Developing a Neural Network Model for a Non-invasive Prediction of Histologic Activity in Inflammatory Bowel Diseases

Iolanda Valentina Popa[1,2], Mircea Diculescu[3,4], Catalina Mihai[1,2], Cristina Cijevschi Prelipcean[1,2], Alexandru Burlacu[1,5,6]

[1]University of Medicine and Pharmacy "Gr. T. Popa", Iași, Romania
[2]Institute of Gastroenterology and Hepatology, Iași, Romania
[3]Department of Gastroenterology, Fundeni Clinical Institute, Bucharest, Romania
[4]'Carol Davila' University of Medicine and Pharmacy, Bucharest, Romania
[5]Department of Interventional Cardiology, Cardiovascular Diseases Institute, Iași, Romania
[6]Romanian Academy of Medical Sciences, Bucharest, Romania

Cite this article as: Popa IV, Diculescu M, Mihai C, Cijevschi Prelipcean C, Burlacu A. Developing a neural network model for a non-invasive prediction of histologic activity in inflammatory bowel diseases. *Turk J Gastroenterol.* 2021; 32(3): 276-286.

ABSTRACT

Background: *Colonoscopy with biopsy is the "gold" standard for evaluating disease activity in inflammatory bowel diseases (IBD). Current research is geared toward finding non-invasive, cost-efficient methods that estimate disease activity. We aimed to develop a neural network (NN) model for the non-invasive prediction of histologic activity in IBD using routinely available clinical-biological parameters.*

Methods: *Standard clinical-biological parameters and histologic activity from 371 ulcerative colitis (UC) and 115 Crohn's disease (CD) patient records were collected. A training set, a test set, and a validation set were used for building/validating 2 models for each disease. All models had binary output predicting the active/inactive histologic disease status. For both diseases, the first model used both clinical and biological inputs, while the second used only biological data.*

Results: *First UC model obtained an accuracy of 95.59% on the test set and 96.67% on the validation set. The second UC model achieved accuracies of 88.24% and 86.67% on the test and validation sets, respectively. The First CD classifier resulted in 90.48% accuracy on the test set and 91.67% on the validation set. Finally, the second CD classifier obtained an accuracy of 85.71% on the test set and 91.67% on the validation set.*

Conclusions: *An accurate and non-invasive artificial intelligence system to predict histologic disease activity in IBD is designed. Our models achieved similar or better results compared to the documented performance of fecal calprotectin (the best non-invasive IBD biomarker to date). Given these favorable results, we anticipate the future utility in the clinical setting of a non-invasive disease activity prediction.*

Keywords: *Inflammatory bowel diseases, artificial intelligence, neural networks, predictive model, histologic disease activity*

INTRODUCTION

Inflammatory bowel diseases (IBD) consist of ulcerative colitis (UC) and Crohn's disease (CD) and are characterized by chronic recurrent inflammation of the gastrointestinal tract. Assessing disease activity in IBD is indispensable for proper treatment management. Disease activity can be evaluated using clinical, endoscopic, and histological criteria. Substantial evidence shows that targeting endoscopic remission in IBD therapy is superior to tailing only clinical remission (concerning relapse rates, hospitalization rates, and the need for surgery).[1] Moreover, histological healing in UC was shown to be a better predictor for reduced relapses, hospitalizations, colectomy rates, and colorectal risk than endoscopic/macroscopic healing alone.[2] Data are more scarce concerning histology in CD, although recent work showed that histologic healing was associated with decreased risk of clinical relapse, medication escalation, or corticosteroid use.[3]

Colonoscopy with biopsy is the "gold" standard for evaluating disease activity in IBD.[4] However, colonoscopy is expensive and harbors the risk for complications due to its invasiveness.[5] The acquisition of biopsies exposes patients to additional risks of bleeding and perforation, and it delays the diagnosis because of lengthy tissue processing.[6]

Corresponding author: **Catalina Mihai**, *e-mail:* **catalinamihai@yahoo.com**
Received: **May 29, 2020** *Accepted:* **August 18, 2020** *Available Online Date: X XX 2021*

DOI: **10.5152/tjg.2021.20420**

Current research is geared toward finding non-invasive, inexpensive, and accessible methods that estimate disease activity in order to overcome the limitations and risks associated with colonoscopy and biopsy sampling. Modern solutions based on artificial intelligence (AI)/machine learning (ML) algorithms are new and powerful tools currently being proposed for various medical topics.[7] Recently, an AI/ML-based deep neural network (NN) was described for evaluating endoscopic and histologic disease activity in UC using only endoscopic images.[8] Another AI/ML solution was proposed for predicting persistent intestinal histologic inflammation using endocytoscopy images.[9]

To our knowledge, no AI/ML solution has yet been described for predicting histologic activity in IBD using only standard clinical and biological variables.

We aimed to develop a NN model for non-invasive prediction of histologic activity in IBD using inexpensive, routinely available clinical and biological parameters and assess the model's generalizability by testing it on unseen data, anticipating the future utility in clinical setting and decision-making.

MATERIALS AND METHODS

Study Design and Participants

An observational retrospective single-center cohort study was conducted on a sample of 486 (371 UC, 115 CD) patient records. All patients were admitted to the hospital between March 2011 and November 2019.

