On the Machine Learning-based Multi-class Classification of Microscopic Colitis

Vivek Tara[1], Dipankar Mitra[2,*], Aditi Muduganti[3], Padmavathi Mali[4], Srabana Maiti[5], Shuvashis Dey[5], and Rahul Gomes[6]

[1]Dept. of Physics and Astronomy, University of Wisconsin-La Crosse, WI, USA
[2]Dept. of Computer Science and Computer Engineering, University of Wisconsin-La Crosse, WI, USA
[3]Onalaska High School, Onalaska, WI, USA
[4]Dept. of Internal Medicine, Gundersen Health System, La Crosse, WI, USA
[5]Dept. of Electrical and Computer Engineering, North Dakota State University, Fargo, ND, USA
Dept. of Computer Science, University of Wisconsin-Eau Claire
[*]Corresponding Author: dmitra@uwlax.edu.

***Abstract*—Standard practice for the classification of Microscopic Colitis (MC) as dictated by European guidelines instructs for a complete colonoscopy opposed to the less invasive, but historically inaccurate flexible sigmoidoscopy. A recent study conducted by researchers at Gunderson Health System sought to analyze historical data of patients diagnosed with MC to determine if a complete colonoscopy truly is required to make a diagnosis, or if a flexible sigmoidoscopy would suffice. The results of the study indicated that analysis of patient data obtained through flexible sigmoidoscopy alone provided inaccurate rates of classification of MC. The researchers concluded the results of the study further reinforced European guidelines of performing a complete colonoscopy for the classification of MC. In this work multi-class classification machine learning models are explored to differentiate subtypes of MC on medical data obtained through flexible sigmoidoscopy at Gunderson Health System. The results of the study show that Machine Learning (ML) algorithms have the potential to automate the classification of the traditionally difficult to diagnose disease, MC.**

Keywords—Machine Learning (ML), Microscopic Colitis, Colonoscopy, Topographic distribution, and Multi-class Classification

I. Introduction

The role of physicians has always been to synthesize medically relevant data to identify diagnostic patterns, leading to eventual treatment of the patient [1]. Widespread digitization of medical records combined with the prominence of high-volume data streams have made the medical field an obvious domain for the application of ML models. In recent history deployment of ML models for a wide range of predictive tasks in the medical field have produced repeated successes [2-5]. Medical experts are faced with an increasing number of responsibilities; preventative, diagnostic, administrative, and fiscal [6]. Utilization of ML models for medical purposes has the potential to automate the detection of diseases, even those traditionally difficult to diagnose [7, 8]. ML models have further been shown to have applications in prediction of patient outcome probability [9-11]. Automation of complex clinical tasks will provide more time for medical experts to focus on treatment of patients, rather than diagnosing form of disease [12-15].

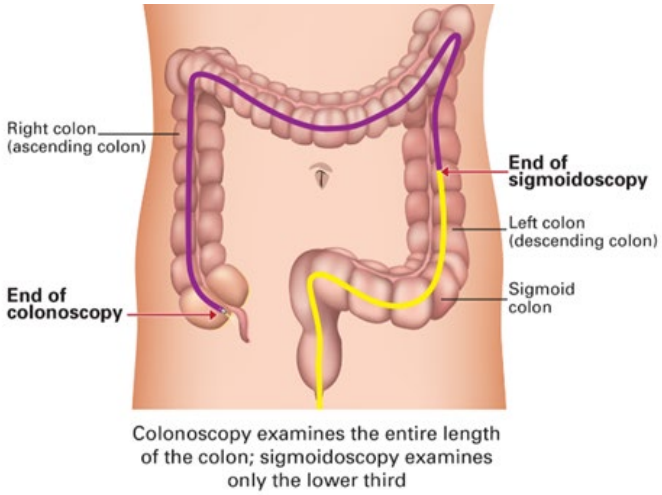

Figure 1. Diagram displaying both Colonoscopy and Sigmoidoscopy Procedure. ©2023 FORCE-Facing Our Risk of Cancer Empowered, Inc. Used with permission.

