J Pediatr Gastroenterol Nutr. 2021 Feb 1;72(2):262-269. doi: 10.1097/MPG.0000000000002956.

Accurate Classification of Pediatric Colonic Inflammatory Bowel Disease Subtype Using a Random Forest Machine Learning Classifier

Jasbir Dhaliwal [1 2], Lauren Erdman [3], Erik Drysdale [3], Firas Rinawi [1 2], Jennifer Muir [4], Thomas D Walters [1 2], Iram Siddiqui [4], Anne M Griffiths [1 2], Peter C Church [1 2]

Affiliations

PMID: 33003163 DOI: 10.1097/MPG.0000000000002956

Abstract

Background: The pediatric inflammatory bowel disease (PIBD) classes algorithm was developed to bring consistency to labelling of colonic IBD, but labels are exclusively based on features atypical for ulcerative colitis (UC).

Aim: The aim of the study was to develop an algorithm and identify features that discriminate between pediatric UC and colonic Crohn disease (CD).

Methods: Baseline clinical, endoscopic, radiologic, and histologic data, including the PIBD class features in 74 colonic IBD (56: UC, 18: colonic CD) patients were collected. The PIBD class features and additional features common to UC were used to perform initial clustering, using similarity network fusion (SNF). We trained a Random Forest (RF) classifier on the full dataset and used a leave-one-out approach to evaluate model accuracy. The top-features were used to build a new classifier, which we tested on 15 previously unused patients. We then performed clustering with SNF on the top RF features and assessed ability to discriminate between UC and colonic-CD independent of a supervised model.

Results: The initial SNF clustering with 58 patients demonstrated 2 groups: group 1 (n = 39, 90% UC) and group 2 (n = 19, 68% colonic-CD). Our RF classifier correctly labelled 97% of the 58 patients based on leave-one-out cross validation and identified the 7 most important features (3 histological and 4 endoscopic) to clinically distinguish these groups. We trained a new RF classifier with the top 7 features and found 100% accuracy in a set of 15 held-out patients. Finally, post hoc clustering with these 7 features revealed 2 groups of patients: group 1 (n = 55, 98% UC) and group 2 (n = 18, 94% colonic-CD).

Conclusions: A combination of supervised and unsupervised analyses identified a short list of features, which consistently distinguish UC from colonic CD. Future directions include validation in other populations.

Comment in

Artificial Intelligence and Gastroenterology: Expanding Technology.
Vanderpool CPB.
J Pediatr Gastroenterol Nutr. 2021 Feb 1;72(2):183. doi: 10.1097/MPG.0000000000002986.
PMID: 33264180 No abstract available.

Related information

MedGen

LinkOut – more resources

Full Text Sources
Ovid Technologies, Inc.
Wolters Kluwer

Other Literature Sources
scite Smart Citations

Medical
MedlinePlus Health Information

Research Materials
NCI CPTC Antibody Characterization Program