

Machine Learning-Based Diagnosis of Inflammatory Bowel Diseases

Hafsa Nur Peker and Gul Didem Batur Sir

Department of Industrial Engineering
Gazi University, Ankara, Türkiye
hnurozturk@gazi.edu.tr, dbatur@gazi.edu.tr

Abstract

Crohn's and ulcerative colitis are two diseases classified under inflammatory bowel diseases that are frequently confused during the diagnostic process. These conditions are chronic diseases that significantly affect patients' quality of life. Since they may present with similar clinical symptoms accurate and timely diagnosis is often difficult. Misdiagnosis may lead to inappropriate treatment and delay disease control, thereby imposing an additional burden on the healthcare system. The diagnosis of inflammatory bowel diseases is currently based on the combined evaluation of medical history, physical examination findings, biochemical tests, radiological assessments, endoscopic procedures, and histopathological examination of biopsy samples. However, this process is both invasive and limited in terms of time and cost. Therefore, there is an increasing need for less invasive and faster diagnostic approaches. The potential to establish a preliminary diagnosis using easily accessible data such as symptom information and blood tests may improve the efficiency of clinical processes. In this study, considering these challenges, machine learning approaches were applied to support the diagnostic process using logistic regression and random forest methods. The results indicate that both methods achieved successful performance in disease classification. However, the random forest method demonstrated better performance with higher accuracy.

Keywords

Inflammatory bowel diseases, Crohn's disease, Ulcerative colitis, Machine learning, Classification

1. Introduction

Crohn's disease and ulcerative colitis are classified under the group of inflammatory bowel diseases (IBD) and are chronic conditions that significantly affect patients' quality of life due to their persistent course and recurrent inflammatory episodes (Baumgart and Cardig, 2007; Baumgart, 2009). In addition, the presence of similar clinical symptoms (such as abdominal pain, diarrhea, weight loss, and fatigue) during the course of the disease and the high similarity in patients' complaints make it quite difficult to distinguish between these two diseases and to establish a clear diagnosis. These challenges in the diagnostic process increase the risk of misdiagnosis or delayed diagnosis, which in turn leads to inappropriate treatment, accelerated disease progression, and additional cost burdens on healthcare systems. In current clinical practice, the diagnosis of IBD is generally established through the combined use of additional procedures such as endoscopic examinations and biopsy analyses following blood tests and symptom evaluation. However, the methods used in practice are invasive, reduce patient comfort, and have significant limitations due to being both time-consuming and costly.

In this context, the development of less invasive, faster, and cost-effective diagnostic approaches has become an important area of research. In particular, the use of easily accessible and non-invasive data sources such as symptom data and blood test results may enable the development of clinical decision support systems. In recent years, machine learning methods have begun to be widely used in the healthcare field, especially in disease classification and diagnostic support systems. These methods can provide higher accuracy compared to traditional statistical approaches due to their ability to learn patterns from large and complex datasets. However, when machine learning applications

in the diagnosis of inflammatory bowel diseases are examined, it is observed that many studies rely on invasive data sources. In addition, the number of studies based on non-invasive and easily accessible data sources such as symptoms and blood tests is limited.

In this study, considering these limitations and the gaps in the literature, a machine learning-based approach is proposed for distinguishing Crohn's disease and ulcerative colitis using non-invasive data sources such as symptom data and blood test results. Within this scope, disease classification was performed using logistic regression and random forest methods, and the performance of these methods was evaluated comparatively. It is planned that the findings obtained will be used as an initial step toward the development of faster, more accessible, and clinically integrable diagnostic support systems.

1.1 Objectives

The primary objective of this study is to investigate the applicability of a machine learning-based classification approach for distinguishing Crohn's disease and ulcerative colitis using only non-invasive and easily accessible clinical data, without the use of invasive diagnostic methods.

In line with this main objective, the specific aims of the study are presented as follows:

1. To apply machine learning models for the classification of Crohn's disease and ulcerative colitis using symptom data and routine blood test parameters,
2. To perform disease classification using logistic regression and random forest algorithms and to compare the performance of these methods using standard evaluation metrics.

