
ISSN: 2321-2152
IJMECE
International Journal of modern electronics and communication engineering
E-Mail
editor.ijmece@gmail.com
editor@ijmece.com
www.ijmece.com

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

Using Machine Learning for IBD Severity Classification and Prognosis

[1] S. Venkateswara Rao, [2]T.Madhuri,

[1]Assistant Professor, Megha Institute of Engineering & Technology for Women, Ghatkesar.
[2] MCA Student, Megha Institute of Engineering & Technology for Women, Ghatkesar.

Abstract

A new method for determining the severity of inflammatory bowel disease in young people based on vitamin D levels is presented in this research. Interstitial bowel disease (IBD) is a term for chronic inflammation of the gastrointestinal tract. In inflammatory bowel disease (IBD) patients in particular, low vitamin D levels are associated with an increased risk of colon cancer. Vitamin D may preserve intestinal health if begun sooner in inflammatory bowel disease patients. In healthcare, machine learning is useful because it can quickly and accurately identify the existence of an illness and because it helps clinicians provide better treatment options to patients with the use of AI. Vitamin D levels, measured by serum 25(OH) D concentration, are examined in this research using an accessible dataset pertaining to inflammatory bowel disease (IBD) patients. Patients are categorized as low risk, moderate risk, or high risk according to the severity index. We train and compare ensemble boosted tree classifiers, tree classifiers, and Support Vector Machines (SVMs). Both healthy individuals and those with inflammatory bowel disease (IBD) ranging in age from 2 to 20 make up the dataset's 31 characteristics. With a maximum accuracy of 98% and an area under the ROC curve of 0.98, the ensemble trees classifier is the most effective.

Keywords

inflammatory bowel disease, dietary vitamin D, support vector machine, tree classifier.

I. INTRODUCTION

The advent of machine learning has revitalized the healthcare sector and raised new expectations. Healthcare providers may benefit from clinical perceptions provided by machines educated with large datasets; these practitioners can then use this information to better plan and provide treatment, as well as to take preventative actions, all of which lead to improved results at lower cost and with more patient satisfaction. The prevalence of Inflammatory Bowel Disease (IBD) is on the rise worldwide. The global incidence of inflammatory bowel disease was 6.8 million in 2017. There is a strong correlation between insufficient vitamin D and inflammatory bowel disease. Using a publicly accessible dataset and machine learning classifiers, this research categorizes vitamin D status according to illness severity. The following is the outline of the paper. Inflammatory bowel disease and its subtypes are introduced in Section II. The methods of machine learning are introduced in Section III. Classifiers such as decision trees, Support Vector Machines, and Ensemble trees are described in depth in Section IV. The paper's design process is detailed in Section V. A comparison of several classifiers is presented in Section VI. After Section VII, the paper is concluded.

II. INSIGHTS INTO IBD

To start, persistent inflammation of the digestive system is known as Inflammatory Bowel Disease (IBD). Irritable bowel syndrome (IBS) is characterized by extreme diarrhea, stomach discomfort, lethargy, and loss of weight. IBD is associated with potentially fatal consequences. Two types of inflammatory bowel disease include Chrohn's disease (CD) and ulcerative colitis (UC). The rectum and colon's inner linings may develop ulcers as a result of Ulcerative Colitis. Swelling begins on the intestinal lining and progresses deep into other tissues in Chrohn's disease (CD). We don't know what causes inflammatory bowel disease (IBD), but a malfunctioning immune system is a key factor. Colon cancer is more common in those with inflammatory bowel disease. [1] Relative to individuals without IBD, those with low vitamin D levels had a higher chance of developing colon cancer, according to the research. Additionally, those with vitamin D deficiencies have a much higher chance of requiring surgery or hospitalization. This

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

means that vitamin D insufficiency should be considered in all inflammatory bowel disease patients and that they should all take vitamin D supplements. Preventative medicine is one area where machine learning may have a significant impact. [2].

