March 5, 2021

Peer-review started: March 5, 2021

First decision: April 17, 2021

Revised: April 26, 2021

Accepted: September 6, 2021

RESULTS

A total of 146 patients were included (median age 36 years, 48% male). Concomitant therapy at anti-TNF commencement included thiopurines (68%), methotrexate (18%), corticosteroids (44%) and aminosalicylates (33%). After 12 mo, 64% had CRP < 5 mg/L. The conventional learning algorithm selected the following baseline variables for the predictive model: Complex disease behavior, albumin, monocytes, lymphocytes, mean corpuscular hemoglobin concentration and gamma-glutamyl transferase, and had a cross-validated AUC of 0.659, 95% confidence interval (CI): 0.562-0.756. A feed-forward artificial neural network using only baseline data demonstrated an AUC of 0.710 (95%CI: 0.622-0.799; $P = 0.25$ vs conventional). A recurrent neural network using repeated biomarker measurements demonstrated significantly higher AUC compared to the conventional algorithm (0.754, 95%CI: 0.674-0.834; $P = 0.036$).

CONCLUSION

Deep learning methods are feasible and have the potential for stronger predictive performance compared to conventional model building methods when applied to predicting remission after anti-TNF therapy in CD.

Key Words: Machine learning; Artificial intelligence; Precision medicine; Personalized medicine; Deep learning

©The Author(s) 2021. Published by Baishideng Publishing Group Inc. All rights reserved.

Core Tip: Deep learning has vast potential, but its clinical utility in predicting outcomes in Crohn's disease (CD) has not been explored. This study showed that deep learning algorithms (a recurrent neural network) using a more complex information structure including repeated biomarker measurements had a better predictive performance compared to a conventional statistical algorithm using only baseline data. This proof-of-concept study therefore paves the way for further research in the use of deep learning methods in clinical prediction in CD.

Citation: Con D, van Langenberg DR, Vasudevan A. Deep learning vs conventional learning algorithms for clinical prediction in Crohn's disease: A proof-of-concept study. *World J Gastroenterol* 2021; 27(38): 6476-6488

URL: <https://www.wjgnet.com/1007-9327/full/v27/i38/6476.htm>

DOI: <https://dx.doi.org/10.3748/wjg.v27.i38.6476>

INTRODUCTION

Crohn's disease (CD) is a heterogeneous chronic inflammatory bowel disease (IBD) that is characterized by intermittent flares, medication changes, the potential need for surgery and substantial psychological morbidity[1,2]. As with many chronic conditions, predicting disease trajectory, outcomes and response to therapies in CD are key components of clinical practice where management is tailored to the individual [3]. Precision medicine has been in part driven by the vast expansion of available electronic health data, genomic data and novel disease biomarkers[3]. However, deciphering the complex relationships between large amounts of information and multiple data types presents new analytical challenges.

Traditional approaches to constructing prediction models rely on multivariable regression approaches, typically logistic regression for classification or proportional hazards regression for longitudinal prediction[4]. The resulting predictive models are thus typically only linear combinations of the included predictors and may have limited ability to learn more complex relationships within the data. The advantage of machine learning and artificial intelligence over traditional predictive tools is the potential ability for computational algorithms to automatically find and learn complex, hidden relationships between predictive markers and outcomes[5,6]. This is especially true for deep learning or artificial neural network (ANN) methods, although their 'black box' approach has been criticized for an inability to produce a causal

Article in press: September 6, 2021

Published online: October 14, 2021

P-Reviewer: Jin B, Yu C

S-Editor: Gao CC

L-Editor: A

P-Editor: Liu JH



explanation between predictors and outcomes[6].

Despite some limitations, there is much interest in developing and testing machine learning and deep learning tools to aid decision making[5,7]. In luminal gastroenterology, machine learning is gaining traction but its use has been relatively limited to automatic image recognition in endoscopy[8-11] as well as feature selection in genomic and microbiomics data[12,13]. Although there has been great interest in predicting clinical outcomes in CD such as response to therapeutics including biologics[14-18] and immunomodulators[19,20], studies investigating the utility of machine learning models for such predictive tasks have been more limited[21-23]. In particular, the utility of deep learning or ANNs specifically in clinical prediction of CD remains unknown[7].

We aimed to evaluate the utility of deep learning algorithms compared with conventional statistical learning algorithms for clinical prediction in this proof-of-concept study. In particular, we aimed to compare these algorithms as methods of learning and prediction in a general sense, rather than to develop any specific predictive model or score.

MATERIALS AND METHODS

Study design

This proof-of-concept study utilized a retrospective longitudinal cohort at a tertiary health network comprising three acute hospitals in Melbourne, Australia. The focus of the study was to compare the ability of two supervised learning algorithms (conventional statistical learning *vs* deep learning) to predict remission after 12 mo of treatment using clinical variables and biomarkers available at baseline. The performance of each algorithm was evaluated using cross-validation. The emphasis of the study was to compare the predictive performance of the two methods of learning rather than any specific model itself. This study was approved by the Eastern Health Office of Research & Ethics (approval number: LR 61/2015).

Study cohort

All adult patients > 18 years with confirmed CD according to standard criteria[24] were included if they were commenced on treatment with an anti-tumor necrosis factor (anti-TNF) agent (adalimumab or infliximab) for luminal CD and received at least one dose of the drug between January 2010 and December 2015. Patients receiving anti-TNF for perianal disease without luminal disease were excluded. Patients were followed up for 12 mo to determine rates of biochemical remission.

Outcomes

Response to anti-TNF was defined as having achieved biochemical remission as per serum C-reactive protein (CRP) < 5 mg/L at 12 mo. This endpoint was chosen because CRP is an accepted biomarker to reflect disease activity and predict outcomes in CD [25,26]. Additionally, normalization of CRP predicts better outcomes in CD patients in remission[27,28]. The first CRP measurement after 12 mo and before 18 mo was used. Patients who did not have a CRP measurement in this time period were excluded.

Data collection and pre-processing

Baseline characteristics were collected *via* hospital and clinic records, including Montreal classification, concomitant baseline therapies, prior anti-TNF exposure and prior surgeries. Biomarker data were collected at two time points: (1) A baseline measurement defined as the most proximate measurement prior to commencing anti-TNF, up to 3 mo before commencement; and (2) A prior measurement defined as the second most proximate measurement, up to 12 mo before commencement. Only patients with complete baseline data were included, while missing prior values were imputed with the respective baseline value. The following variables were log-transformed to correct skewness: serum bilirubin, alanine aminotransferase, alkaline phosphatase and gamma-glutamyl transferase (GGT). The data underlying this article cannot be shared publicly due to privacy and ethical concerns. The data will be shared upon reasonable request to the corresponding author.