MAIN POINTS

- *Colonoscopy with biopsy is the "gold" standard for evaluating disease activity in inflammatory bowel diseases but is expensive and harbors the risk for complications due to its invasiveness.*
- *Our study is the first to propose a non-invasive prediction of histologic activity in inflammatory bowel diseases based on Artificial Intelligence methods using standard clinical and biological parameters.*
- *Two multilayered perceptron models are built to predict histologic activity in ulcerative colitis, and 2 other models are proposed for the prediction of histologic activity in Crohn's disease.*
- *Our models achieved similar or better performance than the documented performance of fecal calprotectin (the best non-invasive IBD biomarker to date).*
- *We anticipate the future utility in the clinical decision-making of a non-invasive disease activity prediction, thus reducing the risks and drawbacks of invasive procedures.*

Pre-diagnosed and newly diagnosed confirmed UC and CD patients who underwent a colonoscopy with biopsy for disease assessment were included. Patients were excluded if they were in evidence with concurrent disorders (infections, autoimmune and inflammatory conditions, cirrhosis, neoplasia, and hemodialysis) capable of influencing medical parameters.

All patients provided written informed consent. The study has full ethical approval from local/regional institutions. No sex-based or racial/ethnic-based differences were present.

Clinical Protocol

Pre-diagnosed UC or CD patients were admitted for treatment monitoring or disease worsening. Newly UC or CD diagnosed cases were admitted for the typical or atypical onset of digestive symptoms, including rectal bleeding, diarrhea, abdominal pain, urgency, and incontinence. According to the European consensus guidelines, UC and CD diagnosis is established by clinical, biochemical, stool, endoscopic, cross-sectional imaging, and histological investigations.[4] Patients underwent a medical history interview, physical examination, routine laboratory tests, and colonoscopy with biopsy to diagnose or assess already diagnosed UC or CD disease. Patients were investigated following the European standard protocols. Only patients with a confirmed diagnosis of UC or CD were included.

Data Collection

Documented clinical parameters were age, gender, smoking status, number of stools/day and presence of diarrhea, tenesmus, lower gastrointestinal bleeding (LGB), abdominal pain, weight loss, asthenia, and pallor. Smoking status was represented as a categorical variable with 3 possible values: 0: smoker, 1: nonsmoker, and 2: former smoker. The number of stools/day was recorded as a continuous variable. The presence of diarrhea, tenesmus, LGB, abdominal pain, weight loss, asthenia, and pallor was represented as binary categorical variables (1: indicating presence and 0: the absence of referred symptom).

Laboratory parameters documented were red blood cells, white blood cells (WBC), platelets (PLT), hemoglobin (HGB), hematocrit (HCT), mean corpuscular hemoglobin concentration (MCHC), plateletcrit (PCT), platelet distribution width (PDW), mean platelet volume (MPV), platelet large cell ratio (PLCR), neutrophils (NEUT), lymphocytes, monocytes (MONO), C reactive protein (CRP),

erythrocyte sedimentation rate/h (ESR), fibrinogen, serum iron (SI), ferritin, total proteins (TP), albumin, alpha 1 globulins (A1G), alpha 2 globulins (A2G), beta 1 globulins, beta 2 globulins, gamma globulins, glucose.

Colonoscopy with biopsy was performed on an EVIS EXERA II endoscopy system (Olympus America). According to current recommendations, the overall assessment of histologic disease severity was based on the most affected biopsy sample.[10] To date, there is no validation or standardization of the histological definition of remission in IBD.[2] In our paper, histological healing was defined as the absence of erosions/ulcerations and neutrophils both in the crypts and lamina propria, according to 1 of the latest proposed definitions cited in a recent systematic review published by the American College of Gastroenterology.[2,11] A patient was considered to have histologic remission if no neutrophils were detected in the worst affected biopsy specimen. Similarly, the active histologic disease was considered when erosions/ulcerations were present, or neutrophils were detected in the biopsy sample either in the crypts or in lamina propria. Each bioptic sample in our study was analyzed independently by 2 experienced pathologists, with good expertise in IBD diagnosis. The lack of consensus was resolved by a third independent experienced pathologist. Subsequent colonoscopies with biopsy conducted in an interval higher than 1 month were recorded separately.

Preprocessing and Management of Missing Values

Documented continuous variables (biological parameters and number of stools/day) were standardized in the range [0,1]. Values of HGB, HCT, SI, and ferritin were processed to resolve the differences between the sexes.

Missing values were assigned using multivariate imputation by chained equations method implemented by the MICE package in R Studio Version 1.2.1335 © 2009-2019 RStudio, Inc. Build 1379 (f1ac3452). Missing continuous variables were assigned by applying the Bayesian regression built-in method, while categorical data were imputed using the logistic regression built-in method.

Standard Statistics for Feature Selection

For UC patients, ANOVA with Holm adjustments in R Studio was used to determine continuous variables for which significant differences between the active histology group and remission group existed. Statistical significance was considered for $P < .0001$. If any 2 of the selected continuous variables had high intercorrelation with a Pearson coefficient ≥ 0.9, 1 of them was removed. Feature selection in CD was based on domain knowledge.

Chi-square test of independence was performed both for UC and CD to examine whether there is a relationship between each categorical parameter and the histologic activity class. The categorical parameter with the best chi-square test statistic was selected for each of the 2 diseases.

Four Neural Network Models—Construction and Evaluation

Initial data of 341 UC patient records were randomly divided into a training set of 273 records (80%) and a test set of 68 records (20%) such that variables distributions in each set were similar to those in the original dataset. The other 30 patient records from the same medical center were added independently to be used as a validation set. Similarly, initial data of 103 CD patient records were randomly divided into a training set of 82 records (80%), and a test set of 21 records (20%) and other 12 independent patient records from the same center constituted the validation set. Histologic activity classes (active/inactive) were not equally represented in the train and test sets. However, both validation sets had a more balanced distribution of histologic activity classes.