III. MACHINE LEARNING

A branch of AI known as "machine learning" allows computers to "learn" new skills and improve existing ones automatically, without human intervention. In order to make better decisions, the learning process begins with data that is analyzed for patterns. There are two main categories of machine learning algorithms: supervised and unsupervised. A. earning Under Supervision Predictions are made using Supervised Learning using trained labeled data. When making decisions, supervised learning employs two methods. For data class predictions, it use classification, and for value predictions, it employs regression. Labeled data makes this strategy shine. Some of the methods used in supervised learning include decision trees, random forests, classification, and regression.Section B: Unsupervised Inference When dealing with unlabeled data, these learning techniques come in handy. It creates clusters to reveal patterns in the data. Among the clusters that are most closely related, it inserts the data. C. Reinforcement Learning in 2021 With this strategy, success is more or less certain. With each affirmative reaction, the agent receives a reward, and with each bad response, they are punished. In order to learn, the model uses incentives, and it can also forecast its own value.

IV. Machine Learning-Based Classification

The goal of data classification is to assign a predetermined label to each piece of inputted information. menu bar. An algorithm for decision trees One common use of decision tree classifiers is pattern recognition. Their great productivity and adaptability allow them to do categorization effectively. For nonlinear classes, this classifier works well. Missing data is no problem for decision tree classifiers. By recognizing lines, it repeatedly breaks the plot into subparts. A decision tree may stand in for any input attribute function that takes a Boolean value. Representation of products added up. A discontinuous normal form is a common way to describe the sum of products (SOP). Any branch that continues from the tree's root to a leaf node that also belongs to the same class is called a conjunction, and any branches that terminate in that class are called disjunctions, or sums.

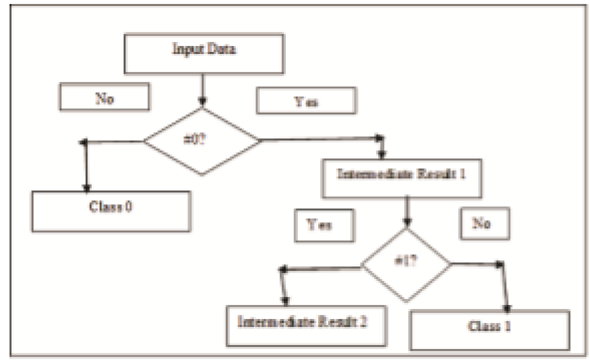
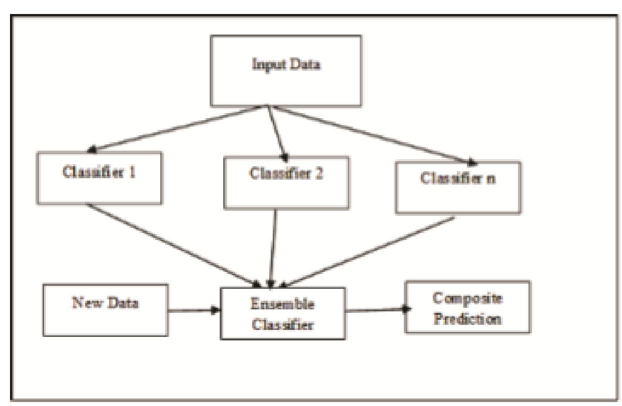


Fig. 1. Decision tree classifier

Classifier B. Support Vector Machine Another way to look at it is as a separating hyperplane that represents a selective classifier. Classification is accomplished by it. Decision boundaries are hyperplanes. Specific properties provided as input determine the hyperplane's dimensions. Data points that affect the hyperplane's location and orientation are called support vectors. As much space as possible is filled in between the data points and the hyperplane in support vector machines. This margin is maximized using the loss function. When the sign of the anticipated and actual values is identical, the cost is set to zero. In that case, the loss value is determined. A very effective algorithm is support vector machine. The SVM is versatile enough to handle a variety of mathematical tasks. SVMs are mainly grouped into two types: linear and gaussian. The linear support vector machine's (SVM) hyperplane separation mathematical function determines whether the model is quadratic, cubic, or linear. Ensemble tree classifier (C) In order to enhance the model's prediction over a single model, the ensemble approach mixes many classifiers. To make the decision tree more consistent, bagging is used.



ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

Fig. 2. Ensemble classifier

V. DESIGN METHODOLOGY

Using an accessible dataset [3], the researchers in this study compared the vitamin D status of individuals with inflammatory bowel disease (IBD) with healthy controls, spanning the ages of 2 to 22 (both sexes). We do the categorization as Age, weight, serum25OH(D) level, erythrocyte sedimentation rate (ESR), severity score, kind of inflammatory bowel disease (IBD), and 22 more characteristics are included in the dataset. The steps involved in training a classifier are shown in Figure 3.

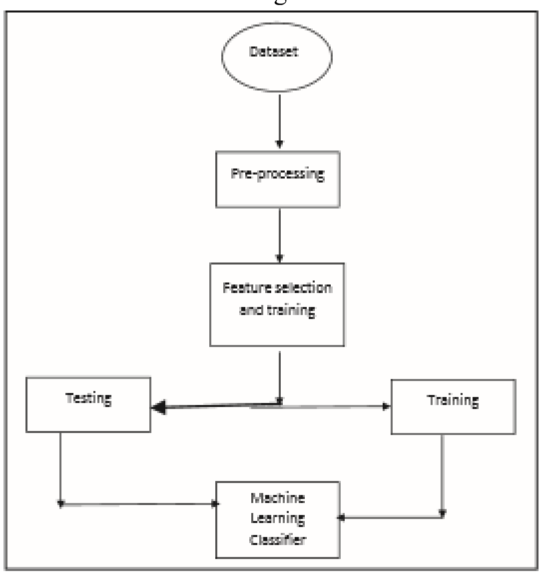


Fig.3. Processof classification using machine learning classifier.

There are three categories used to categorize the dataset. Level 0: the most severe, Level 1: the most severe Level 2 safety. Classification entails the following steps: The first step is to open the data file in MATLAB. .csv format Secondly, launch the app that teaches categorization. Step three: bring data into the app. Choose the attributes and answers that will serve as predictors. 5. Choose the method of cross-validation. 5. After that, you may choose to either activate or disable PCA. 7. Train the classifier of your choice. 8. To begin training, click on the training tab. 9. Examine the outcomes by means of a confusion matrix, ROC curve, and scatter plots.

VI. RESULTS AND DISCUSSION

Several classifiers have been trained on this dataset once it was imported to MATLAB. Based on the severity of inflammatory bowel disease (IBD), the classifier sorts the patients into male and female categories using the dataset. After that, many metrics like accuracy, ROC area, false-negative rate, confusion matrix, true positive rate, etc. are used for the comparison study. dining room I analyze these outcomes comparatively. Screenshots of training outcomes using the classification learner app and MATLAB software are shown in Figures 3–8. Every classifier was trained and had identical results.

TABLE I. CLASSIFIER PERFORMANCE

classifier	*Accuracy*	*AUC*	*True positive rate*		
			Class 0	*Class 1*	*Class 2*
Linear SVM	15.8	0.92	95	25	0
Quadratic SVM	14.4	0.83	86	25	0
Cubic SVM	13.7	0.74	77	38	0
Fine Gaussian SVM	0.7	0.05	5	0	0
Medium Gaussian SVM	11	0.72	73	0	0
Coarse Gaussian SVM	14.4	0.93	95	0	0
Tree Classifier	**98.6**	**0.98**	**95**	**100**	**99**
Ensemble(boosted Trees)	79.5	-	0	0	100
Ensemble(BaggedTrees)	91.8	0.98	86	0	99
RUS (boosted Trees)	20.5	1	100	100	0
Subspace KNN	93.2	0.56	82	25	100