MC is a disease responsible for chronic inflammation of the large intestine associated with heavy symptom burden and an impaired health-related quality of life (HRQoL) [16-18]. The primary symptom of MC is chronic, non-bloody diarrhea, which may be accompanied by abdominal pain, fecal urgency, or weight loss [19, 20]. MC can be categorized into two major histological sub-types: Collagenous Colitis (CC) and Lymphocytic Colitis (LC), but Incomplete Microscopic Colitis (MCi) may occur [21].

The key histological features of LC are increased Intraepithelial Lymphocytes (IELs), increased inflammatory infiltrates in lamina propria, surface epithelial injury, and undistorted crypt architecture [22]. Increased IELs are typically observed in the surface epithelium of colon tissue obtained through biopsy. In normal colonic mucosa IELs are present in ratios < 5 per 100 epithelial cells, whereas the diagnostic criterion for LC is IELs present in ratios > 20 per 100 epithelial cells [23]. High cellular infiltrate density in the lamina propria makes the colonic mucosa appear "bluer" than normal colonic

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

mucosa [24]. Thickening of the sub-epithelial collagen band may be present in LC < 10μm [25, 26].

The primary histological difference between LC and CC is the presence of a thickened subepithelial collagen band > 10μm. Increased ILEs may be present in CC; however, with less regularity than LC < 20 per 100 epithelial cells. Other histological features of CC are shared with LC such as surface epithelial injury, increased inflammatory infiltrates in lamina propria, and undistorted crypt architecture [22 - 26].

Classification of MC requires histological analysis of tissue samples that must be obtained through biopsy [25]. Depiction of Colonoscopy vs flexible Sigmoidoscopy is shown in Figure 1. The topographic distribution of MC within the colon remains controversial. According to the literature, the histological changes, especially in CC, may be patchy and not continuously distributed throughout the colon [26]. Current European guidelines instruct a complete colonoscopy with right and transverse colon biopsies, a highly invasive procedure, opposed to a flexible sigmoidoscopy with rectal biopsies. The reason for this guideline is historically inaccurate rates of classification of MC when tissue samples from rectal biopsies alone are used for detection of the disease [27].

Colonoscopy is a procedure entailing utilization of a long flexible tube with a small video camera attached for examination of the entire length of the colon and rectum. Sedation of the patient is required. Flexible sigmoidoscopy, like the colonoscopy procedure, utilizes a long flexible tube with a small video camera attached, however examination is limited up to the descending colon. Most patients do not require anesthesia for the completion of this medical procedure [28].

A recent study conducted by researchers at Gunderson Health System sought to analyze historical data on patients diagnosed with MC to determine if a complete colonoscopy with right and transverse colon biopsies was truly required for the classification of MC, or if a flexible sigmoidoscopy with rectal biopsies would suffice. The results of the study indicated that 53.8% of the diagnoses for MC would be missed by analysis of patient tissue obtained through rectal biopsy alone. The researchers concluded that their work further reinforced the current European guidelines of performing a complete colonoscopy with right and transverse colon biopsies for the classification of microscopic colitis [29].

Six ML models are trained on medical data obtained through rectal biopsies at Gunderson Health System. The models aim to classify the sub-type of MC a patient has with a greater rate of accuracy than has previously been obtained through human analysis of the same dataset [29]. Successful detection of MC through less-invasive techniques reduces risk to patient health during diagnostic procedures for MC.

II. Materials & Methods

A. Dataset

Study approval and waiver of informed consent were granted by the Human Subjects Committee / Institutional Review Board of Gundersen Clinic, Ltd. and Gundersen Lutheran Medical Center, Inc. The permission to use the data was given under a data use agreement between Gundersen Lutheran Administrative Services, Inc., a Wisconsin non-stock corporation, individually and as agent for Gundersen Lutheran Medical Foundation, Inc., Gundersen Clinic, Ltd. and Gundersen Lutheran Medical Center, Inc. ("Recipient") and Board of Regents of the University of Wisconsin System d/b/a University of Wisconsin – La Crosse ("Covered Entity"), Dr. Dipankar Mitra, acting as Principal Investigator. The purpose of this Agreement is to provide Recipient with access to a Limited Data Set ("LDS") for use in its Research and analyses and for the Health Care Operations of the Covered Entity, in accordance with the HIPAA Regulations.