Four multilayered perceptron classifiers were developed based on the training sets. Classifiers were constructed using the caret::train function in R Studio. Ten-fold cross-validation was used to reduce the problem of overfitting. Synthetic minority over-sampling technique (SMOTE) was used with caret::train function to overcome the issue of imbalanced data.

Two binary classifiers were built for UC predictions. UC classifiers were used to predict whether a UC patient has histologic activity or remission based on all 14 parameters chosen by the feature selection method (first UC classifier) or based only on biological parameters (second classifier).

The other 2 binary classifiers were built for CD predictions. CD classifiers were used to predict whether a CD patient has active or inactive histologic disease based on 2 clinical and 4 biological parameters (first CD classifier) or based on 6 biological parameters (second CD classifier).

Classifiers trained only on biological data were built as it is of interest to construct and evaluate a model based only on objective data.

Developed NNs were evaluated on the test set and validation set on accuracy (ACC) of classification. Where applicable, the area under the receiver operating characteristic curve (AUC), sensitivity (SE), specificity (SP), and positive and negative predictive values (PPV and NPV) were also determined.

RESULTS

Patient Characteristics

Of all 371 UC patient records, 247 (67%) were males and 124 (33%) females. Among the 115 CD patient records, 64 (56%) were males and 51 (44%) females. The age range of the participants was 18-82. The distribution of active/inactive histology classes was imbalanced: 75 UC records were classified with histologic remission, 296 with UC active disease, and 33 CD patient records had inactive histology while 82 had CD active disease.

Selected clinical characteristics and laboratory findings for all patient records, and each activity class are summarized in Table 1 (UC) and Table 2 (CD).

Handled Missing Values

UC patient records had a total of 350 (6.74%) missing values were imputed using MICE package as follows LGB: 1, MONO: 1, MPV: 5, ESR: 51, fibrinogen: 87, SI: 25, CRP: 30, A1G: 75, A2G: 75. For CD records, a total of 62 (7.28%) missing values were imputed as follows PLT: 1, ESR: 13, fibrinogen: 31, SI: 12, CRP: 10.

Feature Selection

Using ANOVA with Holm adjustment, feature selection for UC initially included 17 continuous variables with a significant difference between active histology group and remission group: number of stools/day, WBC, HGB, HCT, MCHC, PLT, MONO, NEUT, MPV, PLCR, PCT, ESR, fibrinogen, SI, CRP, A1G, and A2G. Significance was established at $P < .0001$. Next, highly intercorrelated features were identified and removed. Four strong correlations with a

Table 1. Clinical and Biological Parameters for All UC Patient Records and Each Histologic Activity Class

		Histology	
	All	Remission	Activity
Number of records	371	75	296
Gender (male:female)	247:124	46:29	201:95
Age (years)	44.8 ± 14	43.9 ± 12.3	45 ± 14.4
Number of stools/day	5.1 ± 3.9	1.4 ± 0.8	6 ± 3.8
LGB	250 (67.4%)	1 (1.3%)	249 (84.1%)
WBC, $\times 10^3/\mu L$	8.2 ± 3.3	6.5 ± 1.5	8.6 ± 3.4
HGB, g/dL	13.6 ± 2	14.6 ± 1.2	13.4 ± 2
MCHC, g/dL	33 ± 1.5	33.6 ± 1.3	32.9 ± 1.5
PLT, $\times 10^3/\mu L$	313.6 ± 111.7	259.5 ± 78.1	327.3 ± 114.9
MONO, $\times 10^3/\mu L$	0.68 ± 0.3	0.5 ± 0.2	0.7 ± 0.4
MPV, fL	10.3 ± 1	10.8 ± 0.8	10.1 ± 1
ESR, mm/h	15.9 ± 19.8	4.6 ± 3.2	18.9 ± 21.2
Fibrinogen, mg/dL	389.4 ± 83.5	336 ± 71.7	405.5 ± 80.2
CRP, mg/dL	1.6 ± 3.4	0.2 ± 0.2	2 ± 3.8
SI, µg/dL	72.5 ± 42	97.3 ± 36.6	66 ± 40.9
A1G, %	2.8 ± 1.2	2 ± 0.3	3 ± 1.3
A2G, %	11.3 ± 2.5	10.2 ± 1.6	11.6 ± 2.6

LGB, lower gastrointestinal bleeding; WBC, white blood cells; HGB, hemoglobin; MCHC, mean corpuscular hemoglobin concentration; PLT, platelets; MONO, monocytes; MPV, mean platelet volume; ESR, erythrocyte sedimentation rate/h; CRP, C reactive protein; SI, serum iron; A1G, alpha 1 globulins (A1G); A2G, alpha 2 globulins.

