ISSN: 2347-467X, Vol. 12, No. (3) 2024, Pg. 1284-1299

Current Research in Nutrition and Food Science

www.foodandnutritionjournal.org

Development and Performance Analysis of Machine Learning Methods for Predicting the Occurrence of Constipation and its Risk Factors Among College-aged Girls

JOYETA GHOSH[1]* and POULOMI SANYAL[2]

[1]Department of Dietetics and Applied Nutrition, Amity University Kolkata, Kolkata, India.
[2]Department of Dietetics and Nutrition, NSHM Knowledge campus, Kolkata, India.

Abstract

The present study sought to determine which model was most useful for predicting functional constipation (FC) in college-aged students by examining the applicability of multiple models and evaluating the forecasting accuracy of prediction methods, including regression-based models and machine learning models. This observational descriptive study involved 300 college girls from Kolkata, West Bengal, India, who were randomly chosen using social media (Linkedin,WhatsApp and Face book) and ranged in age from 18 to 25 years. The survey was carried out using an online, standard questionnaire that had been pre-tested. The obtained data were entered into a Microsoft Excel Worksheet (Redwoods, Washington, USA: Microsoft) and reviewed for elimination errors.19 attributes were selected for prediction study. Weka version 3.8.0 software was used for predictive modeling, performance analysis, and the building of FC prediction system. The predictive models were then developed and contrasted using 5 different models as a classifier. We divided our data into training and test datasets, which comprised 70% and 30% of the total sample, respectively, at random for each investigation. Out of 300 occurrences, 96.00 % were correctly classified, while only 4 % were wrongly classified, with a Kappa value of 0.875, and a root mean squared error of 0.19. The model's accuracy was 96.3% weighted precision, 96% true positives, 0.05% false positives, 0.961 F measure, and 0.994ROC(receiver operating characteristic curve). Here 6 different evaluators were used and surprisingly they all predict Bristol's Stool consistency Scale as the number 1 predictor of FC among college girls. Again ‘Pain and discomfort in abdomen’ remains second predictor according to all selected evaluators. Thus, it can be confirmed that ‘Bristol's Stool

Article History
Received: 29 April 2024
Accepted: 07 August 2024

Keywords
Constipation;
College Going Girl;
Machine Learning;
Young Adults.

CONTACT Joyeta Ghosh joyetaghosh01@gmail.com Department of Dietetics and Applied Nutrition, Amity University Kolkata, Kolkata, India.

Doi: https://dx.doi.org/10.12944/CRNFSJ.12.3.23

consistency Scale' and the 'Pain and discomfort in abdomen' are the two significant predictor of FC among college going girls. This machine learning model-based automated approach for predicting functional constipation will assist medical professionals in identifying younger generations who are more likely to experience constipation. Additionally, predictions can be made quickly and efficiently using sociodemographic and morbidity parameters. For further follow-up and care, at-risk patients can be referred to consultant physicians. This will lessen the burden of gastrointestinal-related morbidity and mortality among the younger population.

Abbreviations

Functional constipation (FC), Irritable bowel syndrome (IBS), Artificial intelligence(AI), Machine learning (ML), Body Mass Index (BMI), Receiver Operating Characteristic Curve(ROC)

Introduction

Functional constipation (FC) is a functional gastrointestinal condition that is clinically common and affects both adults and children worldwide.[1, 2] Chronic constipation (FC, sometimes called chronic idiopathic constipation) is characterized by failure or disturbance in the physiological activities of feces, with the exception of irritable bowel syndrome (IBS).[3-5] Meta-analyses revealed that FC was endemic in a variety of countries, and the prevalence varied across different cross-sectional surveys (i.e., different regions).[6] The average global prevalence of constipation, as determined by national and international surveys, was 16% (ranging from 0.7% to 79%),[7] with 10.1% of cases of FC identified by the Roman III criteria.[6] College students are one of the communities most affected by functional constipation (FC). Prolonged constipation can lead to many health problems such as intestinal blockage, hemorrhoids, fissures, and irritability, which can affect academic performance and overall well-being.[9–11]

Existing research has demonstrated that sedentary lifestyles, certain dietary practices, such as consuming little fruit and vegetable fibre, drinking insufficient amounts of water, and having low educational attainment are factors in the rising incidence of FC.[7,12,13] Apart from the recognized variables already discussed, there may be a relationship between sleep quality and the prevalence of FC. According to a World Health Organisation survey, 27% of people worldwide have sleep issues.[14] 77% of college students said they had difficulty sleeping during the preceding 12 months, according to China University's "Students Health Survey." Sleep problems have a direct effect on the lives of individuals.Recent research has shown that persons with gastrointestinal conditions or symptoms frequently experience sleep issues.[15] There are now some techniques or tools available for predicting the occurrence of FC, however, research on predicting the likelihood of FC in college students or the younger generation is still in its infancy. Artificial intelligence (AI) has the potential to completely change the way that clinical decision-making is done in the field of health sciences.[16,19,20] More specifically, AI can assist with the proactive and objective assessment of health symptoms to support diagnosis and therapy delivery that is suited to the needs of each patient, including long-term monitoring and care management. Consequently, machine learning (ML) techniques have been proposed in an effort to increase the accuracy of constipation forecasting, in addition to the use of more conventional regression techniques.

Most existing studies on FC rely on traditional statistical methods rather than advanced machine learning techniques, limiting predictive accuracy.[7,12,13] FC presents with diverse symptoms, making it challenging to develop a one-size-fits-all predictive model. Many studies on FC prediction are limited by small sample sizes, affecting the generalizability of models. Existing models often struggle to effectively integrate diverse factors like diet, lifestyle, and physiological parameters into a single predictive framework.[7,12,13] On this background the primary objective of this study is to develop and implement a novel machine learning model for the diagnosis of FC among young adults.In order to determine the most effective model for predicting FC in young adults, we examined the applicability of a number of models

and evaluated the forecasting accuracy of prediction techniques, such as regression-based models and machine learning models. To date, no prediction-based model has been established for the diagnosis of FC. This study represents the first of its kind, pioneering the development and implementation of a machine learning approach for FC diagnosis. Such development of a predictive model for FC among young adults offers significant benefits for early identification and management of the condition. By incorporating the standard ROME III classification criteria alongside dietary habits, lifestyle factors, and nutritional parameters, these models will provide more comprehensive approach to FC prediction. This integration of diverse factors allows for a more personalized assessment, tailoring predictions to individual characteristics and behaviors. Such a model not only enhances our understanding of FC in the younger population but also paves the way for more targeted and effective interventions.