2. Literature Review

Artificial Intelligence (AI) is widely used in diagnostic processes, as in many stages of diseases. In the diagnosis of IBD, AI applications can be examined under different categories, including the analysis of imaging data, the evaluation of non-imaging clinical and laboratory data, histopathological examinations, molecular and proteomic analyses, as well as microbiota and metabolomic analyses. Imaging is one of the areas where AI is most intensively used in the diagnosis of IBD. Convolutional neural networks and other deep learning methods are widely utilized in the analysis of endoscopic, capsule, and radiological images. In one study, it was shown that AI-assisted systems achieved sensitivity and specificity rates above ninety percent in determining endoscopic scoring standards for ulcerative colitis (Sinonquel et al., 2023). Similarly, real-time AI models have been reported to significantly reduce interobserver variability and improve disease activity assessments (Ahmad et al., 2023). A system developed in the field of capsule endoscopy was reported to significantly shorten image review times while providing a high level of specificity (Brodersen et al., 2023). In a study conducted on radiological images, it was demonstrated that deep learning algorithms could distinguish Crohn's disease and ulcerative colitis with high accuracy (Jiang et al., 2022). In addition, a model developed using endoscopic images was reported to determine endoscopic and histological remission states with high accuracy and, in some cases, reduce the need for biopsy (Takenaka et al., 2020).

Non-imaging approaches also constitute an important area of research in the diagnosis of IBD. The fact that invasive methods such as endoscopy and colonoscopy are costly and time-consuming increases the need for alternative diagnostic methods. Furthermore, delays in the diagnostic process negatively affect disease prognosis, making studies in this area even more critical. In this context, a study demonstrated that a machine learning model developed using data obtained from routinely performed blood, urine, and stool tests could classify Crohn's disease and ulcerative colitis with high accuracy. Results obtained using the random forest algorithm indicate that models developed with non-invasive data can effectively support the diagnostic process. (Kraszewski et al., 2021).

Histopathological analyses hold an important place in the diagnosis of IBD; however, the evaluation process is expert-dependent and time-consuming. Therefore, AI-based approaches provide significant advantages in terms of the automation and standardization of histological analyses. In one study, a model developed to evaluate ulcerative colitis biopsies achieved high levels of sensitivity and specificity and was also able to predict the risk of disease flare-ups (Iacucci et al., 2023). Similarly, another model developed using large datasets was reported to classify histological activity with satisfactory accuracy (Hamamoto et al., 2023).

Molecular and proteomic analyses provide important insights into the early diagnosis of inflammatory bowel diseases and the understanding of disease mechanisms. In this field, a study reported the identification of fifty-one protein biomarkers capable of predicting Crohn's disease approximately five years before clinical diagnosis, achieving a high area under the curve (AUC) value (Torres et al., 2020). In another study based on gene expression profiling, Crohn's disease and ulcerative colitis were distinguished with high accuracy (Burczynski et al., 2006). Additionally, a study based on serum protein analysis reported that these two diseases could be differentiated using surface-enhanced laser desorption ionization mass spectrometry (Meuwis et al., 2007).

The gut microbiota plays a critical role in the pathogenesis of IBD, and AI-based analyses in this area are increasing. In one study, a random forest model developed using microbiota data was found to be highly effective in distinguishing Crohn's disease and ulcerative colitis patients from healthy individuals. Furthermore, a study utilizing metabolomic approaches reported that disease-specific metabolic profiles for Crohn's disease and ulcerative colitis could be identified using stool samples (Marchesi et al., 2007).

When the existing literature is examined, it is evident that AI applications in the diagnosis of IBD have made significant progress. However, it is noteworthy that a large proportion of studies rely on imaging data or histopathological and molecular data obtained through invasive methods. Although studies on non-imaging approaches exist, their number is limited and they often rely on the combined use of different data types. It is considered that analyses based on easily accessible and non-invasive data will fill an important gap in terms of accelerating diagnostic processes, improving patient satisfaction, and reducing costs.

3. Materials and Methods

This study generated a synthetic dataset for the multiclass classification of inflammatory bowel disease. The dataset consists of three classes: non-inflammatory bowel disease individuals, individuals with Crohn's disease, and individuals with ulcerative colitis. A total of 1000 samples were generated for each class, resulting in a balanced dataset comprising 3000 observations. The variables included in the dataset were selected to represent inflammatory processes, hematological status, and clinical symptoms. Continuous variables included C-reactive protein, erythrocyte sedimentation rate, white blood cell count, hemoglobin, platelet count, albumin, and fecal calprotectin. These variables were generated from normal distributions with different means and standard deviations to reflect the clinical characteristics of each class. In the inflammatory bowel disease groups, inflammatory markers were defined to be higher, while hemoglobin and albumin levels were defined to be lower.