Table 2. Clinical and Biological Parameters for All CD Patient Records and Each Histologic Activity Class

		Histology	
	All	Remission	Activity
Number of records	115	33	82
Gender (male:female)	64:51	20:13	44:38
Age (years)	38.4 ± 11.2	41.8 ± 11.2	37 ± 10.9
Number of stools/day	3.7 ± 3	1.8 ± 1.1	4.5 ± 3.1
Diarrhea	59 (51.3%)	3 (9.1%)	56 (68.3%)
WBC, $\times 10^3/\mu L$	8.9 ± 3.7	7.1 ± 1.8	9.7 ± 4.1
PLT, $\times 10^3/\mu L$	353 ± 145.4	268.8 ± 46.2	385.9 ± 157.4
ESR, mm/h	16.4 ± 18.1	5.1 ± 4.4	20.4 ± 19.5
Fibrinogen, mg/dL	408.3 ± 91.7	351 ± 50	432.6 ± 94.8
CRP, mg/dL	3.1 ± 5.4	0.3 ± 0.2	4.3 ± 6.1
SI, µg/dl	67 ± 45	108.4 ± 42.8	51.6 ± 35.2

WB, white blood cells; PLT, platelets; ESR, erythrocyte sedimentation rate/h; CRP, C reactive protein; SI, serum iron.

Pearson coefficient ≥0.9 were identified between WBC and NEUT, HGB and HCT, PLT and PCT, MPV, and PLCR. Thus, the following 4 parameters were removed from the analysis: NEUT, HCT, PCT, and PLCR.

Feature selection in CD was based on domain knowledge and included 7 continuous variables: number of stools/day, WBC, CRP, fibrinogen, ESR, PLT, and SI. Considering that the number of CD patient records in our database is much smaller than in the UC dataset, features selected for CD also have to be fewer.

All categorical parameters were successively fitted with the histology activity variable using the chi-square test, both in UC and CD. The categorical parameter with the best chi-square test statistic was chosen to be included in the analysis for each disease. LGB was the selected categorical feature in UC (X^2 (1, N = 371) = 184.54, $P < .001$) and diarrhea in CD (X^2 (1, N = 115) = 30.684, $P < .001$).

As a result of the feature selection stage, 14 parameters in UC (number of stools/day, LGB, WBC, HGB, MCHC, PLT, MONO, MPV, ESR, fibrinogen, SI, CRP, A1G, A2G) and 8 parameters in CD (number of stools/day, diarrhea, WBC, CRP, fibrinogen, ESR, PLT, SI) were included in further analysis.

Results of the NN Models Construction and Evaluation

Based on the results produced by the feature selection methods, 2 NN models were trained for UC predictions and 2 for CD predictions using the selected parameters as inputs.

UC Classifiers

To build the classifiers, the initial dataset of 341 UC patient records was randomly divided into a training set (273 records) and a test set (68 records). Thirty patient records were added independently to constitute the validation set. Unlike in the training and test set, histology activity classes were more balanced in the validation set.

The first NN model was developed using all 14 variables to predict whether a UC patient has histologic activity or remission. Model performance metrics are shown in Table 3.

The second NN model was built to predict the same binary outcome (histologic remission or activity) as the first classifier using only the 12 biological input parameters in order to investigate a model that uses only objective data. The second model performance metrics are indicated in Table 4.

ROC curves proving models' performance on the train, test, and validation sets are shown in Figure 1 for the first UC classifier and in Figure 2 for the second UC classifier.

CD Classifiers

The initial dataset of 103 CD patient records was randomly divided into a training set (82 records) and a test

Table 3. The First UC Histology Classifier Performance Metrics

	Train Set		Test Set		Validation Set	
	Predictions		Predictions		Predictions	
Actual	Remission	Activity	Remission	Activity	Remission	Activity
Remission	47	13	14	2	12	0
Activity	0	213	1	51	1	17
ACC	95.24%		95.59%		96.67%	
95% CI	(0.9199, 0.9744)		(0.8764, 0.9908)		(0.8278, 0.9992)	
P value	<.001		<.001		<.001	
SE	100%		93.33%		92.31%	
SP	94.25%		96.23%		100%	
PPV	78.33%		87.5%		100%	
NPV	100%		98.08%		94.44%	
AUC	0.9719		0.9786		0.991	

ACC, accuracy; SE, sensitivity; SP, specificity; PPV, positive predictive values; NPV, negative predictive values; AUC, area under the receiver operating characteristic curve.

set (21 records). Twelve patient records were added independently to constitute the validation set.

The other 2 binary classifiers were built for CD predictions. CD classifiers were used to predict whether a CD patient has active or inactive histologic disease based on 2 clinical and 4 biological parameters (first CD classifier) or based on 6 biological parameters (second CD classifier).

The first CD binary model was developed using 6 variables as inputs (number of stools/day, diarrhea, CRP, fibrinogen, ESR, PLT) to predict the histologic activity. Model renderings on each of the 3 datasets are shown in Table 5.

The second CD binary model was built to predict histologic activity using only 6 biological input parameters (WBC, CRP, fibrinogen, ESR, PLT, SI) in order to test a model based only on objective data. The second CD model performance metrics are indicated in Table 6.

Table 4. Second UC Histology Classifier Performance Metrics

	Train Set		Test Set		Validation Set	
	Predictions		Predictions		Predictions	
Actual	Remission	Activity	Remission	Activity	Remission	Activity
Remission	45	20	11	4	11	2
Activity	2	206	4	49	2	15
ACC	91.94%		88.24%		86.67%	
95% CI	(0.8805, 0.9488)		(0.7813, 0.9478)		(0.6928, 0.9624)	
P value	<.001		<.001		<.001	
SE	95.74%		73.33%		84.62%	
SP	91.15%		92.45%		88.24%	
PPV	69.23%		73.33%		84.62%	
NPV	99.04%		92.45%		88.24%	
AUC	0.9508		0.8629		0.9231	

ACC, accuracy; SE, sensitivity; SP, specificity; PPV, positive predictive values; NPV, negative predictive values; AUC, area under the receiver operating characteristic curve.