Statistical learning algorithm (conventional approach)

The conventional approach to developing a predictive clinical model is to run univariable and multivariable regression analysis to find useful and preferably

independent predictors of the outcome of interest (see [Figure 1](#)). Criteria for variable selection usually involves significance testing (P values) or likelihood-based information criterion (such as the Akaike information criterion). In this study, logistic regression was used given the dichotomous nature of the outcome (CRP < 5 mg/L vs CRP \geq 5 mg/L). The conventional approach typically only uses data from a single time-point, therefore we used baseline data only (the most proximate measurement for all biomarkers). For this conventional approach, we employed the following modelling algorithm: (1) Perform univariable logistic regression on each variable and retain all variables with $P < 0.5$; (2) Run backwards stepwise selection on all retained variables with removal criterion $P > 0.2$; and (3) Use the regression coefficients in the remaining multivariable model to derive the predictive score.

Deep learning algorithms (experimental approach)

A basic deep learning algorithm is a feed-forward ANN[6]. An ANN is composed of layers: an input layer (consisting of all the input predictor variables), an output layer (the prediction), and a number of 'hidden' layers (see [Figure 1](#)). Nodes within a hidden layer are called 'neurons'. The hidden layers allow an ANN to learn complex, non-linear relationships between input variables and the outcome of interest. The influence of nodes in a layer on other nodes in subsequent layers is 'trained' or fitted using a mathematical function and ultimately determines how information is propagated through the ANN – this is analogous to fitting a regression line on data in conventional statistics. An ANN with only an input and output layer, without hidden layers, can be analogous to simple logistic regression, although they are not equivalent.

However, like the conventional statistical algorithm, a basic feed-forward ANN is still only able to model relationships between predictors at a single time-point. A recurrent neural network (RNN) is a more advanced deep learning algorithm that is able to model repeated measurements over time. Like a feed-forward ANN, information is propagated from the input layer to the output layer. However, instead of only allowing the information to pass through once, information is fed to the RNN sequentially, or 'recurrently' – that is, each set of repeated measurements is inputted once at a time allowing the RNN to update its knowledge of the relationship between the predictors and the outcome. Therefore, the algorithm is additionally able to learn and utilize the dynamics of biomarkers over time, in a way that cannot be achieved by conventional statistical learning methods.

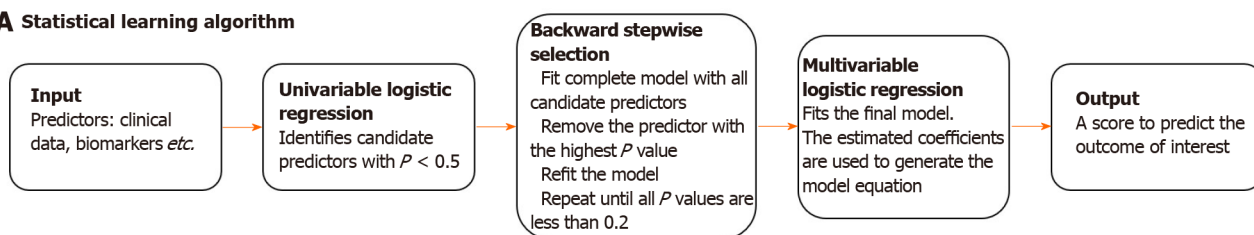
We tested the feed-forward ANN and the RNN in three separate experiments: (1) Using all baseline clinical data in a feed-forward ANN; (2) Using only baseline biomarker data in a feed-forward ANN; and (3) Using repeated biomarker data in an RNN. In this study after hyper-parameter tuning, we used a feed-forward ANN architecture of 3 hidden layers, each with 64 neurons, and an RNN architecture of 1 hidden layer with 64 neurons.

Comparison of algorithms

The predictive performances of the conventional statistical algorithm and the experimental deep learning algorithm (ANN) was defined as their ability to correctly classify 12-mo CRP < 5 mg/L measured using the area under the receiver operator characteristic curve (AUC). Because the learning ability of an ANN can be arbitrarily increased, an overly powerful ANN that is trained such that it has near-perfect prediction on the original training cohort, would suffer from poor predictive ability in an external cohort (this is called 'over-fitting', a well-known phenomenon). Similarly, the same conventional statistical learning algorithm might result in models with different variables when applied to different cohorts. Therefore, it is important to evaluate the ability of a learning algorithm to predict outcomes in patients that are not included in the original training cohort (external validity).

In the absence of external testing cohorts to assess external validity, cross-validation is an internal validation procedure that is suited to this purpose[4]. During cross-validation, the cohort is randomly divided into k equally sized sub-cohorts, known as 'folds' (where k is often 5 or 10 by convention). Then, one fold is set aside to be used to test the algorithm, after the algorithm is first trained on the remaining $k-1$ folds (see [Figure 2](#)). This allows the algorithms to be tested on patients that were not used during training. The process is then repeated for each fold (where each fold takes turns in being the test fold). The average AUC after repeating k times gives the cross-validated AUC. However, this procedure is not free from error, because the partitioning process may have randomly resulted in a better (or worse) than usual performance. Thus it is important to repeat the whole process a number of times, to reduce this error[29].

A Statistical learning algorithm



B Deep learning algorithm

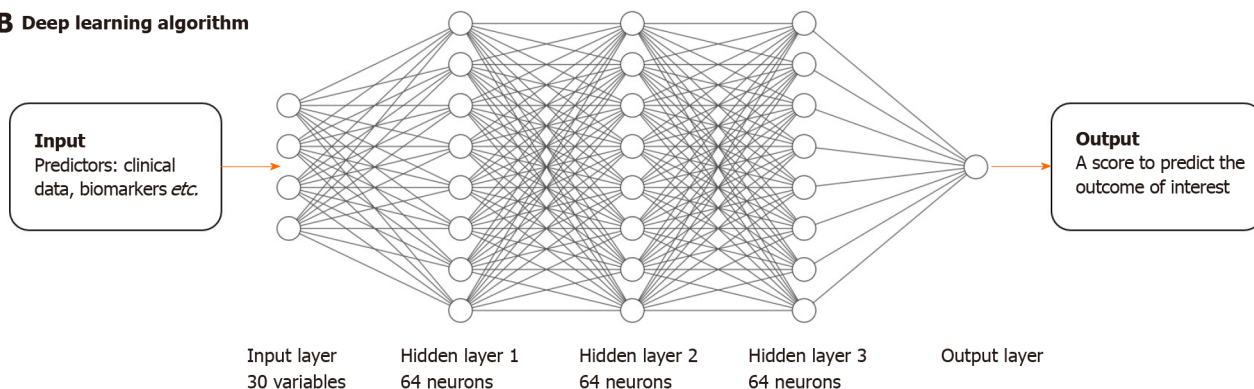


Figure 1 Comparison of the predictive modelling process using two supervised learning algorithms. A: Conventional statistical learning; B: Deep learning.