Prior to modeling, the dataset was divided into independent variables and the target variable. Subsequently, the data were split into training and test sets, with 80 percent allocated for training and 20 percent for testing. Stratified sampling was applied to preserve the class distribution. Due to the scale sensitivity of the logistic regression model, standardization was applied to continuous variables. In this process, the mean and standard deviation were learned from the training dataset, and the same transformation was applied to the test dataset. The random forest algorithm was trained on the raw data, as it is not sensitive to feature scaling. Symptomatic variables, including diarrhea, abdominal pain, weight loss, blood in stool, and fatigue, were generated using a Bernoulli distribution with different probabilities assigned for each class. For example, the probability of blood in stool was defined to be higher in the ulcerative colitis group, whereas abdominal pain and weight loss were modeled with higher probabilities in the Crohn's disease group. The age variable was generated using a uniform distribution between 18 and 70. This parametric approach enabled the creation of clinically meaningful differences between classes, while also allowing partial overlap in variable distributions, thereby resulting in a more realistic classification problem.

In this study, two machine learning algorithms were used for the multiclass classification problem. The logistic regression model was trained using the multinomial approach for the multiclass structure. The maximum number of iterations was set to 1000 to prevent convergence issues in the optimization process. Logistic regression was preferred due to its probabilistic framework based on linear decision boundaries and its interpretability. The random forest model was constructed using an ensemble learning approach based on decision trees. A total of 200 trees were used in the model, and randomness was controlled using a fixed seed value. This model was included due to its ability to capture non-linear relationships and its strong generalization performance. Model performance was evaluated on the test dataset using accuracy, macro-averaged F1 score, precision, and recall metrics. Due to the multiclass structure, all metrics were computed using the macro averaging approach. In addition, specificity values for each class were

calculated based on the confusion matrix, and the average specificity was reported. In the calculation of specificity, true negative and false positive values were considered separately for each class.

4. Results and Discussion

4.1 Numerical Results

The numerical performance of the models on the test dataset is presented in Table 1. Both models demonstrated high classification performance across all evaluation metrics. Logistic regression achieved an accuracy of 0.94, with macro-averaged F1 score, precision, and recall values all approximately equal to 0.94. The random forest model slightly outperformed logistic regression, achieving an accuracy of 0.9433 and a macro F1 score of 0.9433. Similarly, precision and recall values for the random forest model were consistently higher, although the difference between the two models remained marginal. Specificity values were also high for both models, with logistic regression achieving 0.97 and random forest reaching 0.9717. This indicates that both models are highly effective in correctly identifying negative instances across all classes.

Table 1. Test Results

Model	Accuracy	F1-Score	Precision	Recall	Specificity
Logistic Regression	0.940000	0.939998	0.940027	0.940000	0.970000
Random Forest	0.943333	0.943332	0.943361	0.943333	0.971667

Cross-validation results (Table 2.) further support the robustness of both models. The mean macro F1 score for logistic regression was calculated as 0.9358, while the random forest model achieved a slightly lower mean of 0.9332. Although random forest performed better on the test dataset, logistic regression demonstrated slightly more stable performance across different folds. This suggests that logistic regression may have a marginal advantage in terms of generalization consistency. Overall, both models achieved strong and comparable performance, indicating that the generated feature space is highly informative for distinguishing between non-inflammatory bowel disease, Crohn's disease, and ulcerative colitis.

Table 2. Cross Validation

Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean F1
Logistic Regression	0.9437	0.9270	0.9478	0.9333	0.9270	0.9358
Random Forest	0.9375	0.9312	0.9479	0.9163	0.9333	0.9332

4.2 Graphical Results

Figure 1 presents a comparative visualization of the confusion matrices for the Logistic Regression and Random Forest models within the scope of a three-class classification problem. In this study, class 0 represents healthy individuals, class 1 represents Crohn's disease, and class 2 represents ulcerative colitis. Upon examination of the findings, it is observed that both models demonstrate perfect performance for class 0 (healthy individuals), correctly classifying all observations belonging to this class. In contrast, misclassifications are predominantly concentrated between class 1 (Crohn's disease) and class 2 (ulcerative colitis). In the Logistic Regression model, 17 observations belonging to Crohn's disease were misclassified as ulcerative colitis, while 19 observations belonging to ulcerative colitis were misclassified as Crohn's disease. A similar error pattern is observed in the Random Forest model; 16 observations from Crohn's disease were misclassified as ulcerative colitis, whereas 18 observations from ulcerative colitis were misclassified as Crohn's disease. This reciprocal misclassification pattern indicates that the two diseases may exhibit similar clinical and biochemical characteristics, making their distinction more challenging. In conclusion, these findings highlight that model evaluation should not rely solely on overall accuracy, but should also consider class-based performance and error patterns.