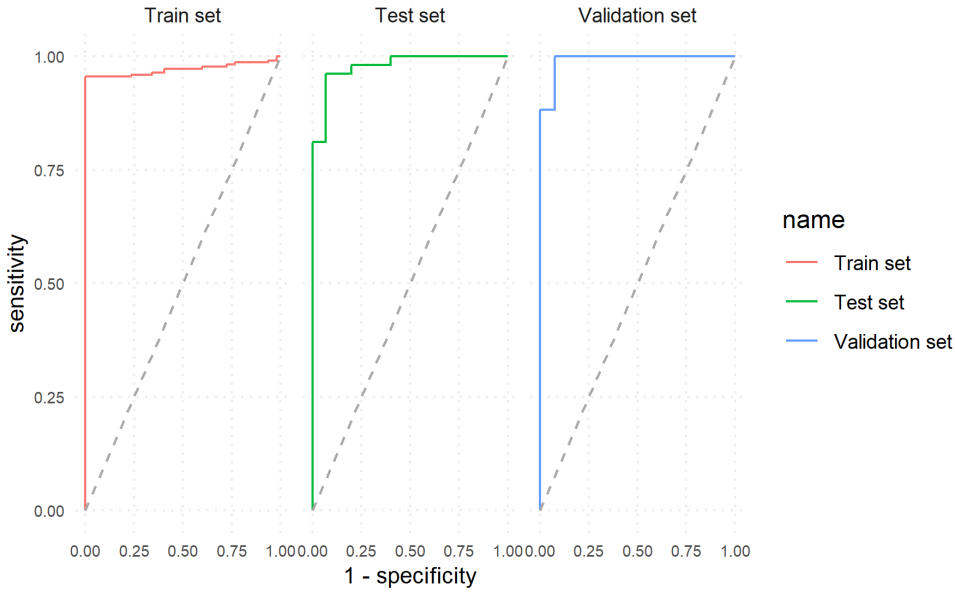

Figure 1. Performance of the first UC classifier to predict histologic remission versus relapse.

ROC curves rendering models' performance on the train, test, and validation sets are shown in Figure 3 for the first CD classifier and in Figure 4 for the second CD classifier.

DISCUSSION

In this paper, the first NN for predicting histologic disease activity in IBD based on routinely available clinical and biological variables was proposed. We showed that using only standard, non-costly, non-invasive parameters, it is possible to differentiate active histologic disease in IBD from inactive status.

Monitoring disease activity in IBD includes clinical, biological, endoscopic, and histologic evaluations and is crucial for disease control, treatment management, and prognosis. Clinical assessment is usually inaccurate if considered alone. Endoscopic and histologic examinations are the

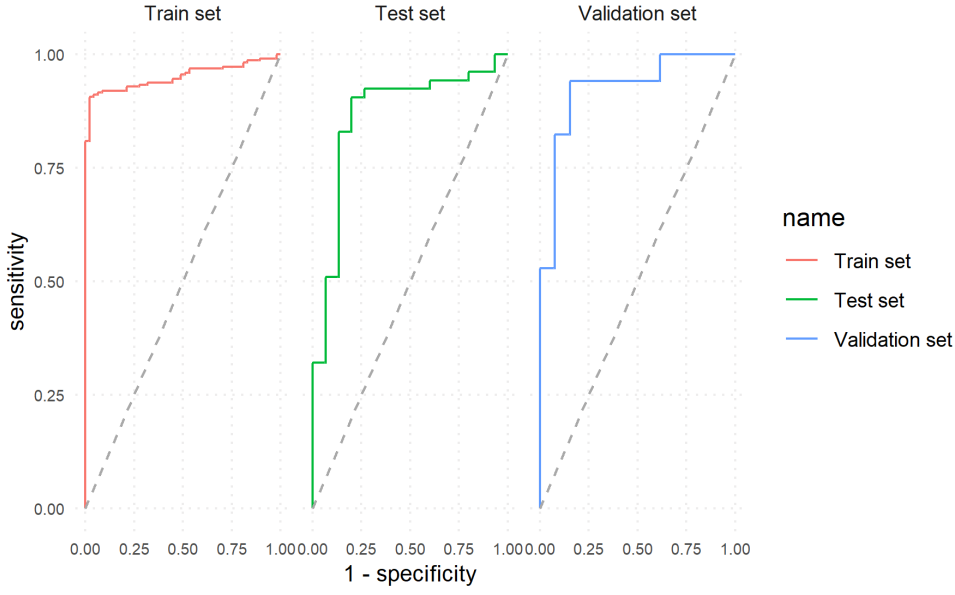

Figure 2. Performance of the second UC classifier to predict histologic remission versus relapse.

Table 5. The First CD Histology Classifier Performance Metrics

	Train Set		Test Set		Validation Set	
	Predictions		Predictions		Predictions	
Actual	Remission	Activity	Remission	Activity	Remission	Activity
Remission	20	7	5	1	5	0
Activity	1	54	1	14	1	6
ACC	90.24%		90.48%		91.67%	
95% CI	(0.8168, 0.9569)		(0.6962, 0.9883)		(0.6152, 0.9979)	
P value	<.001		<.001		<.001	
SE	95.24%		83.33%		83.33%	
SP	88.52%		93.33%		100%	
PPV	74.07%		83.33%		100%	
NPV	98.18%		93.33%		85.71%	
AUC	0.9305		0.9444		0.8889	

ACC, accuracy; SE, sensitivity; SP, specificity; PPV, positive predictive values; NPV, negative predictive values; AUC, area under the receiver operating characteristic curve.

gold standard but are uneasy for the patient and harbor possible complications.