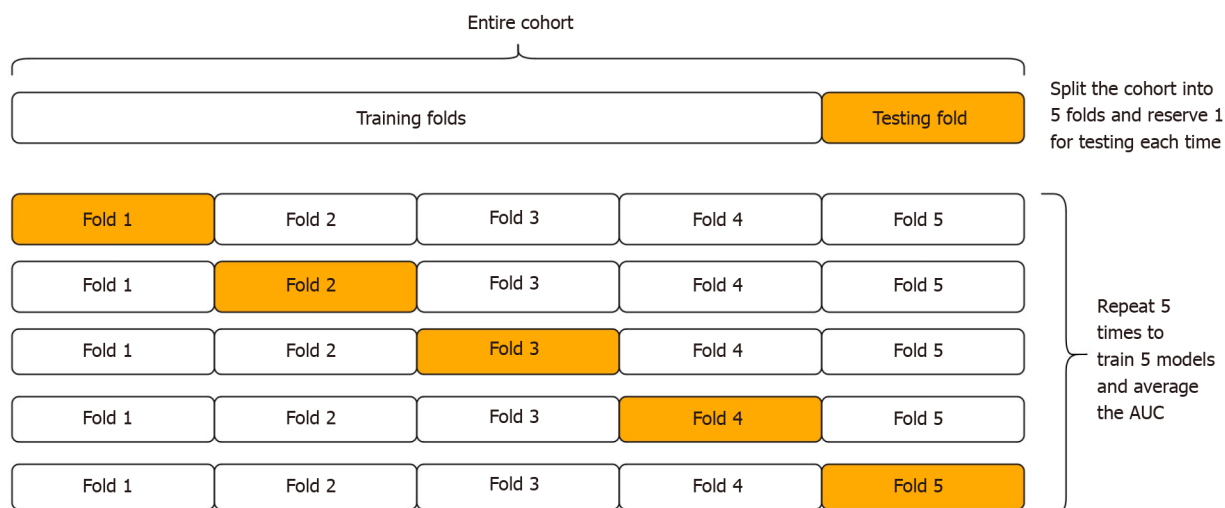


Figure 2 Schematic diagram of *k*-fold cross validation procedure for *k* = 5. This method is considered more reliable than a random train-test split, which would be analogous to training only one model, instead of the average of *k* models. AUC: Area under the receiver operator characteristic curve.

For this study, we used 5-fold cross-validation repeated 10 times to estimate the generalizability of each algorithm on unseen data. Statistical comparison of the cross-validated AUCs of each learning algorithm was made using the variance-corrected repeated *k*-fold *t* test instead of a conventional paired *t* test because of the independency violation from repeated partitioning of the same dataset[29]. For comparison, the naïve or apparent AUC of each model after training and testing on the same entire cohort was given, however this is non-informative. Sample size calculations were conducted only as a guide given the exploratory nature of the study and without prior similar studies on which to base AUC assumptions. The target sample size to detect a 10% difference in AUC with 80% power and 95% significance assuming an AUC variance of 10% was $n = 157$ [30]. To instead detect a 15% difference in AUC under the same conditions, a sample size of $n = 70$ was required. The Python 3.8.4 programming language with the open-source module PyTorch was used to create the deep learning algorithm. Stata/IC 16 (Texas, United States, 2020) was used to create the statistical learning algorithm.

RESULTS

Baseline characteristics

A total of 146 CD patients were included (see [Table 1](#)). Their median age was 36 years [inter-quartile range (IQR) 25-50], 48% were male and median disease duration since diagnosis was 5 years (IQR 1-12). The anti-TNF commenced was infliximab in 58% and adalimumab in 42%. Concomitant therapy at anti-TNF commencement included thiopurines (68%), methotrexate (18%), corticosteroids (44%) and aminosalicylates (33%). Over a quarter of patients (28%) had prior intestinal surgery, while 15% had prior exposure to anti-TNF. After 12 mo, 94 (64%) patients were in biochemical remission (CRP < 5 mg/L).

Statistical learning algorithm

Univariable analysis: Baseline factors associated with biochemical remission at 12 mo on univariable testing included non-complex disease behavior (B1), higher albumin and mean corpuscular hemoglobin concentration (MCHC), and lower platelets, lymphocytes and monocytes (each $P < 0.05$; see [Table 2](#)), while lower neutrophil count was nearly significant ($P = 0.06$). There was no significant association with age, sex, disease location or baseline medical therapies (see [Table 2](#)).

Multivariable analysis: After backward stepwise selection, the following variables remained in the final multivariable model: Complex disease, baseline albumin, monocytes, lymphocytes, MCHC and GGT (see [Table 2](#)). The resulting prediction model was given by the following equation (coefficients correct to two significant figures): $\text{Score} = 0.079 \times (\text{albumin, g/L}) + 0.050 \times (\text{MCHC, mg/L}) - 1.1 \times (\text{monocytes, } 10^9/\text{L}) - 0.43 \times (\text{lymphocytes, } 10^9/\text{L}) - 1.0 \times (\text{complex disease, } y=1 | n=0) - 0.69 \times \log_e (\text{GGT, IU/L})$.

Outcome prediction: After 10×5 -fold cross validation, the average AUC of the statistical learning algorithm was 0.659 [95% confidence interval (CI): 0.562-0.756]. This suggests the statistical learning algorithm is expected to accurately classify 65.9% of patients in external cohorts who have similar characteristics to the study cohort (see [Table 3](#)). The algorithm performed better than chance (AUC > 0.5) 94% of the time and had an AUC > 0.7 in 38% of occasions (see [Figure 3](#)). The apparent naïve AUC (when trained and tested on the same data) of the model was 0.771.

Deep learning algorithms

Feed-forward ANN with complete baseline data: The feed-forward ANN with complete baseline data had a cross-validated AUC of 0.710 (95%CI: 0.622-0.799) (see [Figure 3](#) and [Table 3](#)). This difference was not statistically significant using the variance corrected t test ($P = 0.25$). The algorithm performed better than chance 100% of the time and had good performance (AUC > 0.7) 54% of the time (see [Figure 3](#)). For comparison, the naïve AUC of the model was 0.857.

Feed-forward ANN with baseline biomarker data only: The same feed-forward ANN using only baseline biomarker data had a similar cross-validated AUC of 0.706 (95%CI: 0.621-0.791), which was again not significantly different compared to the conventional algorithm ($P = 0.33$) (see [Table 3](#)). The algorithm performed better than chance 100% of the time and had good performance (AUC > 0.7) 58% of the time (see [Figure 3](#)). The naïve AUC of the model was 0.776.

