

World Journal of *Gastroenterology*

World J Gastroenterol 2022 February 7; 28(5): 502-607



Contents

Weekly Volume 28 Number 5 February 7, 2022

FRONTIER

- 502 Colorectal cancer screening and surveillance in patients with inflammatory bowel disease in 2021
Huguet JM, Ferrer-Barceló L, Suárez P, Sanchez E, Prieto JD, Garcia V, Sempere J

MINIREVIEWS

- 517 Viral hepatitis: Innovations and expectations
Leoni S, Casabianca A, Biagioni B, Serio I

ORIGINAL ARTICLE

Basic Study

- 532 Effect of *Bacillus subtilis*, *Enterococcus faecium*, and *Enterococcus faecalis* supernatants on serotonin transporter expression in cells and tissues
Chen YM, Li Y, Wang X, Wang ZL, Hou JJ, Su S, Zhong WL, Xu X, Zhang J, Wang BM, Wang YM
- 547 Connective tissue growth factor expression hints at aggressive nature of colorectal cancer
Bhat IP, Rather TB, Maqbool I, Rashid G, Akhtar K, Bhat GA, Parray FQ, Syed B, Khan IY, Kazi M, Hussain MD, Syed M

Retrospective Cohort Study

- 570 Abnormal liver chemistries as a predictor of COVID-19 severity and clinical outcomes in hospitalized patients
Krishnan A, Prichett L, Tao X, Alqahtani SA, Hamilton JP, Mezey E, Strauss AT, Kim A, Potter JJ, Chen PH, Woreta TA

CASE REPORT

- 588 Simultaneous endoscopic and video-assisted retroperitoneal debridement in walled-off pancreatic necrosis using a laparoscopic access platform: Two case reports
Lindgaard L, Lauritsen ML, Novovic S, Hansen EF, Karstensen JG, Schmidt PN
- 594 Curative resection with endoscopic submucosal dissection of early gastric cancer in *Helicobacter pylori*-negative Ménétrier's disease: A case report
Fukushi K, Goda K, Kino H, Kondo M, Kanazawa M, Kashima K, Kanamori A, Abe K, Suzuki T, Tominaga K, Yamagishi H, Irisawa A

LETTER TO THE EDITOR

- 602 Artificial intelligence model validation before its application in clinical diagnosis assistance
Vazquez-Zapien GJ, Mata-Miranda MM, Garibay-Gonzalez F, Sanchez-Brito M
- 605 Machine learning models and over-fitting considerations
Charilaou P, Battat R

ABOUT COVER

Editorial Board Member of *World Journal of Gastroenterology*, Guang Ji, MD, PhD, Professor, Institute of Digestive Diseases, Longhua Hospital, Shanghai University of Traditional Chinese Medicine, No. 725 South Wanping Road, Shanghai 200032, China. jiliver@vip.sina.com

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: *Ying-Yi Yuan*, Production Department Director: *Xiang Li*, 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

EDITORIAL BOARD MEMBERS

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

PUBLICATION DATE

February 7, 2022

COPYRIGHT

© 2022 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>



Machine learning models and over-fitting considerations

Paris Charilaou, Robert Battat

ORCID number: Paris Charilaou
[0000-0002-5512-4225](https://orcid.org/0000-0002-5512-4225); Robert Battat
[0000-0002-7421-9764](https://orcid.org/0000-0002-7421-9764).

Author contributions: Charilaou P and Battat R drafted and edited the manuscript, and reviewed the intellectual content.

Conflict-of-interest statement: The authors have no conflict of interest to declare.

Country/Territory of origin: United States

Specialty type: Gastroenterology and hepatology

Provenance and peer review: Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

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

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

Paris Charilaou, Robert Battat, Jill Roberts Center for Inflammatory Bowel Disease - Division of Gastroenterology & Hepatology, Weill Cornell Medicine, New York, NY 10021, United States

Corresponding author: Robert Battat, MD, Assistant Professor, Jill Roberts Center for Inflammatory Bowel Disease - Division of Gastroenterology & Hepatology, Weill Cornell Medicine, 1315 York Avenue, New York, NY 10021, United States. rob9175@med.cornell.edu

Abstract

Machine learning models may outperform traditional statistical regression algorithms for predicting clinical outcomes. Proper validation of building such models and tuning their underlying algorithms is necessary to avoid over-fitting and poor generalizability, which smaller datasets can be more prone to. In an effort to educate readers interested in artificial intelligence and model-building based on machine-learning algorithms, we outline important details on cross-validation techniques that can enhance the performance and generalizability of such models.