The evaluation of histological disease activity is especially critical because persistent microscopical inflammation in quiescent endoscopic disease increases the risk of disease relapse.[12]

The advent of AI/ML cost-effective techniques in healthcare represents a significant step forward in precision medicine. ML models have been proposed in several clinical IBD studies for differential diagnosis, monitoring, prediction of outcomes, and treatment responses.[13,14] Notably, 3 papers described methods for the assessment of histologic activity in IBD. A deep NN was proposed for evaluating endoscopic and histologic disease activity in UC based on endoscopic images only.[8] Another AI/ML solution uses endocytoscopy images to predict intestinal histologic inflammation.[9] Deep learning has also been used to determine histologic inflammation from endoscopic images by

Table 6. The Second CD Histology Classifier Performance Metrics

	Train Set		Test Set		Validation Set	
	Predictions		Predictions		Predictions	
Actual	Remission	Activity	Remission	Activity	Remission	Activity
Remission	20	8	4	1	5	0
Activity	1	53	2	14	1	6
ACC	89.02%		85.71%		91.67%	
95% CI	(0.8018, 0.9486)		(0.6366, 0.9695)		(0.6152, 0.9979)	
P value	<.001		<.001		<.001	
SE	95.24%		66.67%		83.33%	
SP	86.89%		93.33%		100%	
PPV	71.43%		80%		100%	
NPV	98.15%		87.5%		85.71%	
AUC	0.9641		0.6889		1	

ACC, accuracy; SE, sensitivity; SP, specificity; PPV, positive predictive values; NPV, negative predictive values; AUC, area under the receiver operating characteristic curve.

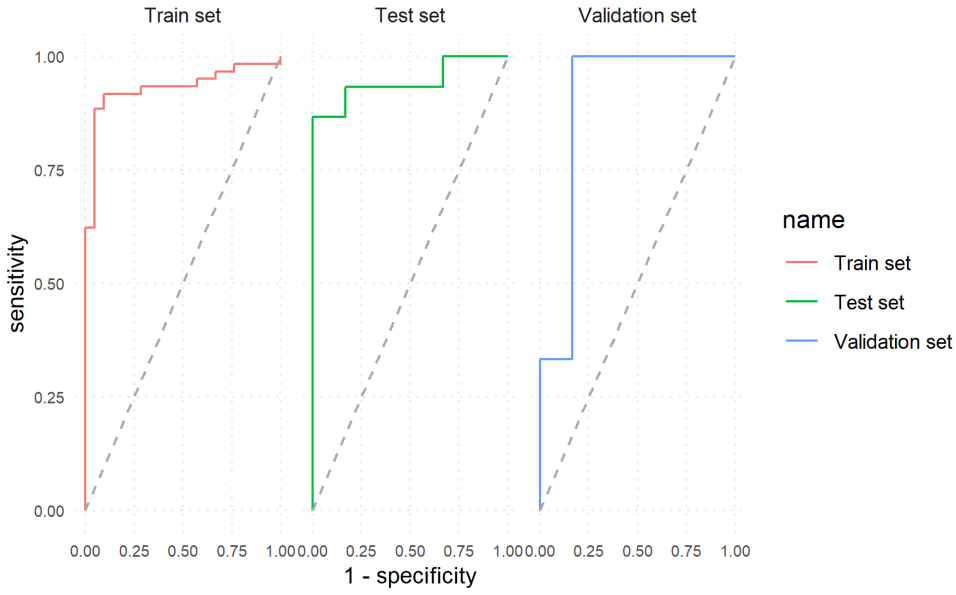

Figure 3. Performance of the first CD classifier to predict active/inactive histologic status.

measuring the redness degree in the red density score in UC.[15] However, all 3 methods necessitate invasive intervention for acquiring endoscopic images.

In the search for non-invasive solutions in IBD monitoring, fecal calprotectin (FC) is, so far, the most encouraging and frequently used fecal biomarker to indicate disease activity in IBD.[16] FC was shown to be a valuable marker for predicting histologic healing and mucosal inflammation in UC[16] and ileocolonic and colonic CD.[17,18] However, FC has not been validated for biomarker-based decision making due to low reliability: significant variability across platforms (which makes it difficult to establish a cut-off value), intra-individual day-to-day variation in FC concentrations, degradation of FC levels at room temperature after stool collection.[19] Additionally, FC is not allowed in Asian and some Western countries[20] and is more expensive than routine *laboratory* investigations.