The study was conducted in a single tertiary care facility located in the Midwest. The electronic health records (EHRs) of all patients with a histological diagnosis of lymphocytic, collagenous, or microscopic colitis from January 1, 1999, through December 18, 2019, were retrospectively reviewed. Patients aged less than 18 were excluded. Demographic variables, including age at diagnosis, sex, race and ethnicity, and tobacco use history and status at diagnosis (i.e., whether a current, former, or never smoker), were collected from the EHR. Clinical data, such as medication use (nonsteroidal anti-inflammatory drug or budesonide) and comorbidities (celiac disease, diabetes type II, hyper- or hypoparathyroidism, fibromyalgia, autoimmune thyroiditis, pernicious anemia, autoimmune hepatitis, ankylosing spondylitis, rheumatoid arthritis, Raynaud syndrome, and Sjogren syndrome), were also collected. Variables of primary interest for the study objective included the results of pathologic evaluation of the patient's biopsy specimens obtained from left-sided and random colonic locations, whether positive or negative for microscopic colitis. The frequency with which pathologic diagnoses of left-sided and random colonic tissue specimens were negative or positive for microscopic colitis was then compared. The Kappa coefficient was used as a measure of interrater reliability for agreement. A p-value of less than .05 was considered significant.

B. Machine Learning Models

Six classification ML models were utilized for determination of subtype of MC on patient data obtained through flexible sigmoidoscopy: Decision Trees, Random Forest Classifiers, Support Vector Machines, K-Nearest Neighbors, Multinomial Naïve Bayes, and Neural Networks. The algorithms that underlie each of the applied models and their selected hyperparameters are outlined below.

C. Decision Trees & Random Forrest Classifiers

Given training vectors $x_i \in \mathbb{R}^n$, $i = 1, \ldots, l$ and a label vector $y \in \mathbb{R}^l$, a decision tree recursively partitions the feature space such that samples with the same labels are grouped together.

Let the data at node m be represented by Q_m, with n_m samples. For each potential split $\theta = (j, t_m)$ consisting of feature j and threshold t_m partition the data into $Q_m^{left}(\theta)$ and $Q_m^{right}(\theta)$ subsets.

$$Q_m^{left}(\theta) = \{(x, y)|x_j \leq t_m|\} \quad (1)$$
$$Q_m^{right}(\theta) = \{(x, y)|x_j \geq t_m|\} \quad (2)$$

The quality of the potential split is then evaluated using an impurity function $H(Q_{m,}, \theta)$.

$$G(Q_m, \theta) = \frac{n_m^{left}}{n_m} \cdot H(Q_m^{left}) + \frac{n_m^{right}}{n_m} \cdot H(Q_m^{right}) \quad (3)$$

Where the impurity function utilized was the entropy function:

$$H(Q_m, \theta) = -\sum_{i=1}^{k} p_i \log_2(p_i) \quad (4)$$

Select the parameters that minimize impurity:

$$\theta^{\cdot} = \min(G(Q_m, \theta)) \quad (5)$$

Recurse for subsets Q_m^{left} and Q_m^{right} until maximum allowable depth is reached, or $n_m = 1$. Random Forest Classifiers are composed of independent decision trees that are ensembled together before prediction.

The maximum allowed depth of the model was selected to be 5. A node had to contain at least 14 examples to be considered a leaf node. The Random Forest classifier was composed of five decision trees. The maximum allowed depth of the model was selected to be 12. The models were allowed to select from a maximum of 4 features before separation of the dataset.