RNN with repeated biomarker data: The same feed-forward ANN using only baseline biomarker data had a similar cross-validated AUC of 0.754 (95%CI: 0.674-0.834), which was significantly higher than the AUC of the conventional algorithm ($P = 0.036$) (see [Table 3](#)). This suggests the RNN is expected to accurately classify 75.4% of patients in external cohorts who have similar characteristics to the study cohort. The RNN algorithm performed better than chance 100% of the time and had good performance (AUC > 0.7) 72% of the time (see [Figure 3](#)). For comparison, the naïve AUC of the model was 0.892.

DISCUSSION

The rapid expansion of available health data has motivated the development of machine learning and deep learning tools to predict useful outcomes in clinical medicine[5,6]. The advent of machine learning and data science techniques is

Table 1 Baseline characteristics of study cohort (*n* = 146)

Characteristic	<i>n</i> (%)
Age, years, median (IQR)	36 (25-50)
Sex	
Female	76 (52)
Male	70 (48)
Smoker (active)	33 (23)
CD behavior	
B1: Non-stricturing, non-penetrating	75 (51)
B2: Stricturing	56 (38)
B3: Penetrating/fistulizing	15 (10)
CD location	
L1: Ileal	41 (28)
L2: Colonic	43 (29)
L3: Ileocolonic	62 (42)
L4: Isolated UGI	0 (0)
Perianal involvement	20 (21)
Initial anti-TNF commenced	
Infliximab	84 (58)
Adalimumab	62 (42)
Baseline thiopurine	99 (68)
Baseline methotrexate	27 (18)
Baseline corticosteroids	64 (44)
Baseline aminosaliclates	48 (33)
Prior anti-TNF	22 (15)
Prior intestinal surgery	41 (28)
Disease duration, yr, median (IQR)	5 (1-12)
Baseline investigations	
CRP, mg/L, median (IQR)	3 (2-8)
Albumin, g/L, median (IQR)	37 (36-41)

IQR: Inter-quartile range; CD: Crohn's disease; CRP: C-reactive protein; TNF: Tumor necrosis factor; UGI: Upper gastrointestinal.

especially applicable to IBD due to the heterogeneity and chronic nature of such conditions and the repeated measures of disease activity over time which provides data that may be more suitable for complex modelling techniques. For instance, those with CD typically present with a wide array of disparate disease phenotypes and underlying pathogeneses, and their response to treatment and the trajectory of their disease course varies substantially and changes based on their response[31]. This study has exhibited the potential of deep learning algorithms in predicting response to anti-TNF therapy in patients with CD. The ability to predict the likelihood of response to a given treatment is crucial for risk-benefit assessment, which in turn is crucial to facilitate shared decision making between clinicians and patients[32]. Further, although biologic therapies have revolutionized management in IBD[31], medical therapy is now the principal driver of healthcare costs[33,34] and health economic considerations will inevitably affect treatment choice. Ideally, patients should receive therapies that are both likely to work and cost-effective. Therefore, there can be no 'one-size-fits-all' strategy to management, and precision and personalized medicine are key objectives.

Table 2 Estimated odds ratios with 95% confidence intervals on univariable and multivariable logistic regression analysis

Predictor	Univariable		Multivariable	
	OR (95%CI)	P value	Adj. OR (95%CI)	P value
Age, per year	0.98 (0.96-1.00)	0.10	-	-
Male (<i>vs</i> female)	1.42 (0.72-2.82)	0.31	-	-
CD behavior				
B1	1.0		Not included	
B2	0.45 (0.22-0.94)	0.034	Not included	
B3	0.42 (0.13-1.29)	0.13	Not included	
CD location				
L1: ileal	1.0		Not included	
L2: colonic	1.33 (0.54-3.31)	0.54	Not included	
L3: ileocolonic	0.91 (0.40-2.06)	0.83	Not included	
Ileal location (L1)	0.94 (0.45-2.00)	0.88	Not included	
Complex disease (B2/B3)	0.44 (0.22-0.89)	0.021	0.36 (0.16-0.80)	0.012
Active smoker	0.76 (0.40-1.47)	0.42	-	-
Perianal involvement	1.14 (0.49-2.65)	0.77	Not included	
Anti-TNF type: infliximab (<i>vs</i> adalimumab)	1.12 (0.56-2.22)	0.75	Not included	
Baseline immunomodulator	1.24 (0.47-3.27)	0.66	Not included	
Baseline corticosteroids	1.10 (0.56-2.18)	0.78	Not included	
Baseline aminosalicylates	1.16 (0.56-2.40)	0.69	Not included	
Prior anti-TNF	0.96 (0.37-2.47)	0.94	Not included	
Prior intestinal surgery	0.71 (0.34-1.48)	0.36	-	-
Disease duration, per log _e year	0.83 (0.65-1.06)	0.14	-	-
Albumin, per g/L	1.12 (1.03-1.22)	0.006	1.08 (0.98-1.20)	0.12
Hemoglobin, per g/L	1.01 (0.99-1.04)	0.32	-	-
HCT, per %	0.91 (0.71-1.16)	0.44	-	-
RCC, per 10 ⁹ /L	1.07 (0.84-1.36)	0.60	Not included	
MCV, per fL	1.01 (0.96-1.07)	0.64	Not included	
MCH, per pg/cell	1.15 (0.99-1.32)	0.06	-	-
MCHC, per mg/L	1.05 (1.02-1.08)	0.002	1.05 (1.02-1.09)	0.004
Platelets, per 100 × 10 ⁹ /L	0.63 (0.43-0.93)	0.020	-	-
Neutrophils, per 10 ⁹ /L	0.91 (0.82-1.00)	0.06	-	-
Lymphocytes, per 10 ⁹ /L	0.66 (0.46-0.93)	0.019	0.65 (0.41-1.02)	0.06
Monocytes, per 10 ⁹ /L	0.23 (0.08-0.63)	0.004	0.34 (0.10-1.16)	0.09
Eosinophils, per 10 ⁹ /L	0.61 (0.08-4.77)	0.64	Not included	
Basophils, per 0.01 × 10 ⁹ /L	0.92 (0.80-1.06)	0.24	-	-
Bilirubin, per log _e μmol/L	1.38 (0.70-2.72)	0.36	-	-
ALT, per log _e IU/L	1.04 (0.60-1.80)	0.90	Not included	
ALP, per log _e IU/L	0.55 (0.18-1.64)	0.28	-	-
GGT, per log _e IU/L	0.71 (0.46-1.09)	0.12	0.69 (0.43-1.11)	0.13

Variables excluded after univariable regression are in grey; variables excluded after stepwise selection are marked with a dash. CI: Confidence interval; OR: Odds ratio; CD: Crohn's disease; TNF: tumor necrosis factor; HCT: Hematocrit; RCC: Red cell count; MCV: Mean corpuscular volume; MCH: Mean corpuscular hemoglobin; MCHC: Mean corpuscular hemoglobin concentration; ALP: Alkaline phosphatase; ALT: Alanine aminotransferase; GGT: Gamma-glutamyl transferase.