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

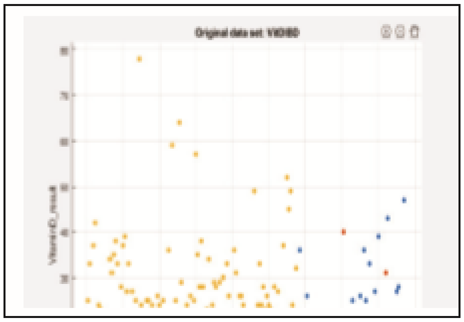


Fig. 3. Scatter Plot of Original Dataset

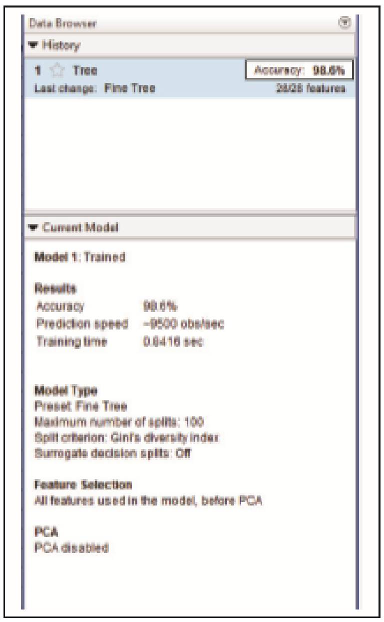


Fig. 4. Screenshot while training data using tree classifier in MATLAB

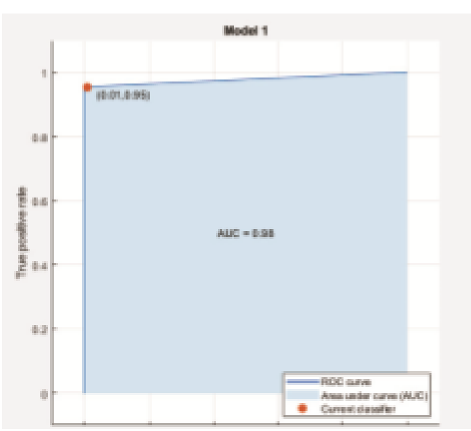


Fig. 5. Screenshot for ROC curve of the tree classifier model

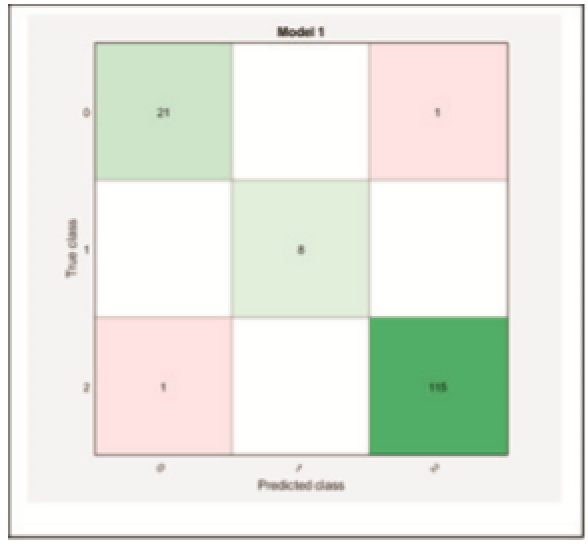


Fig. 6. Screenshot for confusion matrix of atree classifier

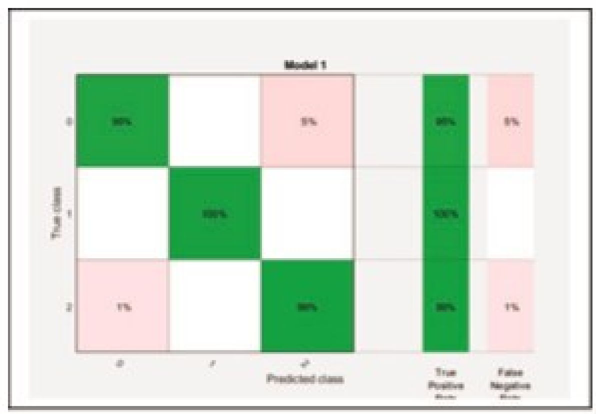


Fig. 7.Screenshot for true positive and false negative rates of tree Classifier

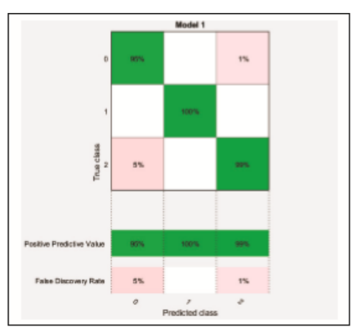


Fig. 8.Screenshot for positive predictive and false discovery values of tree Classifier