Materials and Methods

Data Source

Sample Size Calculation

Sample size was calculated by taking the previous prevalence of self reported constipation was 24.8%[17] and using formula $n = 4pq/L^2$ (where,p = prevalence of malnutrition, q = 100−p, L = 15% of p).[18] It came out to be 539. During the study period, 300 participants agreed to participate in the study based upon their availability, willing to participate, and according to exclusion criteria. Purposive sampling techniques were implemented during selection of the study participants.

Study Design

This observational descriptive study involved 300 college girls in Kolkata, West Bengal, India, who were randomly chosen using social media(LinkedIn, WhatsApp and Facebook) and ranged in age from 18 to 25.The traditional evaluation was already published elsewhere,[21] present study is the updated version of the same, where ML models were applied and Bristol's and ROME-III criteria were incorporated. The study was extended part of one bigger project among adult and institutional ethical clearance was obtained from All India Institute of Hygiene and Public Health, Kolkata, India. Undergraduates and postgraduates in their first through fifth years of study who willingly participated and gave their informed consent at the age of 18 met the following inclusion criteria. The following were the exclusion requirements: (1) individuals who suffer from persistent cardiovascular, hematologic, or digestive problems; (2) those who have substantial lesions in other organs. The study took place between February 2022 to May 2022. The survey was carried out using an online, standard questionnaire that had been pre-tested. The obtained data were entered into a Microsoft Excel Worksheet (Redwoods, Washington, USA: Microsoft) and reviewed for elimination errors.

Selection of Predictors

Students who are about to attend college are at a critical juncture in both their personal and academic development as well as a special transitional period between campus and community life. Unavoidably, they will encounter both positive and negative life events, sometimes referred to as stressful life events, which will affect their emotions in different ways.[22] Certain stressful life events, like poor exam results, arguments with close friends, the end of a romantic relationship, and extended separation from family, are more likely to cause negative emotions and behaviors in college-bound students.[23] These behaviors can lead to gastrointestinal dysfunction and increase the risk of FC. Eighteen variables were included as potential predictors in the current investigation. The predictors were developed using many published studies on FC in young people or college-bound students.[21,24]

Age, BMI, and the five criteria for the ROME-III categorization of FC (the last two were taken because the question of "Manual maneuvers on >25% of defecations" received no response and was therefore removed from the data set) Too-small bowel movement frequency, bleeding or tearing in the colon during or following a bowel movement, abdominal pain and discomfort, Bristol's Scale of Stool Consistency, Exercise Frequency: "Do you exercise?" Hours per day when you sleep, How frequently you eat fruit, how often you eat leafy green vegetables, Daily Water Consumption (L), The 19 characteristics that have been chosen to predict constipation are related to eating behaviors (Table 1). As one of the ROME III criteria has common respond to all the participants which is 'no' or 'absent', therefore this attribute was excluded from the analysis part, hence total 18 attributes were selected.

Table1: Explanation of Questionnaires used as a tool

SL No.	Questions	Attributes used in dataset	Answer
1	Age	Age	Age in number
2	BMI status of the participants as per the measurements?	BMI	(0=Normal BMI,1=Malnutrition)
3	What type of food habit participants has?	Food Habit	(0=Vegetarian,1=Non vegetarian)
4	How much water participant took daily?	Daily Water Intake	(0=>1.5 litr,1=<1.5 litr)
5	What is the frequency of too small bowel movement?	Too-small bowel movement frequency	1 = Mild to Moderate; 0 = Absent.
6	Did the individual have tear or bleeding in the rectal area during or after a bowel movement?	Crying or tearing in the rectal cavity during or after a bowel movement	(0=Absent,1=Mild to Moderate)
7	Did the participant feel any abdominal pain or discomfort?	Abdominal pain and discomfort	(0=Absent,1=Mild to Moderate)
8	What type of stool the participants have according to Bristol's Stool consistency Scale?	Bristol's Stool consistency Scale	(0=Type 1,1=Type 2, 2=Type 3, 3=Type 4, 4=Type 5, 5=Type 6, 6=Type 7)
9	What is the frequency of exercise per week?	Frequency of exercise	(0=No exercise, 1=1 to 2 days per week, 2=3 to 4 days per week, 3=every day)
10	Does the participant do physical exercise without work?	Do you exercise?	(0=No,1=Yes)
11	What are the daily sleeping hours of the participants?	Daily sleeping hours	(0=>6 hours per day, 1=<6 hours per day)
12	What are the frequencies of fruit consumption?	Frequency of fruit consumption	(0=daily, 1= Once/ week, 2= Less than three times/ week, 3=rarely)
13	How frequently do you eat lush green vegetables?	Consumption frequency of leafy green vegetables	(0=every day, 1 = once per week, 2 = less than three times per week, and 3 = seldom)
ROME III Criteria			
14	Restricting on more than 25% of feces	Limiting more than 25% of bowel movements	(0=Absent,1=Present)
15	Incomplete evacuation is perceived in more than 25% of defecations	In more than 25% of cases, there is a sense of incomplete evacuation	(0=Absent,1=Present)

16	Over 25% of defecations result in lumpy or hard stool.	More than 25% of bowel movements result in a lumpy or hard stool.	(0=Absent,1=Present)
17	Sensation of anarectal obstruction/ blockage on >25% of defecations	Sensation of anarectal obstruction/blockage on >25% of defecations	(0=Absent,1=Present)
18	Less than 3 defecation per week per week	Less than 3 defecation	(0=Absent,1=Present)

Statistical Analysis and Disease Prediction System

The research technique or step is shown in Figure 1. The full research methodology is outlined in the order and detail that follows. It gives the working flowchart a clear comprehension.

The association between two qualitative data was calculated by Pearson's Chi-square test and 'P' value was determined. All the statistical analysis was performed by SPSS software (Statistical Package for Social Sciences version 20.0). 'P' value is equal to or less than 0.05 was considered as statistically significant.

Model Construction and Evaluation

In present study Weka version 3.8.0 software was used for predictive modeling, performance analysis, and the building of a FC prediction system. Weka is free software for data mining in the area of machine learning.[25,26] Additionally, Python 3.7 version was used for the exploratory data analysis and visualization. A training data set for data mining in Weka was created using primary data obtained by interviewing the 300 respondents that were chosen. The balancing methods that were applied include SMOTE (for oversampling), Spread Sub Sample (for under sampling) and a combination of SMOTE and Spread Sub sample. The default parameters were used for all the methods except for Spread Sub Sample where the distribution spread, we set as 1.0. The application of these methods alters the number of instances in the training dataset. The predictive models were then developed and contrasted using Weka's 5 different models as a classifier. We divided our data into test and training datasets, which comprised 70% and 30% of the total sample, respectively, at random for each investigation. Furthermore, to avoid over fitted or optimistically biased performance estimates, we applied repeated layered cross-validation, as advised by previous studies.[27] Based on how effectively each predictor contributed to prediction accuracy, we also assessed the relative significance of each predictor for each classifier.