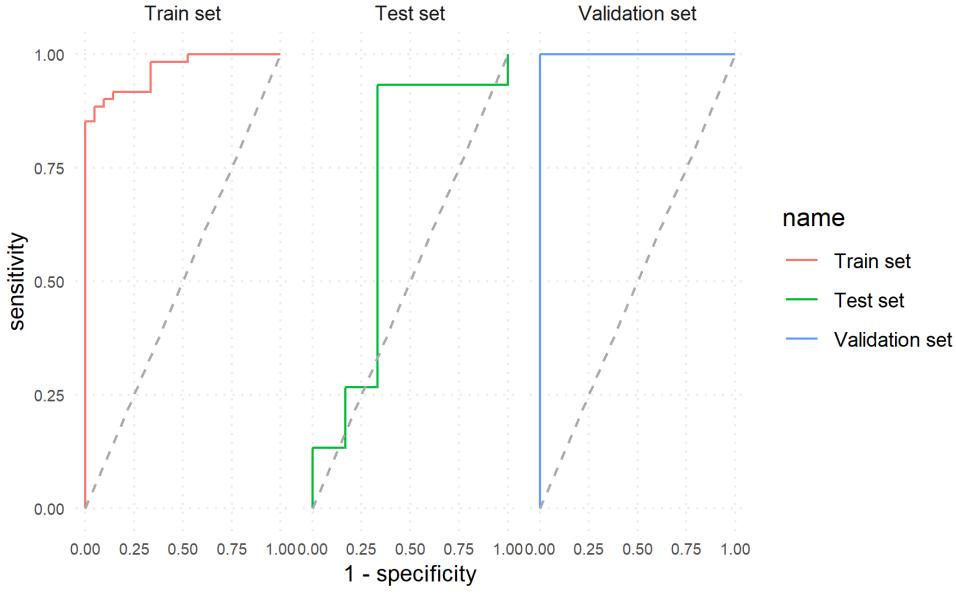

Figure 4. Performance of the second CD classifier to predict active/inactive histologic status.

Thus, we aimed to build a model using only routinely available, non-costly parameters, that could possibly achieve similar or even better performance than FC, given that FC is not readily available in some countries, is slightly more expensive than routine *laboratory* investigations, and has some reliability drawbacks.

Several studies have assessed the performance of FC to estimate histologic activity in IBD, but there is no standardized definition of histologic healing, and this restrains the possibility of comparing performances between different studies. However, there are a few papers that have employed a similar definition to the one used in our research. One of these papers evaluated the performance of FC to predict histologic activity in IBD (both UC and CD) and achieved an AUC of 0.898, an ACC of 85%, SE 77%, SP 100%, PPV 100%, NPV 68% for an FC cut-off level of **200 ng/mL**.[21] Another study assessed FC in estimating histologic remission in colonic IBD and obtained an AUC of 0.95 (SE: 0.03; 95% CI, 0.88-1.02; $P < .001$), SE 100%, SP 77%, PPV 81.2%, NPV 100%, for a FC cut-off level of 100 μg/g.[18] Comparatively, our first UC classifier achieved better performance while the second UC and first CD classifiers had similar behavior as FC in the mentioned studies. This proves the potential of the AI/ML method proposed in this study to be used as a non-invasive and accessible tool in IBD monitoring.

Limitations and Future Perspectives

First, the small size of our dataset, along with the fact that the independent validation set is from the same center, entails rigorous external validation with data from other centers. Notably, the CD dataset is of minimal size and, consequently, fewer features had to be selected for CD models' training. This fact generates a very high risk of bias for CD classifiers. Second, the imbalanced distribution of histologic activity classes predisposes to calculation biases, although the SMOTE function in R was used to reduce these biases significantly. Thirdly, FC was not documented for a direct real-time comparison with the performance of the proposed NN classifiers, due to logistical reasons related to the retrospective nature of our study. Forth, only colonic and ileocolonic activity was assessed, leaving a remaining of up to one-third of the patients with isolated small bowel involvement, due to the reduced accessibility for biopsy acquisition. However, in the current phase of our study, the focus has been on assessing the accuracy and prominence of the new AI/ML-based method.

In the future, these drawbacks could be overcome by employing studies on broader, more diverse, and comprehensive datasets in a center with greater accessibility that would permit organizing a cohort with a balanced distribution of histologic activity classes. The next trials would improve models' performance even further by using different ML algorithms as our patient's database extends and to validate our models through real-time comparison with documented FC levels. Finally, colonic IBD requires dysplasia surveillance and routine biopsies after 8-10 years from disease onset.[22] This task is not tackled in our current models, but future approaches could incorporate more context and data enrichment with specific variables that might be able to confront the burden of dysplasia detection.

Further improving and validating automatic learning methods may lead to more frequent monitoring of subclinical IBD disease activity with significantly fewer invasive procedures, less exposure to inherent risks, and more comfort for the patient.

CONCLUSIONS

In the context of current costly and invasive monitoring methods for IBD, our study proposes a cost-efficient, non-invasive, AI/ML-based tool to predict histologic disease activity in IBD with reasonable accuracy. Our NN model represents a significant upsurge in the non-invasive assessment of IBD microscopical inflammation, leading the way to possible future use in clinical practice. Further validations on external datasets are needed to ensure generalizability and further prospective studies are needed to evaluate whether the proposed classifiers can be used as biomarkers able to replace endoscopies and biopsies.

Ethics Committee Approval: All procedures involving human participants were in accordance with the ethical standards of the local / regional Romanian institutions ("St. Spiridon" Regional Hospital Ethics Committee, no. 54 / 10.2019 and Research Ethics Commission of the "Gr. T. Popa" University of Medicine and Pharmacy, no. 15308 / 07.2019) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

Informed Consent: Informed consent was obtained from all individual participants included in the study.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept – I.V.P., C.C.P., C.M.; Design - I.V.P., A.B.; Supervision - A.B.; Resource - C.C.P., C.M.; Materials - I.V.P., A.B., M.D.; Data Collection and/or Processing - I.V.P.; Analysis and/or Interpretation - I.V.P., A.B.; Writing - I.V.P., A.B.; Critical Reviews - A.B.