Figure 2 shows that the AUC value for class 0 (healthy individual) is 1.00. This indicates that the model can perfectly distinguish healthy individuals from diseased groups and has an excellent discriminative ability for this class. The ROC curve being very close to the upper left corner further supports this finding. For class 1 (Crohn's disease), the AUC value is calculated as 0.98. This suggests that the model can distinguish Crohn's disease with high accuracy, although there is some level of confusion with class 2. Similarly, for class 2 (ulcerative colitis), the AUC value is 0.99, indicating a very strong discriminative performance for this class as well. Overall, the ROC curves for all classes being

located near the upper left corner demonstrate that the model achieves a high true positive rate while maintaining a low false positive rate. However, the slightly lower AUC values for class 1 and class 2 compared to class 0 confirm that distinguishing between these two disease groups is relatively more challenging due to their similar characteristics. In conclusion, the ROC analysis reveals that the model has a very high overall discriminative ability, while the distinction between Crohn's disease and ulcerative colitis remains comparatively more complex than for the other classes.

Figure 3 presents the importance levels of the variables used in the model. Examination of the findings clearly indicates that calprotectin is the most influential variable affecting the model's decisions. Calprotectin is followed by CRP and albumin, respectively. The substantially higher importance values of these three variables compared to the others suggest that they serve as the primary determinants in the model's decision-making mechanism.