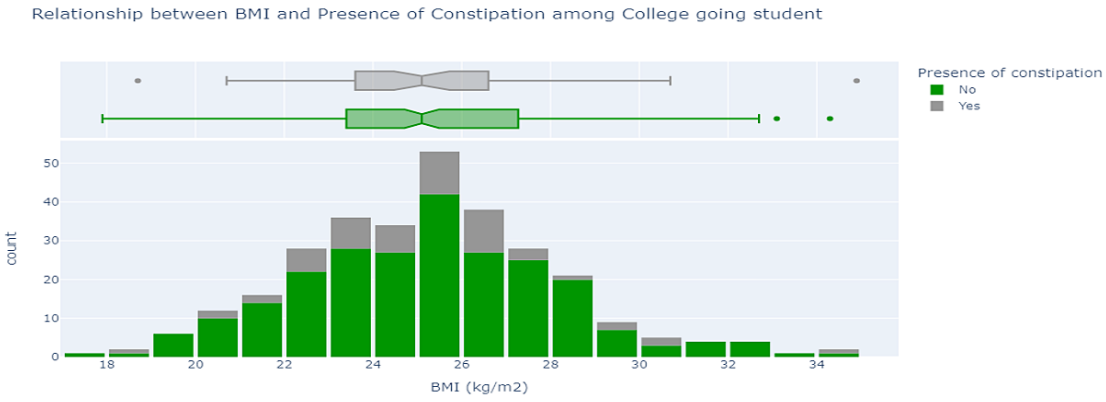

Fig.1: Distribution of BMI and Constipation among targeted respondents (N=300)

Results

This descriptive observational study examined the nutritional, clinical, and other contributing factors to functional constipation in college going youth of India in addition to determining the normal bowel pattern.

Table 2: Distribution of the lifestyle, food habit and nutritional status of the respondents and its association with constipation

Parameters	Participants (%)	Constipation Present (%)		Chi-square test (p value)
	N=300	Yes	No	
Food habit:				
Vegetarian	5(1.67%)	1(1.75%)	56(98.24%)	0.003
Non-vegetarian	295(98.33%)	4(1.64%)	239(98.35%)	(p-0.65)
Nutritional Status (BMI)				
Normal	1(0.33%)	-	1(.41%)	0.369
Underweight	63(21%)	11(19.29%)	52(21.39%)	(p-0.83)
Overweight / Obesity	236(78.67%)-	46(80.70%)	190(78.18%)	
Skipping breakfast:				
None	237(79%)	42(73.68%)	195(80.24%)	1.119
1-3 days	63(21%)	15(26.31%)	48(19.75%)	(p-0.17)
3-6 days	-	-	-	
Everyday	-	-	-	
Daily water intake:				
<1.5L	36(12%)	50(87.71%)	29(11.93%)	136.05
>1.5L	264(88%)	7(12.28%)	214(88.06%)	(p-0.00)
Frequency of green leafy vegetable consumption:				
Daily	15(5%)	-	15(6.17%)	102.870
Once per week	183(61%)	17(29.82%)	166(68.31%)	(p-0.00)
<3times per week	47(15.67%)	3(5.26%)	44(18.10%)	
Very rarely	55(18.33%)	37(64.91%)	18(7.40%)	
Frequency of fruit consumption:				
Daily	30(10%)	2(3.50%)	28(11.52%)	3.829
Once per week	175(58.33%)	36(63.15%)	139(57.20%)	(p-0.02)
<3times per week	71(23.67%)	13(22.80%)	58(23.86%)	
Very rarely	24(8%)	6(10.52%)	18(7.40%)	
Daily sleeping pattern:				
≤6 hours	161(53.67%)	29(50.87%)	132(54.32%)	0.220
7-8 hours	139(46.33%)	28(49.12%)	111(45.67%)	(p-0.37)
Frequency of exercise perweek:				
Daily	-	-	-	0.917
5-6 days/week	-	-	-	(p-0.63)
3-4 days/week	40(13.33%)	6(10.52%)	34(13.99%)	
1-2 days/week	11(3.67%)	3(5.26%)	8(3.29%)	
Rarely	249(83%)	48(84.21%)	201(82.71%)	

Training Model and its Performance Analysis

Machine learning has a variety of classifiers at its disposal. These classifiers are all used to construct particular machine learning-based systems. Every classifier has a unique implementation. A good classifier is essential to a machine learning-based model's effectiveness. Each classifier has some benefits and drawbacks. The classifiers' accuracy

varies depending on the approach, types of data, and dataset. Every classifier offers a variable level of accuracy for various techniques and datasets. Finding a proper classifier for a particular model is a crucial task in machine learning. In present study the 10-fold cross validation method was used to train and evaluate the 5 different models on the primary data set. At the end 3 best model were chosen. The results showed in Table 3.The best fitted model was Naïve Bayes Multinomial Classifier. Out of 300 occurrences, 96.00 % were correctly classified, while only 4 % was wrongly classified, with a Kappa value of 0.875, a root mean squared error of 0.19. The model's accuracy was 96.3% weighted precision, 96% true positives, 0.05% false positives, 0.961 F measure, and ROC was 0.994.

The Naïve Bayes Multinomial Classifier was chosen as the fitted model due to its ability to handle the complex nature of Functional Constipation (FC) diagnosis effectively. The Naïve Bayes algorithm was chosen for its effectiveness in handling categorical data, which is prevalent in medical diagnostics, and its ability to perform well with relatively small datasets.

Table 3: Application of different classification model and their performance in predicting FC

Classification model applied	Correctly classified Instances (%)	Incorrectly classified instances (%)	Kappa Statistics	Root Mean Squared Error	Relative Absolute Error	True Positive Rate	False Positive Rate	Precision	F-Measures	ROC Area
Naïve Bayes Multinominal Classifier	96.0	4.0	0.875	0.199	13.67	0.960	0.05	0.963	0.961	0.994
Large margin Classification using perception algorithm	88.33	11.66	0.513	0.341	38.00	0.883	0.484	0.892	0.863	0.771
Randomizable Filtered Classifier	89.66	10.33	0.626	0.320	34.35	0.897	0.333	0.892	0.890	0.783

The selection of algorithms was guided by the following considerations:

1. Nature of the data: The mix of categorical (e.g., ROME III criteria) and continuous (e.g., dietary intake) variables in FC diagnosis necessitated algorithms capable of handling diverse data types.
2. Interpretability: Naïve Bayes offers good interpretability, which is crucial in medical applications where understanding the reasoning behind predictions is important.
3. Performance with limited data: Given the challenges in obtaining large medical datasets, algorithms that perform well with smaller sample sizes were prioritized.
4. Ability to handle multiple predictors: FC involves various factors, requiring algorithms capable of integrating multiple predictors effectively.
5. Balancing accuracy and computational efficiency: The selected algorithms offer a good balance between predictive accuracy and computational demands.