Depending on the data import type, a significant variation in accuracy was seen. The model could train using 26 characteristics and the tree classifier achieved a 93.8% accuracy rate when the data was loaded as a table. The same tree classifier achieved a 98.6% success rate when trained using data imported as a numeric matrix, which allowed for the use of all 28 features. Importing data in the correct format is therefore equally crucial when doing data categorization. According to the statistics in the table, the Tree classifier achieves the highest prediction accuracy of 98% when applied to this dataset. Minimal accuracy is produced using SVM classifier. By properly predicting the severity of illness in advance, classifiers may help patients with vitamin D deficiency avoid consequences like surgery and hospitalization.

VII. CONCLUSION

In preventative healthcare, AI has the potential to greatly help humans by allowing clinicians to base treatment choices on machine learning. Vitamin D levels are crucial in the treatment of inflammatory bowel disease (IBD), as is evident from prior research [1]. Vitamin D supplements may be administered to people at high risk of inflammatory bowel disease (IBD) in advance to prevent further problems. In order to train and assess the open dataset, this study makes use of several machine learning classification techniques. This dataset yields the lowest accuracy from the SVM classifier. The Decision Tree classifier has an area under the ROC curve of 0.98 and can correctly predict 96 data points for this dataset. Future studies may build on the paper's findings by expanding the dataset. The data might be enhanced with additional characteristics, allowing for comparison of categorization results. Additionally, performance may be evaluated by extending classifier fusion.

REFERENCES

[1] "The Facts About Inflammatory Bowel Diseases.," NY: Crohn's and Colitis Foundation of America, New York, 2014.

[2] A. A. N., "Vitamin D and Inflammatory Bowel Disease.," Gastroenterology & hepatology, vol. 12, no. 8, p. 513–515., 2016.

[3] M. L. F. J. N. B. Veit L, "The Vitamin D Status in Inflammatory Bowel Disease," PLoS One, vol. 9, no. 7, p. e101583, (2014).

[4] H. Hacılar, O. U. Nalbantoğlu and B. Bakir-Güngör, "Machine Learning Analysis of Inflammatory Bowel Disease-Associated Metagenomics Dataset," 2018 3rd International Conference on Computer Science and Engineering (UBMK), Sarajevo, 2018, pp. 434-438.

[5] N. Lower and F. Zhan, "A Study of Ensemble Methods for Cyber Security," 2020 10th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2020, pp. 1001-1009.

[6] Mossotto, E., Ashton, J.J., Coelho, T. et al. Classification of Paediatric Inflammatory Bowel Disease using Machine Learning. Sci Rep 7, 2427 (2017). Edoardo Pasolli et al., "Machine learning metaanalysis of large metagenomic datasets: tools and biological insights", PLoS computational biology, vol. 12.7, pp. el004977, 2016.

[7] Mark Hall et al., "The WEKA data mining software: an update", ACM SIGKDD exploratio ns newsletter, vol. 11.1, pp. 1018, 2009.

[8] M Hanchuan Peng, Long Fuhui, Ding Chris, "Feature selection based on mutual information criteria of max-dependency max-relevance and min-redundancy", IEEE Transactions on pattern analysis and machine intelligence, vol. 27.8, pp. 1226-1238, 2005.

[9] Henderson, P. et al. Rising incidence of pediatric inflammatory bowel disease in Scotland. Inflamm. Bowel Dis. 18, 999–100