Table 3 Comparison of learning algorithms during cross-validation experiments

Algorithm	Dataset ¹	AUC (%)		P value ²
		Mean	SD	
Conventional statistics	Baseline clinical + biomarker data	65.9	9.5	-
Feed-forward ANN	Baseline clinical + biomarker data	71.0	8.7	0.25
Feed-forward ANN	Baseline biomarker data only	70.6	8.3	0.33
Recurrent neural network	Baseline and prior biomarker data	75.4	7.8	0.036

¹Clinical data refers to non-biochemical data such as age, sex, disease characteristics and concurrent treatments. Biomarker data refers to complete blood count, liver function tests and albumin.

²P value for comparison against conventional statistical algorithm, using the variance-corrected repeated *k*-fold *t* test. AUC: Area under the receiver operator characteristic curve; ANN: Artificial neural network.

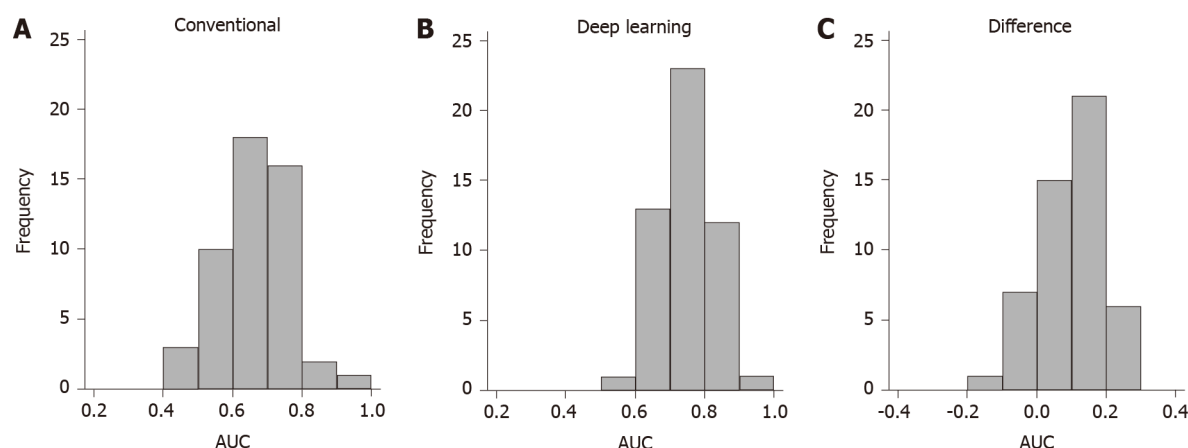


Figure 3 Distribution of area under the receiver operator characteristic curve after 10 × 5 fold cross validation. A: Conventional statistical learning algorithm (mean 0.659, SD 0.095); B: Recurrent neural network (mean 0.754, SD 0.078); C: Head-to-head comparison, matched at each fold and repetition (mean difference, + 0.095, *P* = 0.036). AUC: Area under the receiver operator characteristic curve.

Conventional statistical learning algorithms have generated many useful clinical scores, including the CD activity index[35], the simple endoscopic score for CD[36], scores to predict response to biologic therapies[16], and scores to differentiate CD from intestinal tuberculosis[37]. The advantage of conventional scores is often their simplicity and interpretability. A simple score can be memorized and calculated at the bed side and are intuitive as they utilize important risk factors of the outcome of interest. Yet clinical scores can only apply to a rather generic subgroup of patients and are never specific to any individual, as they utilize relatively few variables. Further, conventional methods are not readily able to model more complex, non-linear or time-dependent health states. With new genomic and microbiomic profiling, as well as the rapid uptake of comprehensive electronic medical records with mass data linkage, the ability of conventional learning algorithms to select useful predictive factors may become redundant[38].

Although the advantages of deep learning for the analysis of non-numerical data types is obvious, such as image data in endoscopy[39–41] and text or speech data in natural language processing[42], the utility of deep learning for the analysis of numerical data is less clear but remains promising. A recent study has demonstrated the utility of machine learning in predicting anti-TNF response in rheumatoid arthritis, but relied on genetic markers in addition to clinical data[43]. Another recent study

used machine learning to predict whether patients with ankylosing spondylitis required anti-TNF therapy, but did not evaluate whether response to therapy could be predicted[44]. It is anticipated that new data science and machine learning techniques are required to handle large amounts of data for use in clinical practice, although the optimal algorithms for this task remain unknown. Nevertheless, with the provision of comprehensive training data, machine learning tools have the potential to aid in individualized risk prediction, although no such model exists in IBD currently. In our cohort, the RNN deep learning algorithm was able to outperform the conventional algorithm after incorporating repeated biomarker measurements and thus additionally learn the non-linear temporal dynamics of the respective biomarkers – a feat that is not possible with conventional prediction models. It is expected that with enough training data, deep learning methods such as the RNN will be able to incorporate the time series data from multiple repeated health states of an individual patient over time. The clear trade-off with deep learning methods is the need for more data coordination and software to execute. However, the continued uptake of automated medical records in routine clinical practice may mitigate this limitation in future. Further, with the ever increasing breadth and volume of information from sources including comprehensive previous medical history, serum and fecal biomarkers, imaging and endoscopic data as well as genetics, the role of machine learning in prediction in chronic diseases including IBD is likely to expand.

This study has also demonstrated the importance of applying model validation techniques during model development[29]. ANNs and other powerful algorithms have the ability to learn intricate differences in data, yet poorly specified models that focus only on learning power have the propensity to learn the random variations or artefacts in the data, which are present only due to chance. This is evidenced by the RNN in this study achieving excellent AUC during training, but a reduced AUC when tested on unseen data (naïve AUC 0.892; cross-validated AUC 0.754). The same phenomenon occurred with the statistical learning algorithm but to a somewhat lesser extent (naïve AUC 0.771; cross-validated AUC 0.659). Therefore, studies developing predictive models should take care to avoid naïvely assessing predictive performance and ensure that effective cross-validation or bootstrapping methods are used for appropriate interval validation[4]. If available, external validation of predictive models in entirely new and different cohorts is the gold standard for model validation[4].

The dataset used in this study was retrospective and from a single center which subjects the results to information bias and limits their external validity. The outcome used was biochemical remission as this is a readily available as a repeated measure which allowed demonstration of more conventional and machine learning models, however it is acknowledged that clinical symptoms and/or mucosal healing are more clinically relevant end-points. Nevertheless, the goal of this study was to demonstrate the feasibility of deep learning methods in clinical prediction in this proof-of-concept study, rather than to develop a specific predictive model. Further, in practice, much larger cohorts will be required to properly train and calibrate deep learning models to maximize their utility in the real world. In future, all studies investigating specific predictive models should be subject to prospective controlled validation prior their application in clinical practice, specifically having shown that outcomes are improved after using predictive models to guide management.