Key Words: Machine learning; Over-fitting; Cross-validation; Hyper-parameter tuning

©The Author(s) 2022. Published by Baishideng Publishing Group Inc. All rights reserved.

Core Tip: Machine learning models are increasingly being used in clinical medicine to predict outcomes. Proper validation techniques of these models are essential to avoid over-fitting and poor generalization on new data.

Citation: Charilaou P, Battat R. Machine learning models and over-fitting considerations. *World J Gastroenterol* 2022; 28(5): 605-607

URL: <https://www.wjgnet.com/1007-9327/full/v28/i5/605.htm>

DOI: <https://dx.doi.org/10.3748/wjg.v28.i5.605>

TO THE EDITOR

Con *et al*[1] explore artificial intelligence (AI) in a classification problem of predicting biochemical remission of Crohn's disease at 12 mo post-induction with infliximab or adalimumab. They illustrate that, after applying appropriate machine learning (ML) methodologies, ML methods outperform conventional multivariable logistic

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: <https://creativecommons.org/licenses/by-nc/4.0/>

Received: October 26, 2021

Peer-review started: October 26, 2021

First decision: December 27, 2021

Revised: December 29, 2021

Accepted: January 14, 2022

Article in press: January 14, 2022

Published online: February 7, 2022

P-Reviewer: Calabro F, Dabbakuti JRKK, Guo XY, Stoyanov D

S-Editor: Gong ZM

L-Editor: A

P-Editor: Gong ZM



regression (a statistical learning algorithm). The area-under-the-curve (AUC) was the chosen performance metric for comparison and cross-validation was performed.

Their study elucidates a few important points regarding the utilization of ML. First, the use of repeated k-fold cross-validation, which is primarily utilized to prevent over-fitting of the models. This technique, while common in ML, it has not been traditionally used in conventional regression models in the literature so far. Especially in small datasets, such as in their study ($n = 146$), linear (and non-linear, in the case of neural networks) relationships risk being “learned” by chance, leading to poor generalization of the models when applied to previously “unseen” or future data points. It was evident from their analysis that the “naïve” AUCs (training the model on all the data), was significantly higher than the mean cross-validated AUCs, in all 3 models, suggestive of “over-fitting” when one does not cross-validate. Smaller datasets tend to be more susceptible to over-fitting as they are less likely to accurately represent the population in question.

Second, the authors utilized “hyper-parameter tuning” for their neural network models, where the otherwise arbitrarily selected “settings” (or hyper-parameters, such as the number of inner neuron layers and number of neurons per layer) of the neural network are chosen based on performance. Hyper-parameters cannot be “learned” or “optimized” by simply fitting the model (as it happens with predictor coefficients), and the only way to discover the best values is by fitting the model with various combinations and assessing its performance. The combinations can be evaluated stochastically (randomly or *via* a Bayes-based approach) or using a grid approach (*e.g.*, for 3 hyper-parameters that take 5 potential values, there are $5 \times 5 \times 5 = 5^3 = 125$ combinations to evaluate) over k times. One may ask, if one was to fit a model $125 \times k$ times, on 146 observations, is not there a risk for over-fitting the “optimal” hyper-parameter values? To avoid such a problem, nested k-fold cross-validation must be performed: within each repeated k-fold training data subset, a sub-k-fold “inner” training/validation must be done to evaluate each hyper-parameter combination. In this way, we overcome potential bias to optimistic model performance, which can occur when we use the same cross-validation procedure and dataset to both tune the hyper-parameters and evaluate the model’s performance metrics (*e.g.*, AUC)[2]. The authors did not elaborate on how the hyperparameter tuning was performed.