Table 4: Distribution of Attributes as per their ranking using different evaluators

Attributes Ranking	Attributes Evaluator					
	Symmetrical Uncertainty Ranking Filter	**Relief Ranking Filter**	**Gain Ratio Feature Evaluator**	**One R Feature Evaluator**	**Correlation Ranking Filter with Correlation Values**	**Information gained ranking filter**
1	Bristol's Stool consistency Scale	Bristol's Stool consistency Scale	Bristol's Stool consistency Scale	Bristol's Stool consistency Scale	Bristol's Stool consistency Scale (1)	Bristol's Stool consistency Scale
2	Pain and discomfort in abdomen	Pain and discomfort in abdomen	Pain and discomfort in abdomen	Pain and discomfort in abdomen	Pain and discomfort in abdomen (0.9122)	Pain and discomfort in abdomen
3	Straining on >25% of defecations	Straining on >25% of defecations	Straining on >25% of defecations	Sensation of incomplete evacuation on >25% of defecations	Sensation of incomplete evacuation on >25% of defecations (0.7301)	Sensation of incomplete evacuation on >25% of defecations
4	Less than 3 defec-ation per week	Frequency of green leafy vegetable consumption	Less than 3 defecation per week	Sensation of anarectal obstruction/blockage on >25% of defecations	Frequency of too small bowel movement (0.7299)	Lumpy or hard stool on >25% of defecations
5	Frequency of too small bowel movement	Daily sleeping hours	Frequency of too small bowel movement	Less than 3 defecation per week	Sensation of anarectal obstruction/blockage on >25% of defecations (0.7293)	Sensation of anarectal obstruction/blockage on >25% of defecations
6	Sensation of anare -ctal obstruction/ blockage on >25% of defecations	Do you exercise	Sensation of anarectal obstruction/blockage on >25% of defe-cations	Lumpy or hard stool on >25% of defecations	Lumpy or hard stool on >25% of defecations (0.7293)	Less than 3 defecation per week
7	Lumpy or hard stool on >25% of defecations	Frequency of exercise	Lumpy or hard stool on >25% of defecations	Frequency of too small bowel movement	Less than 3 defecation per week (0.71162)	Frequency of too small bowel movement
8	Sensation of inco -mplete evacuation on >25% of defecations	Sensation of inco -mplete evacuation on >25% of defecations	Sensation of incom-plete evacuation on >25% of defecations	Straining on >25% of defecations	Straining on >25% of defecations (0.7009)	Straining on >25% of defecations

9	Frequency of green leafy vegetable consumption	Lumpy or hard stool on >25% of defecations	Rectal bleeding or tearing during or after bowel movement	Frequency of green leafy vegetable consumption	Frequency of green leafy vegetable consumption (0.5023)	Frequency of green leafy vegetable consumption
10	Rectal bleeding or tearing during or after bowel movement	Sensation of anarectal obstruction/blockage on >25% of defecations	Frequency of green leafy vegetable consumption	Rectal bleeding or tearing during or after bowel movement	Rectal bleeding or tearing during or after bowel movement (0.4028)	Rectal bleeding or tearing during or after bowel movement
11	BMI	Less than 3 defecation per week	BMI	BMI	Age (0.0784)	Food Habit
12	Food Habit	BMI	Food Habit	Daily water intake	Frequency of fruit consumption (0.0686)	BMI
13	Daily sleeping hours	Frequency of too small bowel movement	Daily sleeping hours	Daily sleeping hours	Frequency of exercise (0.0571)	Daily sleeping hours
14	Daily water intake	Frequency of fruit consumption	Daily water intake	Daily water intake	Daily sleeping hours (0.02709)	Daily water intake
15	Frequency of exercise	Daily water intake	Do you exercise	Frequency of fruit consumption	BMI (0.01095)	Frequency of exercise
16	Do you exercise	Age	Frequency of exercise	Frequency of exercise	Do you exercise (0.01561)	Do you exercise
17	Frequency of fruit consumption	Rectal bleeding or tearing during or after bowel movement	Frequency of fruit consumption	Age	Daily water intake (0.0096)	Frequency of fruit consumption
18	Age	Food Habit	Age	Do you exercise	Food Habit (0.0032)	Age

These algorithms, particularly the Naïve Bayes Multinomial Classifier, address the research question by enabling the integration of diverse FC indicators into a cohesive predictive model. This approach allows for a more comprehensive and personalized diagnosis of FC, potentially improving upon traditional diagnostic methods that may not fully capture the multifaceted nature of the condition.

Analysis of Attributes Ranking

Table 4 represents the attributes ranking status of the study. Here 6 different evaluators were used and surprisingly they all predict Bristol's Stool consistency Scale as the number 1 predictor of FC among the college going girls. Again 'Pain and discomfort in abdomen' remains second predictor according to all selected evaluators. Thus it can be confirmed that 'Bristol's Stool consistency Scale' and the 'Pain and discomfort in abdomen' are the two significant predictor of FC among college going girls. Furthermore 'Straining on >25% of defecations' was predicted as third important risk factors of FC by Symmetrical Uncertainty Ranking Filter, Relief Ranking Filter,& Gain Ratio Feature Evaluator. Whereas, 'More than 25% of defecations, have a sense of incomplete evacuation' is ranked as third predictor by other three evaluator named: One R Feature Information obtained ranking filter, correlation ranking filter, and evaluator. According to the Symmetrical Uncertainty Feature Evaluator with Gain Ratio and Ranking Filter, "Less than 3 defecation per week" was the fourth important predictor, whereas the Relief Ranking Filter suggests that "Frequency of green leafy vegetable consumption" is the fourth major predictor of FC. Again '>25% of defecations have the sensation of an anarectal obstruction or blockage.' is considered as 4th predictor by One R Feature Evaluator. 'Frequency of too small bowel movement' is predicted by Correlation Ranking Filter. 'Lumpy or hard stool on >25% of defecations' is predicted by Information gained ranking filter as 4th predictor.