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

Financial Disclosure: The authors declared that this study has received no financial support.

REFERENCES

1. Zallot C, Peyrin-Biroulet L. Deep remission in inflammatory bowel disease: looking beyond symptoms. Curr Gastroenterol Rep. 2013;15(3):315. [CrossRef]

2. Bryant RV, Winer S, Travis SP, Riddell RH. Systematic review: histological remission in inflammatory bowel disease. Is 'complete' remission the new treatment paradigm? An IOIBD initiative. J Crohns Colitis. 2014;8(12):1582-1597. [CrossRef]

3. Christensen B, Erlich J, Gibson PR, Turner JR, Hart J, Rubin DT. Histologic healing is more strongly associated with clinical outcomes in ileal Crohn's disease than endoscopic healing. Clin Gastroenterol Hepatol. 2020;18(11):2518.e1-2525.e1. [CrossRef]

4. Maaser C, Sturm A, Vavricka SR, et al. ECCO-ESGAR guideline for diagnostic assessment in IBD part 1: initial diagnosis, monitoring of known IBD, detection of complications. J Crohns Colitis. 2019;13(2):144-164. [CrossRef]

5. Ko CW. Colonoscopy risks: what is known and what are the next steps? Gastroenterology. 2018;154(3):473-475. [CrossRef]

6. Bateman AC, Patel P. Lower gastrointestinal endoscopy: guidance on indications for biopsy. Frontline Gastroenterol. 2014;5(2):96-102. [CrossRef]

7. Chan YK, Chen YF, Pham T, Chang W, Hsieh MY. Artificial intelligence in medical applications. J Healthc Eng. 2018;2018:4827875. [CrossRef]

8. Takenaka K, Ohtsuka K, Fujii T, et al. Development and validation of a deep neural network for accurate evaluation of endoscopic images from patients with ulcerative colitis. Gastroenterology. 2020;158(8):2150-2157. [CrossRef]

9. Maeda Y, Kudo SE, Mori Y, 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(2):408-415. [CrossRef]

10. Pai RK, Jairath V, Vande Casteele N, Rieder F, Parker CE, Lauwers GY. The emerging role of histologic disease activity assessment in ulcerative colitis. Gastrointest Endosc. 2018;88(6):887-898. [CrossRef]

11. Chateau T, Feakins R, Marchal-Bressenot A, Magro F, Danese S, Peyrin-Biroulet L. Histological remission in ulcerative colitis: under the microscope is the cure. Am J Gastroenterol. 2020;115(2):179-189. [CrossRef]

12. Riley SA, Mani V, Goodman MJ, Dutt S, Herd ME. Microscopic activity in ulcerative colitis: what does it mean? Gut. 1991;32(2):174-178. [CrossRef]

13. Li J, Qian JM. Artificial intelligence in inflammatory bowel disease: current status and opportunities. Chin Med J (Engl). 2020;133(7):757-759. [CrossRef]

14. Seyed Tabib NS, Madgwick M, Sudhakar P, Verstockt B, Korcsmaros T, Vermeire S. Big data in IBD: big progress for clinical practice. Gut. 2020;69(8):1520-1532. [CrossRef]

15. Bossuyt P, Nakase H, Vermeire S, et al. Automatic, computer-aided determination of endoscopic and histological inflammation in patients with mild to moderate ulcerative colitis based on red density. Gut. 2020;69(10):1778-1786. [CrossRef]

16. Theede K, Holck S, Ibsen P, Ladelund S, Nordgaard-Lassen I, Nielsen AM. Level of fecal calprotectin correlates with endoscopic and histologic inflammation and identifies patients with mucosal healing in ulcerative colitis. Clin Gastroenterol Hepatol. 2015;13(11):1929.e1-1936.e1. [CrossRef]

17. Sipponen T, Kärkkäinen P, Savilahti E, et al. Correlation of faecal calprotectin and lactoferrin with an endoscopic score for Crohn's disease and histological findings. Aliment Pharmacol Ther. 2008;28(10):1221-1229. [CrossRef]

18. Zittan E, Kelly OB, Kirsch R, et al. Low fecal calprotectin correlates with histological remission and mucosal healing in ulcerative colitis and colonic Crohn's disease. Inflamm Bowel Dis. 2016;22(3):623-630. [CrossRef]

19. Dulai PS, Peyrin-Biroulet L, Danese S, et al. Approaches to integrating biomarkers Into clinical trials and care pathways as targets for the treatment of inflammatory bowel diseases. Gastroenterology. 2019;157(4):1032.e1-1043.e1. [CrossRef]

20. Shiga H, Abe I, Onodera M, et al. Serum C-reactive protein and albumin are useful biomarkers for tight control management of Crohn's disease in Japan. Sci Rep. 2020;1:511.

21. Vieira A, Fang CB, Rolim EG, et al. Inflammatory bowel disease activity assessed by fecal calprotectin and lactoferrin: correlation with laboratory parameters, clinical, endoscopic and histological indexes. BMC Res Notes. 2009;2:221. [CrossRef]

22. Eaden JA, Mayberry JF, British Society for Gastroenterology, Association of Coloproctology for Great Britain and Ireland. Guidelines for screening and surveillance of asymptomatic colorectal cancer in patients with inflammatory bowel disease. Gut. 2002;51(suppl 5):V10-V12. [CrossRef]