D. Mulit – Class Support Vector Machine

Given training vectors $x_i \in \mathbb{R}^n$, $i = 1, \ldots, l$, in two classes, and a label vector $y \in \{-1, 1\}^l$ s, SVM produces $w \in \mathbb{R}^n$ and $b \in \mathbb{R}^n$ such that the prediction given by $sign(w^T \cdot X + b)$ is correct for most samples.

The SVM algorithm solves the following primal problem:

$$\min_{w, b, \varsigma} \frac{1}{2}||w||^2 + C\sum_{i=1}^{l} \zeta_i \quad (6)$$

$$subject\ to\ y_i(w^T \cdot X + b) \geq 1 - \varsigma_i, \quad (7)$$

$$\varsigma_i \geq 0, i = 1, \ldots, n$$

Where the objective is to minimize the value of $||w||^2$ which can be shown to be inversely proportional to magnitude of the margin. Where the constraint is forcing the binary classifications onto opposing sides of "positive" and "negative" hyperplanes given a small error ζ that also being minimized [30, 31].

3 Binary SVM Classifiers were implemented, namely (i) LC vs CC, (ii) LC vs MCi, and (ii) CC vs MCi. To make a prediction on the test set an example is processed by each of the binary classifiers. Each classifier will predict one of the two possible outcomes and the class with the most votes will be predicted by the model. The model was implemented using a 6th degree polynomial kernel. The value of C was set to 2.00.

E. K – Nearest Neighbors

KNNs are supervised classification ML algorithms. A given example in the test set will be allocated to the class that is most prevalent amongst its k-nearest neighbors in the feature space, where k must be a positive integer. The Minkowski distance metric was utilized to calculate distance between samples in the feature space:

$$\sum_{i=1}^{n}((x_i - y_i)^p)^{\frac{1}{p}} \quad (8)$$

where, $p = 5$. The k-nearest neighbors' model was implemented with $k = 13$ and $p = 5$ using the Minkowski distance metric [30].

F. Multinomial Naïve Bayes

Multinomial Naïve Bayes is a supervised classification ML algorithm. The distribution is parameterized by vectors $\vec{\theta}_y = (\theta_{y1}, \ldots, \theta_{yn})$ for each class y, where n is the number of features, and θ_{yi} is the probability $P(x_i|y)$ of feature i appearing in a sample belonging to class y. The parameters θ_y are estimated using a smoothed version of maximum likelihood.

$$\theta_{yi} = \frac{N_{yi} + \alpha}{N_y + \alpha n} \quad (9)$$

Where N_{yi} is the number of times feature i appears in a sample of class y in the training set and N_y is the total number of features. Laplace smoothing was used. The model then divides the test set according to the following classification rule:

$$\hat{y} = argmax P(y) \prod_{i=1}^{n} P(x_i|y) \quad (10)$$

G. Artificial Neural Networks

ANNs are supervised ML models whose structure is inspired by the biological brain. The fundamental unit of an ANN, alike to the biological brain, is a single computational element called a neuron. The purpose of a neuron is to perform a non-linear transformation on the weighted sum, or activation, of its inputs.

$$Neuron_{output} = NonLinearTransformation(\overline{w} \cdot \overline{x} + b) \quad (11)$$

Where the weight vector $\overline{w}$ and bias term b are the trainable parameters associated with the neuron.

ANNS are comprised of layered arrangements of neurons. Dense layers were utilized to construct an ANN structure in which each neuron in each layer had connections to every neuron in the previous layer. Each neuron in the input layer is representative of one of the features in the dataset. Each neuron in the output layer is representative of one subtype of MC, or more generically, one possible prediction of the model.

The Rectified Linear Unit (ReLu) activation function was utilized in the hidden layers of the implemented model to introduce non-linearity into the fit decision boundary.

$$f(x) = \max(0, x) \quad (12)$$

The SoftMax activation function was utilized in the output layer of the implemented model to convert the raw output of the ANN into a vector of probabilities.

$$\text{SoftMax}(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{x_j}} \quad (13)$$

Training of an ANN can be broken down into two phases: the forward pass and the backwards pass. During the forward pass of an ANN, a vector containing each of the features of a given sample enters the model through the input layer. These features propagate down the network, transformed using equations 10 and 11 as activation functions in their respective layers. After passing through the output layer the forward pass of the model is completed and the ANN has made a prediction based on the current values of its trainable parameters.