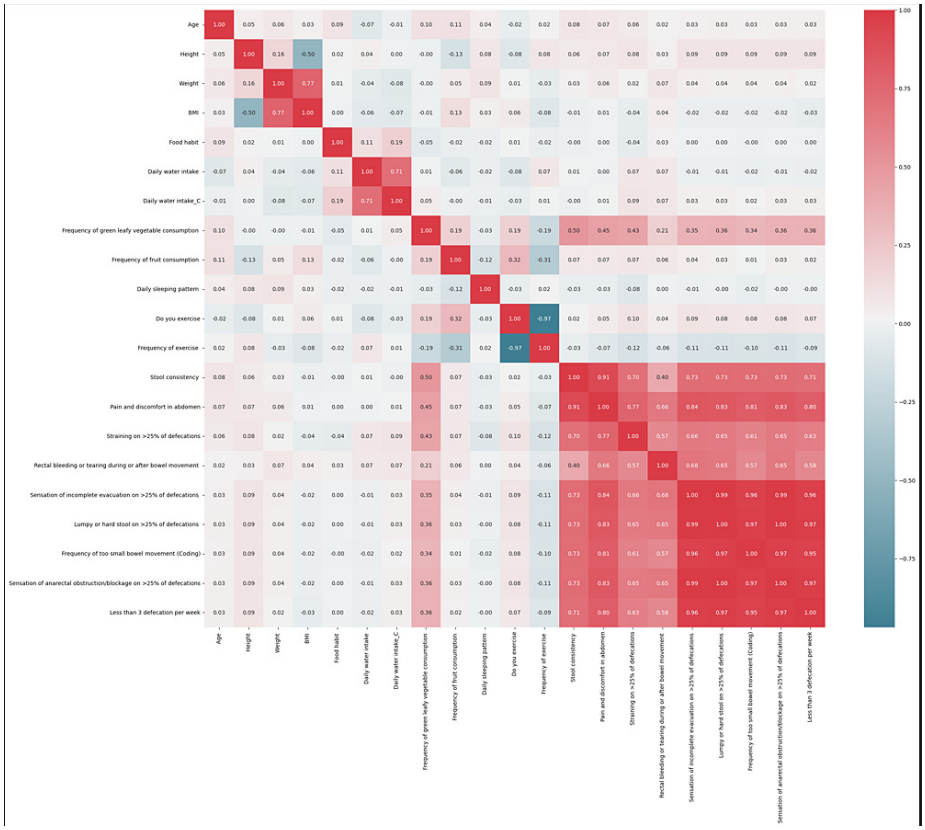

Fig.2: Generated heat map showing correlation value between different attributes of Constipation (FC)

Discussion

The mean age of the targeted adult population was 21.65 ±1.53 years. The targeted respondents' mean weight was 64.72 kg and their average height was 160.49± 5.89 cm; their average BMI was also 25.22 kg/m^2. A total of 300 responders were present, and 19% of them reported having constipation (Table 2, Figure 1). Based on the anthropometric evaluation, 0.33% of the participants were underweight, 21.67% were overweight, 46.33% were obese in the first grade, and 10.67% were obese in the second grade. Once more, just 19.29% of those who participated in this survey were classified as normal, and 80.70% as malnourished. Among the responders, 98.33% of people were Non-vegetarian, while 1.67% was vegetarian. As one of the most prevalent gastroenterological conditions, FC can be addressed by making specific lifestyle changes, such as increasing the amount of fibre and water consumed each day. This fibre aids in improved bowel movements by acting as a bulking agent and aiding in the binding of water.[28] Numerous risk factors may contribute to the cause of constipation. They consist of sociodemographic (female gender), lifestyle (physical activity), and medical (surgery, specific drugs) aspects.[29] According to the respondents' daily water consumption, 12% of them drank less than 1.5 litres per day, while 88% drank more than 1.5 litres per day. Out of these, 12.28% drank more than 1.5L of water each day, compared to 87.71% of the constipation sufferers. Daily water intake is ranked 14th in the ranking attributes by 4 evaluators (Table 4).According to Rajput and Saini (2014),[17] 51.4% of people who consumed less fluid had constipation. Drinking water significantly impacts defecation frequency, stool type, the presence of blood in stools, and the likelihood of obstruction.[29] A report found that women with the highest dietary fiber intake (median intake: 20 g/day) were less likely to experience constipation compared to those with the lowest fiber intake.[32] Regarding the amount of green leafy vegetables that the respondents ate, 5 percent did so every day, 61% did so once a week, 15.67% did so three times a week or less, and 18.33% did so very infrequently. A sedentary lifestyle is also a contributory factor to functional constipation (FC) and poor heart health.[30] Of them, 64.91% of the constipation adults ate leafy green vegetables only infrequently, 5.26% consumed them three or more times a week, and 29.82% consumed them just once. In addition to aiding with constipation, leafy greens also support proper brain functioning, which is necessary for the focused demographic.[33] Consumption of green leafy vegetable is one of the important predictor (ranked 4th) according to Relief Ranking Filter(Table 4).Overall, a healthy diet not only aids in relieving constipation but also has a considerable positive impact on immunological function.[34] Young adults who consumed more whole grains, rice/pasta, and vegetables had a reduced rate of constipation.[35] Increasing dietary fiber intake by about one gram per day could help reduce healthcare expenses related to constipation.[36] According to certain studies,[17,29,37] a non-vegetarian diet is associated with a higher prevalence of constipation. Patients self-managed the majority of cases of constipation; 22% sought medical intervention, mostly from primary care physicians (>50%) and gastroenterologists (14%); this led to substantial costs for diagnostic tests and treatment.[38] Furthermore, the development of healthy eating habits depends on these mindful eating techniques .39, 40, 41 Of all the respondents, 53.67% reported sleeping for 5 to 6 hours each day, while 46.33% reported sleeping for 7 to 8 hours each day. Of those, 54.32% slept for 5 to 6 hours each night while the remaining 49.12% slept for 7 to 8 hours each night. Researcher noted that the signs and effects of constipation may vary from patient to patient as well as depending on the age group.[42] Women who engage in daily moderate exercise have a 44% lower risk of constipation compared to those with poor bowel movement rates.[32] Middle-aged patients with persistent constipation experienced improved defecation patterns with regular physical activity.[43] According to a study done on teenagers, constipation was linked to both excessive sedentary behavior and a lack of moderate-to-vigorous exercise.[37] In terms of daily activity, 17% of respondents reported exercising, compared to 83% who reported not exercising. Out of those, 15.78% of the constipation sufferers exercised, while the remaining 84.28% did not. Now, if we take into account how often respondents exercised each week, we observed that 13.33% of them exercised three to four days per week, 3.67% exercised one to two days per week, and 83% did not exercise at all. Among them, 84.21% of the constipation sufferers exercised infrequently, 5.26% exercised once or twice a week, and 10.52% exercised three or four times a week. Exercise and its frequency are still

considered by some evaluators to be the 15th best predictors (Table 4).