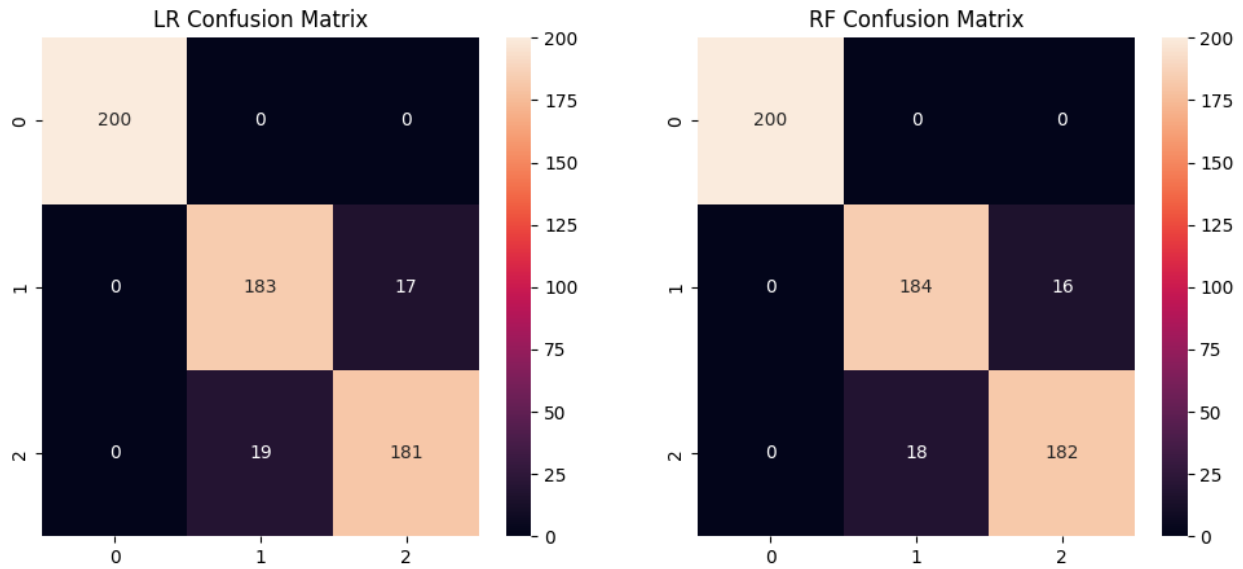


Figure 1. Confusion Matrix of Logistic Regression and Random Forest

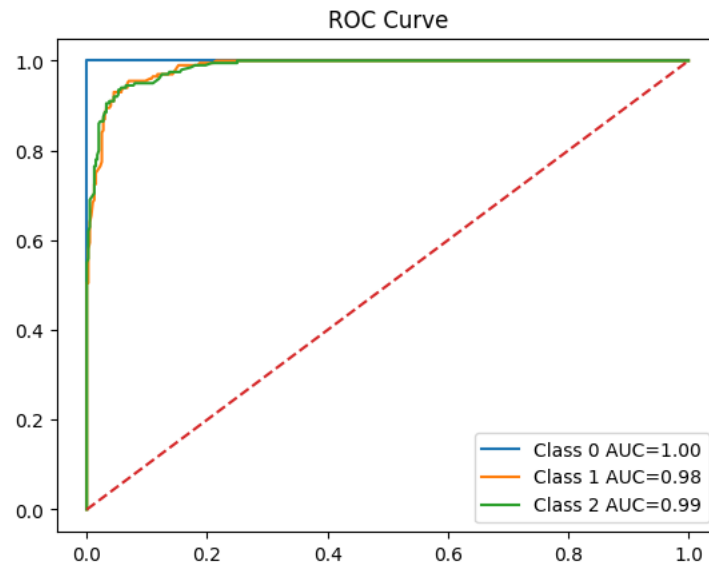


Figure 2. ROC Curve Analysis

Figure 4 presents the correlation relationships among the variables used in the model. Examination of the correlation matrix indicates moderate positive correlations among CRP, ESR, and WBC, suggesting that these parameters reflect similar biological processes and tend to increase together. In contrast, parameters such as hemoglobin and albumin show negative correlations with inflammatory markers. This indicates that as inflammation increases, these values tend to decrease, reflecting a clinically expected relationship.

When symptom-based variables (diarrhea, abdominal pain, weight loss, and bloody stool) are examined, it is observed that correlations among these variables are generally low to moderate. This suggests that symptoms have a more heterogeneous structure and may vary considerably across individuals.