During the backwards pass of the model, separation between predicted and real values is quantized with the use of a loss function. The loss function is a function of the trainable parameters. The purpose of ANNs is to minimize the value of the loss function, in our case Categorical Cross Entropy, across all examples in the training data by iteratively updating the trainable parameters in the model after each forward pass.

$$Categorical\ Cross\ Entropy = \sum_{i=1}^{C} t_i \log(soft\max(z)_i) \quad (14)$$

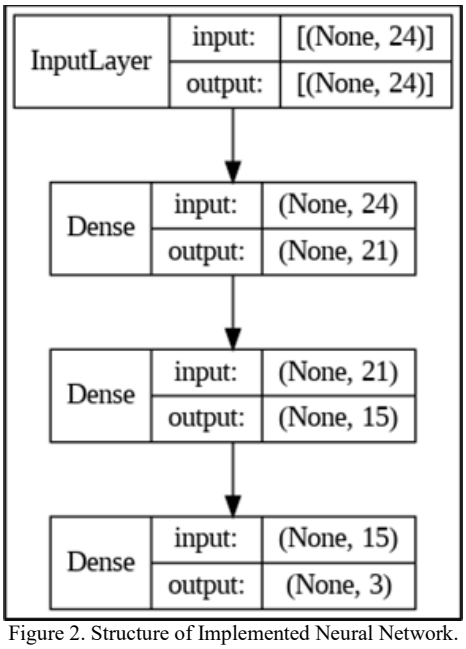

Figure 2. Structure of Implemented Neural Network.

Because $t_i = 1$, indicating a positive true label, for one of the classes, and $t_i = 0$, indicating a negative true label, for the rest. The loss function can be simplified:

$$Categorical\ Cross\ Entropy = -\log\left(\frac{e^{z_i}}{\sum_{j=1}^{C} e^{x_j}}\right) \quad (15)$$

The trainable parameters in the model are then updated through the iterative algorithm gradient descent. The gradient of the loss function is taken, providing a vector pointing in the direction of greatest increase of the function, given an infinitesimal adjustment of the model's parameters. The trainable parameters are then updated according to the following equations:

$$w_j = w_j - \alpha \frac{\partial J(\overline{w}, \overline{b})}{\partial w_j} \quad (16)$$

$$b_j = b_j - \alpha \frac{\partial J(\overline{w}, \overline{b})}{\partial b_j} \quad (17)$$

Where α is the learning rate, a hyperparameter determining rate of change of the parameters after each iteration.

During the training phase of the neural network, the model cyclically produces predictions in the forward pass, calculates the error with a loss function, and adjusts trainable parameters via gradient descent. This process iterates until the loss decreases below a threshold hyperparameter or reaches the maximum allowed iterations. The model fits a mathematical function to the training data, and its validity is assessed using testing data. The implemented neural network model is illustrated in Figure 2.

III. Results & Discusion

Six ML models—Decision Trees, Random Forest Regressors, Support Vector Machines, K-Nearest Neighbors, Multinomial Naïve Bayes, and Neural Networks—were applied to classify the dataset. Categorical accuracy scores for each of the models are presented in Table 1.

Table 1. Categorical Accuracy scores for each of the implemented models.

Model	Categorical Accuracy
Decision Tree	39.70%
Random Forrest Classifier	58.55%
Support Vector Machine	55.41%
K-Nearest Neighbors	38.26%
Multinomial Naïve Bayes	55.88%
Neural Network	47.06%

A. Discussion

This study, to the authors' knowledge, is the first to document classification models differentiating sub-types of MC using patient data from flexible sigmoidoscopies with rectal biopsies. ML models (Decision Trees, Random Forest Classifiers, Support Vector Machines, K-Nearest Neighbors, Multinomial Naïve Bayes, and Neural Networks) were programmed to maximize categorical accuracy scores.