According to the Bristol's Stool consistency Scale, type 1 stool consistency, which indicates severe constipation, was reported by 1.5% of respondents. 17.5% of the participants reported type 2 stool consistencies and mild constipation. As a result, it was anticipated that 19% of participants in the current study experienced constipation overall. In 24.5% and 33% of the participants, respectively, the type 3 and type 4 of normal stool consistency was reported. Bristol's Stool consistency Scale ranked 1st by all the evaluators (Table 4). A 2016 report found that constipation affected 16.2% of students, with women being more likely to experience it (17.4%) compared to men (12.5%).Constipation was common in Asia, affecting 15% to 23% of women and 11% of males.[11] Bristol stool types 1 or 2 were found in 20.5% of people in a cross-sectional survey of the urban South Indian population, while types 3 and 4 were found in 35.6% and 32.5% of participants, respectively.[45]

Similarly, 82% of the participants said they had no stomach pain or discomfort, 15% said they had mild pain, and 3% said they had moderate pain or discomfort. Of these, 5.26 percent of people who had constipation reported having no abdominal discomfort or pain, 78.94 percent reported having small amounts of pain or unease, and 15.78% reported having substantial ache or distress. In table 4 'Pain and discomfort in abdomen' is the 2nd ranking predictor confirmed by all selected evaluator. When it comes to rectal burning, it was shown that 89.67% of respondents did not experience it and that 10.33% only experienced minor rectal burning. Out of those, 54.38% reported only minor rectal burning, compared to 45.61% of the constipation patients who had no rectal burning at all. While thinking 96.33% of respondents reported having no rectal bleeding or ripping, while 3.67% reported having only light bleeding or tearing. Out of those, 19.29% had just little rectal bleeding or tears, whereas 80.70% of the constipation patients did not experience any of either. 82.67% of respondents had no incomplete evacuation, 13.67% had mild incomplete evacuation, and 3.67% had moderate incomplete evacuation when incomplete evacuation was taken into account. Out of those, 56.14% had mild constipation, 19.29% had significant constipation, and the remaining 24.56% did not have incomplete evacuation. Overall the factors of ROM-III criteria ranked top position by all the selected evaluators. The heat map in Figure 2 shows the relationship between several Constipation (FC) characteristics, demonstrating once more how strongly the ROME-III criteria are connected with both FC and one another. Interestingly the green leafy consumption has correlation with ROME-III criteria, which is also visible. 82.67% of the people who responded reported no hardening of the stool, 14% reported mild hardening, and 3.33% reported moderate hardening. Out of those, 57.89% had mild constipation, 17.54% had significant constipation, and the remaining 24.56% did not experience hardness of the stool. Now, considering about how often people have too few bowel movements 82.67% of the participants reported having enough bowel movements, compared to 15.67% who reported mild and 1.67% who experienced too few. Out of those, 66.67% of the constipation sufferers experienced mild symptoms, 8.77% experienced moderate symptoms, and 24.56% of the other people did not encounter "too small bowel movement". Regarding the pushing or straining during defecation: 82.67% of participants reported no straining or pressing, 14% did so in a mild manner, and 3.33% did so in a moderate manner. Out of those, 57.89% experienced mild constipation, 17.54% had significant constipation, and the remaining 24.56% had no constipation at all. Again, when it came to frequency of daily feces, 82.67% of participants excreted stool once day, whereas14% two times daily, and 3.33% three times daily. Among them, 24.56% of the individuals with FC defecated just one times, while 59.64% defecated two times, and 15.78% defecated three times per day.

Limitation of the Study

In present study the models developed for specific demographics may not perform well across different age groups or ethnicities, necessitating more robust, cross-population validation. Current models typically rely on retrospective data, with few capable of real-time prediction of FC onset or progression, an area that requires further development. Low sample size is another concern in this case.

Conclusion

The study found that 19% of participants reported constipation, with a high prevalence of overweight and obesity (78.67%) among respondents. Key

findings include: (a) Bristol Stool Scale was identified as the top predictor for constipation. (b) Abdominal pain/discomfort and other Rome III criteria were strongly associated with constipation. (c) Green leafy vegetable consumption showed a correlation with constipation symptoms. (d) Water intake and physical activity levels were considered relevant factors, though ranked lower in predictive value. The study highlighted the importance of lifestyle factors in managing constipation, including dietary fiber intake, hydration, and exercise. It also emphasized the potential for self-management and the economic impact of constipation on healthcare systems.

The study used a 10-fold cross-validation method to train and evaluate 5 different models on the primary dataset. The Naïve Bayes Multinomial Classifier emerged as the best-fitted model among these. This model shows a good balance across various performance metrics, indicating its robustness in classifying instances of functional constipation among college girls. The high accuracy, low error rate, and strong performance across multiple metrics suggest that the Naïve Bayes Multinomial Classifier is well-suited for this particular dataset and classification task, making it the best choice for predicting functional constipation in this study.

These findings suggest the need for targeted interventions focusing on dietary habits, physical activity, and awareness of constipation symptoms among young adults. Future research could explore the long-term effects of lifestyle modifications on constipation prevalence and severity in this demographic, as well as the potential benefits of early intervention in preventing chronic constipation. This machine learning model-based automated approach for predicting functional constipation will assist medical professionals in identifying younger generations who are more likely to experience constipation. Additionally, predictions can be made quickly and efficiently using sociodemographic and morbidity parameters. For further follow-up and care, at-risk patients can be referred to consultant physicians. This will lessen the burden of gastrointestinal-related morbidity and mortality among the younger population.

Acknowledgement

The authors would like to express their sincere gratitude to Department of Dietetics and Nutrition, NSHM Knowledge Campus Kolkata, and Amity University Kolkata for providing the necessary facilities and support. The authors are also profoundly grateful to all the participants in this study for their invaluable contributions.

Funding Sources

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Conflict of Interest

The authors do not have any conflict of interest.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Ethics Statement

The present study was started after getting approval from institutional ethics committee of All India Institute of Hygiene and Public Health, Kolkata, and it is part of a bigger project published elsewhere.18

Informed Consent Statement

Informed consent was collected before conducting the study.