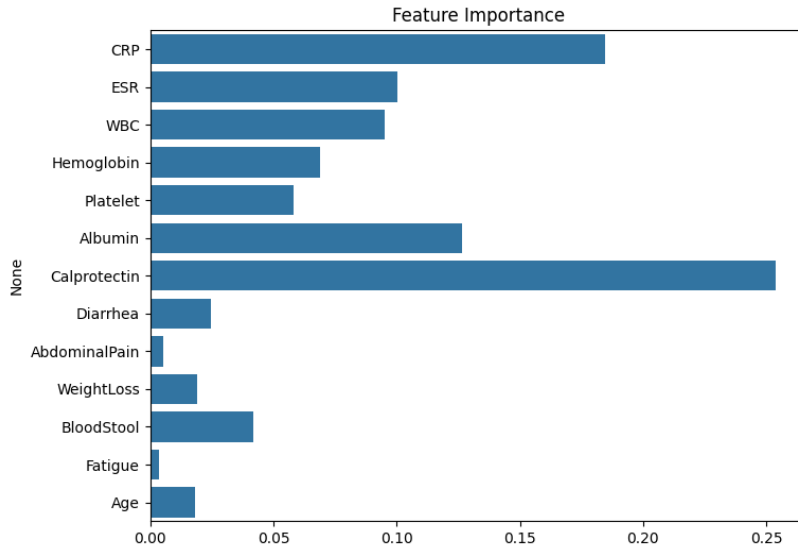


Figure 3. Feature Importance

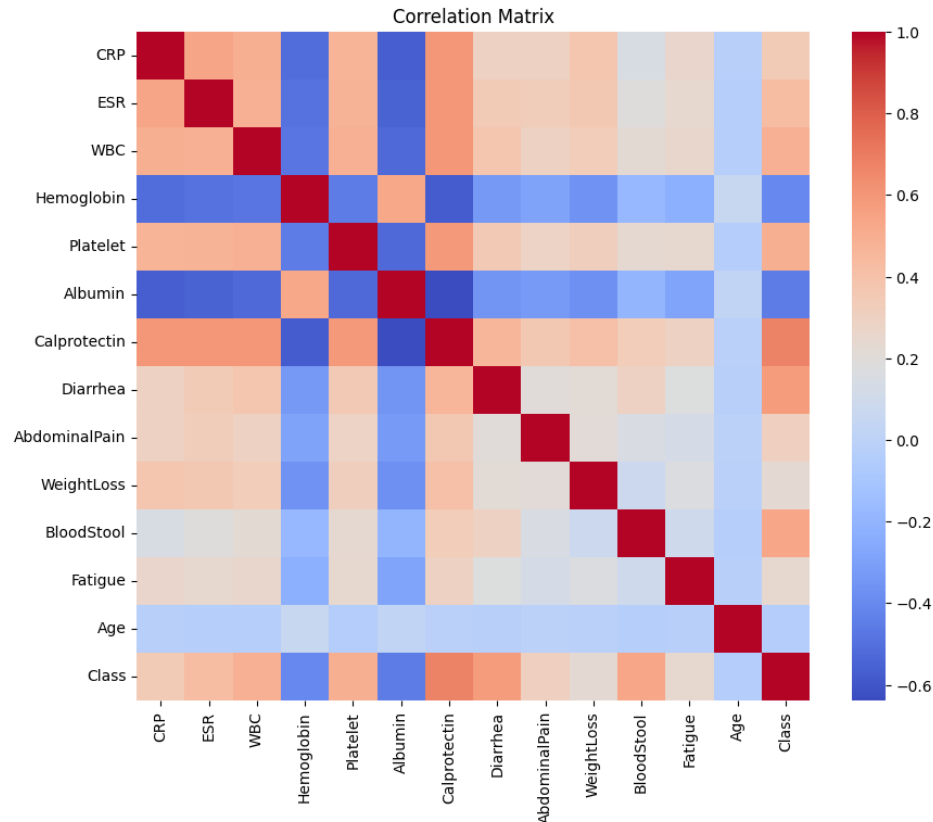


Figure 4. Correlation Matrix

4.3 Proposed Improvements

Although the models achieved high performance, several improvements can be proposed to enhance both methodological rigor and clinical applicability. First, incorporating real-world clinical datasets would significantly improve the external validity of the models. Synthetic data, while useful for controlled experimentation, may not fully capture the complexity and noise present in real clinical environments. Second, additional machine learning models such as gradient boosting algorithms or deep learning-based approaches could be explored to further improve classification performance, particularly in distinguishing between Crohn's disease and ulcerative colitis. Third, hyperparameter optimization techniques could be applied to both models to identify optimal configurations and potentially improve performance. In this study, default or near-default parameters were used, which may not represent the best-performing settings. Furthermore, incorporating temporal data or longitudinal patient records may also improve the ability to capture disease progression patterns.

5. Conclusion

In this study, machine learning models were developed using non-invasive clinical and laboratory data to classify IBD. Analyses conducted with Logistic Regression and Random Forest algorithms demonstrated that healthy individuals can be clearly distinguished from other groups, whereas a notable overlap exists between Crohn's disease and ulcerative colitis. This can be explained by the similar clinical and biochemical characteristics of these two conditions and is consistent with the diagnostic challenges reported in the literature. Overall, this study shows that inflammatory bowel diseases can be classified with high accuracy using non-invasive data. However, the relatively complex nature of distinguishing between Crohn's disease and ulcerative colitis indicates that future studies should focus on more detailed biomarkers, multimodal data sources, or advanced modeling approaches.