Another point to consider in k-fold cross-validation in small datasets is the number of k-folds used, specifically in classification problems (*i.e.*, yes/no binary outcomes). In this study[1], the outcome prevalence was 64% ($n \approx 93$). With a chosen $k = 5$, the training folds would comprise 80% of that data, leading to approximately 74 positive cases of biochemical remission. The number of positive outcomes in each training fold must be considered, especially in logistic regression, where the rule of thumb recommends at least ten positive events per independent predictor, to minimize over-fitting[3]. In this study[1], six predictors were eventually used in the multivariable model, making over-fitting less likely from a model-specification standpoint. Finally, k-folds are recommended to be stratified by the outcome, so the outcome prevalence is equal among the training and testing folds. This becomes crucial when the prevalence of outcome of interest is $< 10\%-20\%$ (imbalanced classification problem). While imbalanced classification is not an issue in this study[1], the authors did not mention whether they used outcome-stratified k-folds.

Lastly, the endpoint utilized, CRP normalization, has poor specificity for endoscopic inflammation in Crohn’s disease[4]. More robust endpoints would include endoscopic inflammation and/or deep remission using validated disease activity indices[5].

We congratulate the authors for their effort, which acts both as a proof-of-concept for using ML in improved prediction of outcomes in IBD, but also for the methodologies outlined to reduce over-fitting. In general, with the advent of AI and specifically ML-based models in IBD[6], it is important to recognize that while now we have the tools to construct more accurate models and enhance precision medicine, most ML-based models, such as artificial neural networks, lack in being intuitively interpretable (*i.e.*, “black-box”). Efforts in “explainable AI” are under way[7], hopefully eliminating the “black-box” concept in future clinical decision tools. Applying these to validated disease activity assessments will be essential for prediction models in future studies.

REFERENCES

- 1 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: 6476-6488 [PMID: 34720536 DOI: 10.3748/wjg.v27.i38.6476]

- 2 **Cawley GC**, Talbot NLC. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J Mach Learn Res* 2010; **11**: 2079-2107
- 3 **Peduzzi P**, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996; **49**: 1373-1379 [PMID: 8970487 DOI: [10.1016/s0895-4356\(96\)00236-3](https://doi.org/10.1016/s0895-4356(96)00236-3)]
- 4 **Mosli MH**, Zou G, Garg SK, Feagan SG, MacDonald JK, Chande N, Sandborn WJ, Feagan BG. C-Reactive Protein, Fecal Calprotectin, and Stool Lactoferrin for Detection of Endoscopic Activity in Symptomatic Inflammatory Bowel Disease Patients: A Systematic Review and Meta-Analysis. *Am J Gastroenterol* 2015; **110**: 802-19; quiz 820 [PMID: 25964225 DOI: [10.1038/ajg.2015.120](https://doi.org/10.1038/ajg.2015.120)]
- 5 **Turner D**, Ricciuto A, Lewis A, D'Amico F, Dhaliwal J, Griffiths AM, Bettenworth D, Sandborn WJ, Sands BE, Reinisch W, Schölmerich J, Bemelman W, Danese S, Mary JY, Rubin D, Colombel JF, Peyrin-Biroulet L, Dotan I, Abreu MT, Dignass A; International Organization for the Study of IBD. STRIDE-II: An Update on the Selecting Therapeutic Targets in Inflammatory Bowel Disease (STRIDE) Initiative of the International Organization for the Study of IBD (IOIBD): Determining Therapeutic Goals for Treat-to-Target strategies in IBD. *Gastroenterology* 2021; **160**: 1570-1583 [PMID: 33359090 DOI: [10.1053/j.gastro.2020.12.031](https://doi.org/10.1053/j.gastro.2020.12.031)]
- 6 **Chen G**, Shen J. Artificial Intelligence Enhances Studies on Inflammatory Bowel Disease. *Front Bioeng Biotechnol* 2021; **9**: 635764 [PMID: 34307315 DOI: [10.3389/fbioe.2021.635764](https://doi.org/10.3389/fbioe.2021.635764)]
- 7 **Linardatos P**, Papastefanopoulos V, Kotsiantis S. Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy (Basel)* 2020; **23** [PMID: 33375658 DOI: [10.3390/e23010018](https://doi.org/10.3390/e23010018)]



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>