Traditional methods for classifying MC involve meticulous collection, processing, and evaluation of samples by skilled medical professionals. European guidelines recommend a comprehensive colonoscopy with right and transverse colon biopsies, which is invasive compared to flexible sigmoidoscopy with rectal biopsies, leading to inconsistencies in disorder recognition. This approach is labor-intensive and susceptible to individual subjectivity influencing results. ML models in medical applications have the potential to automate disease detection, even in traditionally challenging diagnoses, playing a significant role in reducing subjective errors, automating complex clinical decisions, and enhancing diagnostic precision.

The dataset was divided into 80% training and 20% testing sets. Six ML models—Decision Trees, Random Forest Classifiers, Support Vector Machines, K-Nearest Neighbors,

Naïve Bayes, and Neural Networks—were applied to classify medical data. The Random Forrest Classification model achieved the highest test set accuracy at 58.55%. Furthermore, the Support Vector Machine, Multinomial Naïve Bayes, and Artificial Neural Network models were shown to correctly diagnose the subtype of MC a patient had with accuracy greater than 46.2%, what was obtained through human analysis of the same dataset. All other models performed equally in terms of categorical accuracy. Results suggest that ML models can differentiate histological features of the major MC sub-types, CC, and LC, in their respective feature spaces. This study demonstrates the potential of combining medical data from flexible sigmoidoscopy with rectal biopsies and ML models to automate MC subtype classification, minimizing risks to patient care.

There are some limitations to this study. Successful implementation of ML models typically requires large datasets to fit complex and robust mathematical decision boundaries. The sample size is limited to 337 patients with previous histological diagnosis of MC. Furthermore, MC is typically classified into LC and MC based on pathology findings. The use of patient characteristics to train the ML models limits the potential for accurate classification of MC. The overall weakness of this study can be attributed to limitations of the dataset. Additional research with larger sample populations and utilization of pathological findings may provide greater rates of classification than seen in this study.

Using ML, three MC sub-types were classified; the random forest classifier achieved the best results, distinguishing between sub-types with 58.55% categorical accuracy. Implementing ML in MC diagnosis has the potential to minimize patient health risks and automate the labor-intensive, bias-prone diagnostic process. The multi-class model shows great promise in characterizing and classifying diverse MC subtypes using patient data from flexible sigmoidoscopy with rectal biopsies.

IV. Conclusion

In conclusion, this study marks a significant milestone as the first to document the application of classification models to differentiate sub-types of MC using patient data obtained through flexible sigmoidoscopies with rectal biopsies. The utilization of ML models, including Decision Trees, Random Forest Classifiers, Support Vector Machines, K-Nearest Neighbors, Multinomial Naïve Bayes, and Neural Networks, showcased their potential in maximizing categorical accuracy scores. Traditional methods for MC classification, often labor-intensive and invasive, were contrasted with the automated and potentially more precise approach offered by ML models. The study's findings, particularly the success of the random forest classier in achieving the highest test set accuracy at 58.55%, underscore the potential of ML in distinguishing histological features of major MC sub-types, such as CC and LC. Despite the acknowledged limitations, including the relatively small sample size of 337 patients, this research highlights the viability of combining medical data from flexible sigmoidoscopy with rectal biopsies and ML models to automate MC subtype classification. Moving forward, further research with larger sample populations holds the promise of enhancing the classification rates observed in this study, paving the way for more effective and automated MC diagnoses that can minimize risks to patient care.

Acknowledgment

Data was collected for the study titled "Utility of Rectal Biopsies Compared to Random Colon Biopsies in Microscopic Colitis" under IRB Protocol Number 2-19-12-007. Dipankar Mitra and Vivek Tara express gratitude to UW-La Crosse's College of Science and Health for research support through the Dean's Distinguished Fellowship (DDF).