CONCLUSION

In conclusion, we have demonstrated the feasibility of deep learning algorithms for clinical prediction in CD, which demonstrated an improved predictive performance compared to conventional methods. However, conventional statistical methods retain the advantage of simplicity and intuitiveness, allowing their use at the bedside. Yet with the rapid expansion of available health data, machine learning models have the potential to supersede currently conventional methods and greatly improve the development of tools for the clinical prediction of patient outcomes.

ARTICLE HIGHLIGHTS

Research background

Machine learning and artificial intelligence have the potential to revolutionize precision care in inflammatory bowel diseases. The greatest area of interest has been

the application of deep learning methods in automatic tumor detection during endoscopy, yet the application of such techniques in clinical outcome prediction has been lacking.

Research motivation

Traditional approaches to clinical prediction rely on conventional statistical algorithms such as regression, which are not suitable for more complex data such as repeated biomarker measurements.

Research objectives

To determine and compare the utility of deep learning with conventional algorithms in predicting response to anti-tumor necrosis factor (anti-TNF) therapy in Crohn's disease (CD).

Research methods

A retrospective cohort of CD patients commenced on anti-TNF therapy was used to experimentally develop and cross-validate three supervised learning algorithms: (1) Statistical learning algorithm; (2) Feed-forward artificial neural network; and (3) Recurrent neural network with repeated data. Predictive utility was quantified using the area under the receiver operator characteristic curve (AUC).

Research results

Within our cohort of 146 patients, the conventional statistical learning algorithm had the weakest performance [AUC 0.659, 95% confidence interval (CI): 0.562-0.756], compared to the feed-forward artificial neural network (AUC 0.710, 95% CI: 0.622-0.799; $P = 0.25$ *vs* conventional) and the recurrent neural network using repeated biomarker measurements (AUC 0.754, 95% CI: 0.674-0.834; $P = 0.036$ *vs* conventional).

Research conclusions

Deep learning methods are feasible and have the potential for stronger predictive performance compared to conventional model building methods when applied to predicting remission after anti-TNF therapy in CD.

Research perspectives

This has been the first study to investigate the utility of deep neural networks in predicting clinical outcomes using repeated clinical data in inflammatory bowel disease. Future studies should incorporate additional data types such as genetic, imaging and endoscopic factors.

REFERENCES

- 1 Podolsky DK. Inflammatory bowel disease. *N Engl J Med* 2002; **347**: 417-429 [PMID: 12167685 DOI: 10.1056/NEJMra020831]
- 2 Jackson BD, Con D, Gorelik A, Liew D, Knowles S, De Cruz P. Examination of the relationship between disease activity and patient-reported outcome measures in an inflammatory bowel disease cohort. *Intern Med J* 2018; **48**: 1234-1241 [PMID: 29663629 DOI: 10.1111/imj.13937]
- 3 Denson LA, Curran M, McGovern DPB, Koltun WA, Duerr RH, Kim SC, Sartor RB, Sylvester FA, Abraham C, de Zoeten EF, Siegel CA, Burns RM, Dobes AM, Shtraizent N, Honig G, Heller CA, Hurtado-Lorenzo A, Cho JH. Challenges in IBD Research: Precision Medicine. *Inflamm Bowel Dis* 2019; **25**: S31-S39 [PMID: 31095701 DOI: 10.1093/ibd/izz078]
- 4 Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015; **162**: W1-73 [PMID: 25560730 DOI: 10.7326/M14-0698]
- 5 Chen H, Sung JY. Potentials of AI in medical image analysis in Gastroenterology and Hepatology. *J Gastroenterol Hepatol* 2021; **36**: 31-38 [PMID: 33140875 DOI: 10.1111/jgh.15327]
- 6 Le Berre C, Sandborn WJ, Aridhi S, Devignes MD, Fournier L, Smail-Tabbone M, Danese S, Peyrin-Biroulet L. Application of Artificial Intelligence to Gastroenterology and Hepatology. *Gastroenterology* 2020; **158**: 76-94.e2 [PMID: 31593701 DOI: 10.1053/j.gastro.2019.08.058]
- 7 Kohli A, Holzwanger EA, Levy AN. Emerging use of artificial intelligence in inflammatory bowel disease. *World J Gastroenterol* 2020; **26**: 6923-6928 [PMID: 33311940 DOI: 10.3748/wjg.v26.i44.6923]
- 8 Takenaka K, Ohtsuka K, Fujii T, Negi M, Suzuki K, Shimizu H, Oshima S, Akiyama S, Motobayashi M, Nagahori M, Saito E, Matsuoka K, Watanabe M. Development and Validation of a