Author Contributions

- **Joyeta Ghosh:** Conceptualization, Methodology, Writing – Original Draft.
- **Poulomi Sanyal:** Data Collection, Analysis, Writing – Review & Editing.

References

1. Vilanova-Sanchez A , Levitt MA. Surgical Interventions for Functional Constipation: An Update. *Eur J Pediatr Surg* 2020; 30(5):413-419. DOI: https://doi.org/10.1055/s-0040-1708061

2. Vriesman MH, Koppen I, Camilleri M, Di Lorenzo C, Benninga MA. Management of functional constipation in children and adults. *Nat Rev Gastroenterol Hepatol* 2020; 17(1):21-39. DOI: https://doi.org/10.1038/

s41575-019-0222-y

3. Mearin F, Ciriza C, Mínguez M, Rey E, Mascort JJ, Peña E, Cañones P, Júdez J. Clinical Practice Guideline: Irritable bowel syndrome with constipation and functional constipation in the adult. *Rev Esp Enferm Dig.* 2016 Jun;108(6):332-63. DOI: https://doi.org/10.17235/reed.2016.4089/2015
4. Aziz I, Whitehead WE, Palsson OS, Tornblom H, Simren M. An approach to the diagnosis and management of Rome IV functional disorders of chronic constipation. *Expert Rev Gastroenterol Hepatol* 2020; 14(1):39-46. DOI: https://doi.org/10.1080/17474124.2020.1702461
5. Shin JE, Park KS, Nam K. Chronic Functional Constipation. *Korean J Gastroenterol.* 2019; 73(2):92-98. DOI: https://doi.org/10.4166/kjg.2019.73.2.92
6. Barberio B, Judge C, Savarino EV, Ford AC. Global prevalence of functional constipation according to the Rome criteria: A systematic review and meta-analysis. *Lancet Gastroenterol Hepatol* 2021; 6(8):638-648. DOI: https://doi.org/10.1016/S2468-1253(21)00105-5
7. Forootan M, Bagheri N, Darvishi M. Chronic constipation: A review of literature. *Medicine* (Baltimore). 2018; 97(20):e10631. DOI: https://doi.org/10.1097/MD.0000000000010631
8. Adili A, Gulichekran E, Ruxianguli A, Han G. Analysis of factors influencing constipation among college students in a medical school in Xinjiang. *Health Med Res Pract* 2021; 18(1):39-43. DOI: https://doi.org/10.34172/hmrp.2021.08
9. The cost of constipation. *Lancet Gastroenterol Hepatol.* 2019; 4(11):811. DOI: https://doi.org/10.1016/S2468-1253(19)30327-4
10. Brochard C, Chambaz M, Ropert A, l'Héritier AM, Wallenhorst T, Bouguen G, Siproudhis L. Quality of life in 1870 patients with constipation and/or fecal incontinence: Constipation should not be underestimated. *Clin Res Hepatol Gastroenterol* 2019 Nov;43(6):682-687. DOI: https://doi.org/10.1016/j.clinre.2019.07.010
11. Lim YJ, Rosita J, Chieng JY, Hazizi AS. The Prevalence and Symptoms Characteristic of Functional Constipation Using Rome III Diagnostic Criteria among Tertiary Education Students. *PLoS One* 2016; 11(11):e0167243. DOI: https://doi.org/10.1371/journal.pone.0167243
12. Liu X, Liu Y, Chen J, Wang H, Wang Q, Niu Z, Yun Z, Ma B, Yao S. Effectiveness and safety of light vegetarian diet and Qingjiang Tiaochang Recipe for functional constipation: An exploratory study protocol for randomized controlled trial. *Medicine* (Baltimore). 2020 Sep 25;99(39):e21363. DOI: https://doi.org/10.1097/MD.0000000000018325
13. Li L, Huang AP, Wang LQ, Yu XL. Empirically derived dietary patterns and constipation among a middle-aged population from China, 2016-2018. *Nutr J* 2019; 18(1):88. DOI: https://doi.org/10.1186/s12937-019-0510-2
14. Shinjyo N, Waddell G, Green J. Valerian Root in Treating Sleep Problems and Associated Disorders-A Systematic Review and Meta-Analysis. *J Evid Based Integr Med* 2020; 25:2515690X20967323. DOI: https://doi.org/10.1177/2515690X20967323
15. Orr WC, Fass R, Sundaram SS, Scheimann AO. The effect of sleep on gastrointestinal functioning in common digestive diseases. *Lancet Gastroenterol Hepatol* 2020; 5(7):616-624. DOI: https://doi.org/10.1016/S2468-1253(20)30018-3
16. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. In: Artificial Intelligence in Healthcare. *Academic Press;* 2020:25-60. DOI: https://doi.org/10.1016/B978-0-12-818438-7.00002-2
17. Rajput M, Saini SK. Prevalence of constipation among the general population: a community-based survey from India. *Gastroenterol Nurs* 2014; 37(6):425-429. DOI: https://doi.org/10.1097/SGA.0000000000000083
18. Ghosh J, Chaudhuri D, Saha I, Chaudhuri AN. Prevalence of metabolic syndrome, vitamin D level, and their association among elderly women in a rural community of West Bengal, India. *Med J DY Patil Vidyapeeth* 2020; 13(3):315-320. DOI: https://doi.org/10.4103/mjdrdypu.mjdrdypu_229_19
19. Ghosh J, Choudhury SR, Singh K, Koner S. Application of machine learning algorithm and artificial intelligence in improving metabolic syndrome related complications: A review. *Int J Adv Life Sci Res.* 2024; 7(2):41-67. DOI https://doi.org/10.31632/ijalsr.2024.