References

[1] Quer G, Arnaout R, Henne M, and Arnaout R., "Machine Learning and the Future of Cardiovascular Care: JACC State-of-the-Art Review", J Am Coll Cardiol. 2021 Jan 26;77(3):300-313.

[2] Huertas-Fernández I, García-Gómez FJ, García-Solís D, Benítez-Rivero S, Marín-Oyaga VA, Jesús S, Cáceres-Redondo MT, Lojo JA, Martín-Rodríguez JF, Carrillo F, and Mir P., "Machine learning models for the differential diagnosis of vascular parkinsonism and Parkinson's disease using [(123)I]FP-CIT SPECT.", Eur J Nucl Med Mol Imaging. 2015 Jan;42(1):112-9.

[3] Gautam D, Ahmed M, Meena YK, and Ul Haq A., "Machine learning-based diagnosis of melanoma using macro images", Int J Numer Method Biomed Eng. 2018 May;34(5):e2953.

[4] Wu M, Krishna S, Thornhill RE, Flood TA, McInnes MDF, and Schieda N., "Transition zone prostate cancer: Logistic regression and machine-learning models of quantitative ADC, shape and texture features are highly accurate for diagnosis", J Magn Reson Imaging. 2019 Sep;50(3):940-950.

[5] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, and Thrun S., "Dermatologist-level classification of skin cancer with deep neural networks", Nature. 2017 Feb 2;542(7639):115-118.

[6] Pedersen KM, Andersen JS, and Søndergaard J., "General practice and primary health care in Denmark", J Am Board Fam Med. 2012 Mar;25 Suppl 1:S34-8.

[7] Dros JT, Bos I, Bennis FC, Wiegersma S, Paget J, Seghieri C, Barrio Cortés J, and Verheij RA., "Detection of primary Sjögren's syndrome in primary care: developing a classification model with the use of routine healthcare data and machine learning", BMC Prim Care. 2022 Aug 9;23(1):199.

[8] Iwamura H, Mizuno K, Akamatsu S, Hatakeyama S, Tobisawa Y, Narita S, Narita T, Yamashita S, Kawamura S, Sakurai T, Fujita N, Kodama H, Noro D, Kakizaki I, Nakaji S, Itoh K, Tsuchiya N, Ito A, Habuchi T, Ohyama C, and Yoneyama T., "Machine learning diagnosis by immunoglobulin N-glycan signatures for precision diagnosis of urological diseases", Cancer Sci. 2022 Jul;113(7):2434-2445.

[9] Zampieri FG *et al.*, "ICU staffing feature phenotypes and their relationship with patients' outcomes: an unsupervised machine learning analysis", Intensive Care Med. 2019 Nov;45(11):1599-1607.

[10] Chun X, Pan L, Lin Y, Ye L, Liang H, Tao J, and Luo Y., "A model for predicting 7-day pressure injury outcomes in paediatric patients: A machine learning approach", J Adv Nurs. 2021 Mar;77(3):1304-1314.

[11] Pfob A, Mehrara BJ, Nelson JA, Wilkins EG, Pusic AL, and Sidey-Gibbons C., "Machine learning to predict individual patient-reported outcomes at 2-year follow-up for women undergoing cancer-related mastectomy and breast reconstruction (INSPiRED-001)", Breast. 2021 Dec;60:111-122.

[12] Obermeyer Z, Emanuel EJ., "Predicting the Future - Big Data, Machine Learning, and Clinical Medicine", N Engl J Med. 2016 Sep 29;375(13):1216-9.

[13] Chen JH, Asch SM., "Machine Learning and Prediction in Medicine - Beyond the Peak of Inflated Expectations", N Engl J Med. 2017 Jun 29;376(26):2507-2509.

[14] Quer G, Arnaout R, Henne M, and Arnaout R., "Machine Learning and the Future of Cardiovascular Care: JACC State-of-the-Art Review", J Am Coll Cardiol. 2021 Jan 26;77(3):300-313.