- Deep Neural Network for Accurate Evaluation of Endoscopic Images From Patients With Ulcerative Colitis. *Gastroenterology* 2020; **158**: 2150-2157 [PMID: [32060000](#) DOI: [10.1053/j.gastro.2020.02.012](#)]
- 9 **Otani K**, Nakada A, Kurose Y, Niikura R, Yamada A, Aoki T, Nakanishi H, Doyama H, Hasatani K, Sumiyoshi T, Kitsuregawa M, Harada T, Koike K. Automatic detection of different types of small-bowel lesions on capsule endoscopy images using a newly developed deep convolutional neural network. *Endoscopy* 2020; **52**: 786-791 [PMID: [32557474](#) DOI: [10.1055/a-1167-8157](#)]
 - 10 **Klang E**, Barash Y, Margalit RY, Soffer S, Shimon O, Albshesh A, Ben-Horin S, Amitai MM, Eliakim R, Kopylov U. Deep learning algorithms for automated detection of Crohn's disease ulcers by video capsule endoscopy. *Gastrointest Endosc* 2020; **91**: 606-613.e2 [PMID: [31743689](#) DOI: [10.1016/j.gie.2019.11.012](#)]
 - 11 **Sze SF**, Cheung WI, Wong WC, Hui YT, Lam JTW. AmplifEYE assisted colonoscopy vs standard colonoscopy: A randomized controlled study. *J Gastroenterol Hepatol* 2021; **36**: 376-382 [PMID: [33141979](#) DOI: [10.1111/jgh.15331](#)]
 - 12 **Abbas M**, Matta J, Le T, Bensmail H, Obafemi-Ajayi T, Honavar V, El-Manzalawy Y. Biomarker discovery in inflammatory bowel diseases using network-based feature selection. *PLoS One* 2019; **14**: e0225382 [PMID: [31756219](#) DOI: [10.1371/journal.pone.0225382](#)]
 - 13 **Bodein A**, Chapleur O, Droit A, Lê Cao KA. A Generic Multivariate Framework for the Integration of Microbiome Longitudinal Studies With Other Data Types. *Front Genet* 2019; **10**: 963 [PMID: [31803221](#) DOI: [10.3389/fgene.2019.00963](#)]
 - 14 **Matsuoka K**, Hamada S, Shimizu M, Nanki K, Mizuno S, Kiyohara H, Arai M, Sugimoto S, Iwao Y, Ogata H, Hisamatsu T, Naganuma M, Kanai T, Mochizuki M, Hashiguchi M. Factors predicting the therapeutic response to infliximab during maintenance therapy in Japanese patients with Crohn's disease. *PLoS One* 2018; **13**: e0204632 [PMID: [30286108](#) DOI: [10.1371/journal.pone.0204632](#)]
 - 15 **Ding NS**, Malietzis G, Lung PFC, Penez L, Yip WM, Gabe S, Jenkins JT, Hart A. The body composition profile is associated with response to anti-TNF therapy in Crohn's disease and may offer an alternative dosing paradigm. *Aliment Pharmacol Ther* 2017; **46**: 883-891 [PMID: [28881017](#) DOI: [10.1111/apt.14293](#)]
 - 16 **Barber GE**, Yajnik V, Khalili H, Giallourakis C, Garber J, Xavier R, Ananthakrishnan AN. Genetic Markers Predict Primary Non-Response and Durable Response To Anti-TNF Biologic Therapies in Crohn's Disease. *Am J Gastroenterol* 2016; **111**: 1816-1822 [PMID: [27596696](#) DOI: [10.1038/ajg.2016.408](#)]
 - 17 **Ward MG**, Warner B, Unsworth N, Chuah SW, Brownclarke C, Shieh S, Parkes M, Sanderson JD, Arkir Z, Reynolds J, Gibson PR, Irving PM. Infliximab and adalimumab drug levels in Crohn's disease: contrasting associations with disease activity and influencing factors. *Aliment Pharmacol Ther* 2017; **46**: 150-161 [PMID: [28481014](#) DOI: [10.1111/apt.14124](#)]
 - 18 **Mortensen JH**, van Haaften WT, Karsdal MA, Bay-Jensen AC, Olinga P, Grønbaek H, Hvas CL, Manon-Jensen T, Dijkstra G, Dige A. The Citrullinated and MMP-degraded Vimentin Biomarker (VICM) Predicts Early Response to Anti-TNF α Treatment in Crohn's Disease. *J Clin Gastroenterol* 2021; **55**: 59-66 [PMID: [32301833](#) DOI: [10.1097/MCG.0000000000001341](#)]
 - 19 **Con D**, Parthasarathy N, Bishara M, Lubner RP, Joshi N, Wan A, Rickard JA, Long T, Connoley DJ, Sparrow MP, Gibson PR, van Langenberg DR, Vasudevan A. Development of a Simple, Serum Biomarker-based Model Predictive of the Need for Early Biologic Therapy in Crohn's Disease. *J Crohns Colitis* 2021; **15**: 583-593 [PMID: [32949458](#) DOI: [10.1093/ecco-jcc/jjaa194](#)]
 - 20 **Cornish JS**, Wirthgen E, Däbritz J. Biomarkers Predictive of Response to Thiopurine Therapy in Inflammatory Bowel Disease. *Front Med (Lausanne)* 2020; **7**: 8 [PMID: [32064265](#) DOI: [10.3389/fmed.2020.00008](#)]
 - 21 **Waljee AK**, Wallace BI, Cohen-Mekelburg S, Liu Y, Liu B, Sauder K, Stidham RW, Zhu J, Higgins PDR. Development and Validation of Machine Learning Models in Prediction of Remission in Patients With Moderate to Severe Crohn Disease. *JAMA Netw Open* 2019; **2**: e193721 [PMID: [31074823](#) DOI: [10.1001/jamanetworkopen.2019.3721](#)]
 - 22 **Waljee AK**, Lipson R, Wiitala WL, Zhang Y, Liu B, Zhu J, Wallace B, Govani SM, Stidham RW, Hayward R, Higgins PDR. Predicting Hospitalization and Outpatient Corticosteroid Use in Inflammatory Bowel Disease Patients Using Machine Learning. *Inflamm Bowel Dis* 2017; **24**: 45-53 [PMID: [29272474](#) DOI: [10.1093/ibd/izx007](#)]
 - 23 **Noh SM**, Oh EH, Park SH, Lee JB, Kim JY, Park JC, Kim J, Ham NS, Hwang SW, Yang DH, Byeon JS, Myung SJ, Yang SK, Ye BD. Association of Faecal Calprotectin Level and Combined Endoscopic and Radiological Healing in Patients With Crohn's Disease Receiving Anti-tumour Necrosis Factor Therapy. *J Crohns Colitis* 2020; **14**: 1231-1240 [PMID: [32157278](#) DOI: [10.1093/ecco-jcc/jjaa042](#)]
 - 24 **Maaser C**, Sturm A, Vavricka SR, Kucharzik T, Fiorino G, Annese V, Calabrese E, Baumgart DC, Bettenworth D, Borralho Nunes P, Burisch J, Castiglione F, Eliakim R, Ellul P, González-Lama Y, Gordon H, Halligan S, Katsanos K, Kopylov U, Kotze PG, Krustinš E, Laghi A, Limdi JK, Rieder F, Rimola J, Taylor SA, Tolan D, van Rheenen P, Verstockt B, Stoker J; European Crohn's and Colitis Organisation [ECCO] and the European Society of Gastrointestinal and Abdominal Radiology [ESGAR]. ECCO-ESGAR Guideline for Diagnostic Assessment in IBD Part 1: Initial diagnosis, monitoring of known IBD, detection of complications. *J Crohns Colitis* 2019; **13**: 144-164 [PMID: [30137275](#) DOI: [10.1093/ecco-jcc/jjy113](#)]
 - 25 **Porter AC**, Aubrecht J, Birch C, Braun J, Cuff C, Dasgupta S, Gale JD, Hinton R, Hoffmann SC, Honig G, Linggi B, Schito M, Castele NV, Sauer JM. Biomarkers of Crohn's Disease to Support the