v07i02.004

20. Ghosh J. Recognizing and predicting the risk of malnutrition in the elderly using artificial intelligence: A systematic review. *Int J Adv Life Sci Res* 2024; 7(3):1-14. DOI https://doi.org/10.31632/ijalsr.2024.v07i03.001
21. Ghosh J, Sanyal P, Singh K, Roy Choudhury S, Koner S. Prevalence of Constipation and its Relationship with Dietary Habits Among College Going Girls in the Age Group of 18-25 Years of Kolkata, West Bengal, India. *Acta Sci Gastrointest Disord* 2023; 6(3):3-13. DOI:https://actascientific.com/ASGIS/pdf/ASGIS-06-0516.pdf
22. He L, Xu SL. A study of life events and their psychological impact among university students—A visual analysis based on CiteSpace. *Educ Watch* 2021; 10(1):5-8. DOI:https://doi.org/10.3390/nu14214590
23. Marum G, Clench-Aas J, Nes RB, Raanaas RK. The relationship between negative life events, psychological distress and life satisfaction: A population-based study. *Qual Life Res.* 2014; 23(2):601-611. DOI: https://doi.org/10.1007/s11136-013-0512-8
24. Zhang Y, Lin Q, An X, Tan X, Yang L. Factors Associated with Functional Constipation among Students of a Chinese University: A Cross-Sectional Study. *Nutrients* 2022; 14(21):4590. DOI: https://doi.org/10.3390/nu14214590
25. Alpaydin E. Introduction to Machine Learning. MIT Press; 2014.Ethem Alpaydin-Introduction to Machine Learning-The MIT Press (2014).pdf (matlabyar.com)
26. Ghosh J, Choudhury SR, Singh K, Koner S. Application of machine learning algorithm and artificial intelligence in improving metabolic syndrome related complications: A review. *International Journal of Advancement in Life Sciences Research* 2024; 7(2):41-67. DOI:10.31632/ijalsr.2024.v07i02.004
27. Krstajic D, Buturovic LJ, Leahy DE, Thomas S. Cross-validation pitfalls when selecting and assessing regression and classification models. *J Cheminform* 2014; 6(1):10. DOI: https://doi.org/10.1186/1758-2946-6-10
28. Bellini M, Tonarelli S, Barracca F, Rettura F, Pancetti A, Ceccarelli L, Ricchiuti A, Costa F, de Bortoli N, Marchi S, Rossi A. Chronic Constipation: Is a Nutritional Approach Reasonable? *Nutrients.* 2021 Sep 26;13(10):3386. DOI: https://doi.org/10.3390/nu13103386
29. Werth BL,Williams KA,Fisher MJ,Lisa GPont. Defining constipation to estimate its prevalence in the community: results from a national survey. *BMC Gastroenterol*. 2019; 19(1):1-7. DOI: https://doi.org/10.1186/s12876-019-1007-z
30. Ghosh J. A review on understanding the risk factors for coronary heart disease in Indian college students. *Int J Non Commun Dis* 2023; 8:117-28.DOI: 10.4103/jncd.jncd_68_23
31. Jangid V, Godhia M, Sanwalka N, Shukla A. Water intake, dietary fibre, defecatory habits and its association with chronic functional constipation. *Curr Res Nutr Food Sci.* 2016; 4(2):90-95. DOI: https://doi.org/10.12944/CRNFSJ.4.2.06
32. Dukas L. Association between physical activity, fiber intake, and other lifestyle variables and constipation in a study of women. *Am J Gastroenterol.* 2003; 98(8):1790-1796. DOI: https://doi.org/10.1111/j.1572-0241.2003.07591.x
33. Choudhury SR, Ghosh J, Singh K,Koner S, Bera A. Traditional Indian Food for Improving Brain Cognition. *Acta Sci Neurol.* 2022; 5(12):23-29. DOI: 10.31080/ASNE.2022.05.0561
34. Ghosh J, Singh K, Choudhury SR, Basu N. Impact of Diet and Nutrition on Memory T Cell Development, Maintenance and Function in the Context of a Healthy Immune System. *Acta Sci Nutr Health* 2022; 6(8):110-119.DOI: 10.31080/ASNH.2022.06.1108
35. Karabudak E, Koksal E, Macit M. The relationship between body weight, fiber and fluid intake status and functional constipation in young adults. *Nutr Food Sci* 2019; 49(1):129-140. DOI: https://doi.org/10.1108/NFS-03-2018-0090
36. Abdullah MMH, Gyles CL, Marinangeli CPF, Carlberg JG, Jones PJH. Dietary fibre intakes and reduction in functional constipation rates among Canadian adults: A cost-of-illness analysis. *Food Nutr Res.* 2015; 59:28646. DOI: https://doi.org/10.3402/fnr.v59.28646
37. Huang R, Ho SY, Lo WS, Lam TH. Physical activity and constipation in Hong Kong

adolescents *PLoS One*. 2014; 9(2):e90193. DOI: https://doi.org/10.1371/journal.pone.0090193

38. Bharucha AE, Lacy BE. Chronic constipation: Mechanisms, evaluation and management. *Gastroenterology* 2020; 158(5):1232-1249.e4. DOI: https://doi.org/10.1053/j.gastro.2020.01.016
39. Baradia R, Ghosh J. Impact of Mindful Eating among Adolescent. *Int J Sci Res.* 2021; 10(11):11-15.Impact of Mindful Eating among Adolescent (ijsr.net)
40. Das, Poulomi,Banka R,Ghosh J,Singh K,Roychaudhury S,Koner S. "Synergism of Diet, Genetics, and Microbiome on Health." Nutrition Controversies and Advances in Autoimmune Disease, edited by Srikanta Patnaik, IGI Global, 2024, pp. 131-189. https://doi.org/10.4018/979-8-3693-5528-2.ch006
41. Shakil S, Ghosh J, Singh K, Chaudhury SR. Comparative analysis of nutritional status among institutionalized and community-dwelling elderly women and its association with mental health status and cognitive function. *J Fam Med Prim Care.* 2024; 13(8):3078-3083. DOI: 10.4103/jfmpc.jfmpc_1932_23.
42. Lacy BE. Update on the management of chronic idiopathic constipation. *Am J Manag Care.* 2019; 25(4 Suppl):S55-S62.Update on the management of chronic idiopathic constipation - PubMed (nih.gov)
43. De Schryver AM, Keulemans YC, Peters HP, Akkermans LM, Smout AJ, De Vries WR, van Berge-Henegouwen GP. Effects of regular physical activity on defecation pattern in middle-aged patients complaining of *chronic constipation.* Scand J Gastroenterol. 2005; 40(4):422-429. DOI: https://doi.org/10.1080/00365520510011641
44. Gwee KA, Ghoshal UC, Gonlachanvit S, Chua AS, Myung SJ, Rajindrajith S, Patcharatrakul T, Choi MG, Wu JC, Chen MH, Gong XR, Lu CL, Chen CL, Pratap N, Abraham P, Hou XH, Ke M, Ricaforte-Campos JD, Syam AF, Abdullah M.Primary care management of chronic constipation in Asia: The ANMA Chronic Constipation Tool. *J Neurogastroenterol Motil.* 2013; 19(2):149-160. DOI: https://doi.org/10.5056/jnm.2013.19.2.149
45. Srinivas M, Srinivasan V, Jain M, Rani Shanthi CS, Mohan V, Jayanthi V. A cross-sectional study of stool form (using Bristol stool chart) in an urban South Indian population. *J Gastroenterol Hepatol.* 2019; 3(6):464-467. DOI: https://doi.org/10.22271/27069567.2019.v3.i6e.160