[15] Lina Zhou, Shimei Pan, Jianwu Wang, and Athanasios V. Vasilakos, “Machine learning on big data: Opportunities and challenges”, Neurocomputing, Volume 237, 2017, Pages 350-361.

[16] Miehlke S. *et al.*, “European guidelines on microscopic colitis: United European Gastroenterology and European Microscopic Colitis Group statements and recommendations”, United European Gastroenterol J. 2021 Feb 22;9(1):13–37.

[17] Burke KE *et al*., “Identification of Menopausal and Reproductive Risk Factors for Microscopic Colitis-Results From the Nurses' Health Study”, Gastroenterology. 2018 Dec;155(6):1764-1775.

[18] Tachibana M, Hanaoka T, Watanabe S, Matsushita M, Isono T, and Tsutsumi Y., “Colitis nucleomigrans: The third type of microscopic colitis (part 1)”, Pathol Int. 2020 Oct;70(10):752-760.

[19] Münch A, Aust D, Bohr J, Bonderup O, Fernández Bañares F, Hjortswang H, Madisch A, Munck LK, Ström M, Tysk C, and Miehlke S, “Microscopic colitis: Current status, present and future challenges: statements of the European Microscopic Colitis Group”, J Crohns Colitis. 2012 Oct;6(9):932-45.

[20] Ek M, Roth B, Bengtsson M, and Ohlsson B., “Gastrointestinal Symptoms in Women With Endometriosis and Microscopic Colitis in Comparison to Irritable Bowel Syndrome: A Cross-Sectional Study”, Turk J Gastroenterol. 2021 Oct;32(10):819-827.

[21] Langner C *et al*., “Histology of microscopic colitis-review with a practical approach for pathologists”, Histopathology. 2015 Apr;66(5):613-26.

[22] Gentile NM, Yen EF., “The Incidence of Microscopic Colitis: Microscopic No More”, Dig Dis Sci. 2017 Jun;62(6):1394-1395.

[23] Singh H, Demers AA, and Bernstein CN., “Screening colonoscopy vs flexible sigmoidoscopy”, JAMA. 2010 Nov 10;304(18):2017; author reply 2017-8. doi: 10.1001/jama.2010.1585. PMID: 21063010.

[24] Yuan L, Wu TT, and Zhang L., “ Microscopic colitis: lymphocytic colitis, collagenous colitis, and beyond”, Hum Pathol. 2023 Feb;132:89-101.

[25] Yu, Lean *et al.* “Missing Data Preprocessing in Credit Classification: One-Hot Encoding or Imputation?” *Emerging Markets Finance and Trade* 58 (2020): 472 - 482.

[26] Guo, Zhiqiang *et al.* “Novel computer-aided lung cancer detection based on convolutional neural network-based and feature-based classifiers using metaheuristics.” *International Journal of Imaging Systems and Technology* 31 (2021): 1954 - 1969.

[27] Setia, Namrata *et al.* “Lymphocytic colitis: pathologic predictors of response to therapy.” *Human pathology* 78 (2018): 1-7 .

[28] A.-M. K. Fiehn *et al.*, “Distribution of histopathological features along the colon in microscopic colitis,” International Journal of Colorectal Disease, vol. 36, pp. 151–159, 2021.

[29] M. Olesen, S. Eriksson, J. Bohr, G. Järnerot, and C. Tysk, “Microscopic colitis: a common diarrhoeal disease : an epidemiological study in Örebro, Sweden, 1993–1998,” *Gut,* vol. 53, no. 3, pp. 346–350, 2004.

[30] S. Kalyani and K. Swarup, “Classification of Static Security Status Using Multi-Class Support Vector Machines”, *Jou. Eng. Res.*, vol. 9, no. 1, pp. 21–30, Jun. 2012.

[31] B. Vidya, Sasikumar P, “Gait based Parkinson’s disease diagnosis and severity rating using multi-class support vector machine,” *Applied Soft Computing*, Volume 113, Part B, 2021, 107939, ISSN 1568-4946.