References

- Ahmad, H., East, J., Panaccione, R., Travis, S., Canavan, J., Usiskin, K., & Byrne, M. F., Artificial intelligence in inflammatory bowel disease: Implications for clinical practice and future directions, *Intestinal Research*, vol. 21, no. 3, pp. 283–294, 2023.
- Baumgart, D. C., & Carding, S. R., Inflammatory bowel disease: Cause and immunobiology, *The Lancet*, vol. 369, no. 9573, pp. 1627–1640, 2007.
- Baumgart, D., The diagnosis and treatment of Crohn's disease and ulcerative colitis, *Deutsches Arzteblatt International*, vol. 106, no. 8, pp. 123–133, 2009.
- Brodersen, J. B., Jensen, M. D., Leenhardt, R., Kjeldsen, J., Histace, A., Knudsen, T., & Dray, X., Artificial intelligence-assisted analysis of pan-enteric capsule endoscopy in patients with suspected Crohn's disease: A study on diagnostic performance, *Journal of Crohn's and Colitis*, vol. 18, no. 1, pp. 75–81, 2024.
- Burczynski, M., Peterson, R., Twine, N., Zuberek, K., Brodeur, B., Casciotti, L., Reddy, P. S., Strahs, A., Spinelli, W., Schwertschlag, U., Slager, A., Cotreau, M., & Dorner, A., Molecular classification of Crohn's disease and ulcerative colitis patients using transcriptional profiles in peripheral blood mononuclear cells, *The Journal of Molecular Diagnostics*, vol. 8, no. 1, pp. 51–61, 2006.
- Hamamoto, Y., Kawamura, M., Uchida, H., Hiramatsu, K., Katori, C., Asai, H., ... & Yoshida, K., The histological detection of ulcerative colitis using a no-code artificial intelligence model, *International Journal of Surgical Pathology*, vol. 32, no. 5, pp. 890–894, 2024.
- Iacucci, M., Parigi, T. L., Del Amor, R., Mesequer, P., Mandelli, G., Bozzola, A., ... & Villanacci, V., Artificial intelligence enabled histological prediction of remission or activity and clinical outcomes in ulcerative colitis, *Gastroenterology*, vol. 164, no. 7, pp. 1180–1188, 2023.
- Jiang, F., Fu, X., Kuang, K., & Fan, D., [Retracted] Artificial Intelligence algorithm-based differential diagnosis of Crohn's disease and ulcerative colitis by CT image, *Computational and Mathematical Methods in Medicine*, vol. 2022, no. 1, Article 3871994, 2022.
- Kraszewski, S., Szczurek, W., Szymczak, J., Reguła, M., & Neubauer, K., Machine learning prediction model for inflammatory bowel disease based on laboratory markers: Working model in a discovery cohort study, *Journal of Clinical Medicine*, vol. 10, no. 20, Article 4745, 2021.
- Marchesi, J. R., Holmes, E., Khan, F., Kochhar, S., Scanlan, P., Shanahan, F., Wilson, I. D., & Gibson, G. R., Rapid and noninvasive metabonomic characterization of inflammatory bowel disease, *Journal of Proteome Research*, vol. 6, no. 2, pp. 546–551, 2007.
- Meuwis, M., Fillet, M., Geurts, P., de Seny, D., Lutteri, L., Chapelle, J., Bours, V., Wehenkel, L., Belaiche, J., Malaise, M., Louis, E., & Merville, M., Biomarker discovery for inflammatory bowel disease using proteomic serum profiling, *Biochemical Pharmacology*, vol. 73, no. 9, pp. 1422–1433, 2007.
- Sinonquel, P., Schilirò, A., Verstockt, B., Vermeire, S., & Bisschops, R., Evaluating the potential of artificial intelligence in ulcerative colitis, *Expert Review of Gastroenterology & Hepatology*, vol. 17, no. 2, pp. 145–153, 2023.
- Takenaka, K., Ohtsuka, K., Fujii, T., Negi, M., Suzuki, K., Shimizu, H., ... & Watanabe, M., Development and validation of a deep neural network for accurate evaluation of endoscopic images from patients with ulcerative colitis, *Gastroenterology*, vol. 158, no. 8, pp. 2150–2157, 2020.
- Torres, J., Petralia, F., Sato, T., Wang, P., Telesco, S. E., Strauss, R., ... & Colombel, J. F., Serum biomarkers identify patients who will develop inflammatory bowel diseases up to 5 years before diagnosis, *Gastroenterology*, vol. 159, no. 1, pp. 96–104, 2020.

Biographies

Hafsa Nur Peker is a Research Assistant at the Department of Industrial Engineering, Gazi University, Türkiye. She obtained her BSc degree from Gazi University in 2019, and her MSc degree from TOBB University of Economics and Technology in 2022. She has been pursuing her PhD studies at Gazi University since 2022. Her research interests include modeling and analysis of production systems, optimization theory and methods, and machine learning applications.

Gul Didem Batur Sir received the B.Sc. degree in industrial engineering from Gazi University, Ankara, Türkiye, in 2006, the M.Sc. degree in industrial engineering from Bilkent University, Ankara, Türkiye, in 2009, and the Ph.D. degree in industrial engineering from Gazi University, Ankara, Türkiye, in 2014. Her major field of study is industrial engineering. She has held various academic and research positions throughout her career. She is currently a Professor at Gazi University, Ankara, Türkiye. Her research interests include scheduling, production and service systems, and

*Proceedings of the 16th International Conference on Industrial Engineering and Operations Management
Ankara, Türkiye, June 23-25, 2026*

multi-criteria decision making. She has authored/co-authored several research articles in international journals and conferences.