- Development of New Therapeutic Interventions. *Inflamm Bowel Dis* 2020; **26**: 1498-1508 [PMID: 32840322 DOI: 10.1093/ibd/izaa215]
- 26 **Ma C**, Battat R, Parker CE, Khanna R, Jairath V, Feagan BG. Update on C-reactive protein and fecal calprotectin: are they accurate measures of disease activity in Crohn's disease? *Expert Rev Gastroenterol Hepatol* 2019; **13**: 319-330 [PMID: 30791776 DOI: 10.1080/17474124.2019.1563481]
- 27 **Lin X**, Qiu Y, Feng R, Chen B, He Y, Zeng Z, Zhang S, Chen M, Mao R. Normalization of C-Reactive Protein Predicts Better Outcome in Patients With Crohn's Disease With Mucosal Healing and Deep Remission. *Clin Transl Gastroenterol* 2020; **11**: e00135 [PMID: 32463625 DOI: 10.14309/ctg.000000000000135]
- 28 **Click B**, Vargas EJ, Anderson AM, Proksell S, Koutroubakis IE, Ramos Rivers C, Hashash JG, Regueiro M, Watson A, Dunn MA, Schwartz M, Swoger J, Baidoo L, Barrie A 3rd, Binion DG. Silent Crohn's Disease: Asymptomatic Patients with Elevated C-reactive Protein Are at Risk for Subsequent Hospitalization. *Inflamm Bowel Dis* 2015; **21**: 2254-2261 [PMID: 26197446 DOI: 10.1097/MIB.0000000000000516]
- 29 **Bouckaert RR**, Frank E. Evaluating the Replicability of Significance Tests for Comparing Learning Algorithms, in Advances in Knowledge Discovery and Data Mining. In: Dai H, Srikant R, Zhang C. Lecture Notes in Computer Science. Springer: Berlin, Heidelberg, 2004
- 30 **Hajian-Tilaki K**. Sample size estimation in diagnostic test studies of biomedical informatics. *J Biomed Inform* 2014; **48**: 193-204 [PMID: 24582925 DOI: 10.1016/j.jbi.2014.02.013]
- 31 **Torres J**, Mehndru S, Colombel JF, Peyrin-Biroulet L. Crohn's disease. *Lancet* 2017; **389**: 1741-1755 [PMID: 27914655 DOI: 10.1016/S0140-6736(16)31711-1]
- 32 **Con D**, Jackson B, Gray K, De Cruz P. eHealth for inflammatory bowel disease self-management - the patient perspective. *Scand J Gastroenterol* 2017; **52**: 973-980 [PMID: 28598210 DOI: 10.1080/00365521.2017.1333625]
- 33 **van der Valk ME**, Mangen MJ, Severs M, van der Have M, Dijkstra G, van Bodegraven AA, Fidder HH, de Jong DJ, van der Woude CJ, Romberg-Camps MJ, Clemens CH, Jansen JM, van de Meeberg PC, Mahmmoud N, van der Meulen-de Jong AE, Ponsioen CY, Bolwerk C, Vermeijden JR, Siersema PD, Leenders M, Oldenburg B; COIN study group and the Dutch Initiative on Crohn and Colitis. Evolution of Costs of Inflammatory Bowel Disease over Two Years of Follow-Up. *PLoS One* 2016; **11**: e0142481 [PMID: 27099937 DOI: 10.1371/journal.pone.0142481]
- 34 **Jackson B**, Con D, Ma R, Gorelik A, Liew D, De Cruz P. Health care costs associated with Australian tertiary inflammatory bowel disease care. *Scand J Gastroenterol* 2017; **52**: 851-856 [PMID: 28509590 DOI: 10.1080/00365521.2017.1323117]
- 35 **Best WR**, Beckett JM, Singleton JW, Kern F Jr. Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study. *Gastroenterology* 1976; **70**: 439-444 [PMID: 1248701]
- 36 **Daperno M**, D'Haens G, Van Assche G, Baert F, Bulois P, Maunoury V, Sostegni R, Rocca R, Pera A, Gevers A, Mary JY, Colombel JF, Rutgeerts P. Development and validation of a new, simplified endoscopic activity score for Crohn's disease: the SES-CD. *Gastrointest Endosc* 2004; **60**: 505-512 [PMID: 15472670 DOI: 10.1016/s0016-5107(04)01878-4]
- 37 **Limsrivilai J**, Pausawasdi N. Intestinal tuberculosis or Crohn's disease: a review of the diagnostic models designed to differentiate between these two gastrointestinal diseases. *Intest Res* 2021; **19**: 21-32 [PMID: 32311862 DOI: 10.5217/ir.2019.09142]
- 38 **Sung JJ**, Stewart CL, Freedman B. Artificial intelligence in health care: preparing for the fifth Industrial Revolution. *Med J Aust* 2020; **213**: 253-255.e1 [PMID: 32892395 DOI: 10.5694/mja2.50755]
- 39 **Iwagami H**, Ishihara R, Aoyama K, Fukuda H, Shimamoto Y, Kono M, Nakahira H, Matsuura N, Shichijo S, Kanesaka T, Kanzaki H, Ishii T, Nakatani Y, Tada T. Artificial intelligence for the detection of esophageal and esophagogastric junctional adenocarcinoma. *J Gastroenterol Hepatol* 2021; **36**: 131-136 [PMID: 32511793 DOI: 10.1111/jgh.15136]
- 40 **East JE**, Rittscher J. Artificial intelligence for colonoscopic polyp detection: High performance vs human nature. *J Gastroenterol Hepatol* 2020; **35**: 1663-1664 [PMID: 33043510 DOI: 10.1111/jgh.15262]
- 41 **Parasher G**, Wong M, Rawat M. Evolving role of artificial intelligence in gastrointestinal endoscopy. *World J Gastroenterol* 2020; **26**: 7287-7298 [PMID: 33362384 DOI: 10.3748/wjg.v26.i46.7287]
- 42 **Shung D**, Tsay C, Laine L, Chang D, Li F, Thomas P, Partridge C, Simonov M, Hsiao A, Tay JK, Taylor A. Early identification of patients with acute gastrointestinal bleeding using natural language processing and decision rules. *J Gastroenterol Hepatol* 2021; **36**: 1590-1597 [PMID: 33105045 DOI: 10.1111/jgh.15313]
- 43 **Guan Y**, Zhang H, Quang D, Wang Z, Parker SCJ, Pappas DA, Kremer JM, Zhu F. Machine Learning to Predict Anti-Tumor Necrosis Factor Drug Responses of Rheumatoid Arthritis Patients by Integrating Clinical and Genetic Markers. *Arthritis Rheumatol* 2019; **71**: 1987-1996 [PMID: 31342661 DOI: 10.1002/art.41056]
- 44 **Lee S**, Eun Y, Kim H, Cha HS, Koh EM, Lee J. Machine learning to predict early TNF inhibitor users in patients with ankylosing spondylitis. *Sci Rep* 2020; **10**: 20299 [PMID: 33219239 DOI: 10.1038/s41598-020-75352-7]



Published by **Baishideng Publishing Group Inc**
7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA

Telephone: +1-925-3991568

E-mail: bpgoffice@wjgnet.com

Help Desk: <https://www.f6publishing.com/helpdesk>

<https://www.wjgnet.com>

