Bayesian Network for Modeling Causal Relationships Between Risk Factors and Gut-brain Axis Disorders in Honduran Communities

Ethel Flores[1]; Victor Funez[2]; Rene Gonzales[3]; Karen Oliva[4]; Diana Zelaya[5]; Oneyda Zelaya[6]; Jorge Urmeneta[7]

[1,2,3,4,7]Faculty of Medical Sciences, National Autonomous University of Honduras, Honduras, *ethel.flores@unah.edu.hn, victorhugofunezm@gmail.com, rene.gonzales@unah.edu.hn, karen.oliva@unah.edu.hn, jorge.urmeneta@unah.edu.hn*

[5,6]Policia Nacional de Honduras, Honduras, *lizethzelaya2010@gmail.com, oneydaespinoza1996@gmail.com*

Abstract*– Gut-brain Axis Disorders (GBAD) are a major public health issue in poor nations, although Central American epidemiology and risk factor modeling are lacking. Bayesian Network (BN) models were developed and validated to characterize probability correlations between demographic, behavioral, and clinical characteristics related with GBAD in Honduran communities. A cross-sectional survey was undertaken in September–December 2025 among 1,838 12-49-year-olds in six health regions. Use of gastrointestinal literature with Maximum Likelihood Estimation to create the BN structure. The model validation used 5-fold cross-validation, resulting in an AUC-ROC of 0.686 (±0.028). Total GBAD prevalence was 56.0%. Conditional probability analysis showed that prior medical diagnosis (P(GBAD|Diagnosis)=0.840), smoking (P(GBAD|Smoking)=0.815), and alcohol intake each had the highest relationships with GBAD. A comprehensive healthy lifestyle reduced GBAD likelihood by 32.0% relative to baseline in what-if intervention analysis. Prior diagnosis (MI=0.047), alcohol intake (MI=0.017), and smoking (MI=0.016) were the most informative predictors. The suggested BN framework helps public health practitioners evaluate intervention scenarios and allocate community GBAD preventive resources in a clear, interpretable manner.

Keywords*-- Bayesian networks, Gut-brain Axis Disorders, risk factors, probabilistic modeling, public health informatics.

I. Introduction

Gut-brain Axis Disorders (GBAD) include abdominal pain, heartburn, bloating, and bowel changes; they are a global public health issue. The Rome IV criteria describe these illnesses as chronic digestive symptoms without structural or biochemical abnormalities [1]. Global epidemiological studies estimate that 11%–15% of the population has a functional gastrointestinal condition, with IBS being the most common [2]. These situations are particularly difficult in Latin American countries like Honduras due to limited healthcare infrastructure, socioeconomic inequities, and environmental factors that may affect disease manifestation [3].

The progressive increase in obesity triggers multiple problems linked to the risk of gastrointestinal issues. This, coupled with insufficient intake of healthy foods and excessive consumption of ultra-processed foods, becomes a risk factor that triggers morbidity throughout the digestive system [4], [5]. The use of artificial intelligence is necessary due to the increase in patients, the variability in symptoms, and, above all, the lack of sufficient information, which makes it difficult to make scenario-based decisions [6]-[8].

GBAD is caused by numerous biological, psychological, and social factors. Diet, physical activity, tobacco and alcohol usage, stress, and past gastrointestinal disorders have been linked [9]. Traditional statistical methods struggle to identify causal pathways and conditional connections among these factors. Conventional regression models can quantify individual risk factors, but they typically fail to reflect the complex interactions that characterize real-world health phenomena [10].

Bayesian Networks (BNs) describe complex dependencies probabilistically and graphically. A directed acyclic network with nodes representing random variables, edges conditional dependencies, and CPTs quantifying these interactions is a BN [11]. BNs have various advantages over traditional methods: (i) explicit causal assumptions, (ii) uncertainty-based inference, (iii) natural handling of missing data, and (iv) what-if scenario analysis for public health intervention planning [12]. BNs are ideal for medical informatics applications that require interpretability and uncertainty quantification [13].

The Honduran National Health Model considers Primary Health Care (PHC) the foundation for equal healthcare access [14]. Community-based health efforts need evidence-based methods to identify at-risk people and assess preventive treatments. Epidemiological research on Honduran GBAD prevalence and associated variables is scarce despite this need. This information gap limits primary care GBAD prevention and management measures.

We developed and validated a Bayesian Network model to quantify the probabilistic links between demographic, behavioral, and clinical characteristics associated with GBAD in a representative sample of Honduran communities to fill this gap. The objectives are to (i) estimate GBAD prevalence in the study population, (ii) construct a BN model encoding causal assumptions from domain knowledge, (iii) validate the model using cross-validation, (iv) simulate public health intervention effects, and (v) identify the most informative risk factors through mutual information analysis.

II. Materials and Methods

A. Study Design and Population

A cross-sectional epidemiological study was undertaken in communities assigned to the sixth and seventh rotations of Public Health IV at the National Autonomous University of Honduras from September to December 2025. The target population comprised individuals aged 12 to 49 years residing in communities across six health regions: Intibucá (RS10), Colón (RS2), Comayagua (RS3), Olancho (RS15), El Paraíso (RS7), and Atlántida (RS1).

Inclusion criteria were: (1) age between 12 and 49 years; (2) minimum residence of 6 months in the community; and (3) provision of informed consent (or parental assent for minors). Exclusion criteria included: (1) prior diagnosis of organic gastrointestinal disease; (2) recent treatment (within 3 months) for gastrointestinal infections; and (3) incomplete survey responses.

B. Data Collection Instrument

Trained sixth-year medical students administered a standardized questionnaire face-to-face under instructor supervision. Six components made up the instrument: Social demographics (age, sex, community, education level); Gut-brain Axis Disorders (abdominal pain in last 30 days, heartburn in last 2 weeks, gastric reflux frequency, post-prandial abdominal inflammation); dietary habits (fiber, water); lifestyle factors (physical activity, smoking, alcohol); self-perceived digestive health status; and prior medical diagnoses (IBS, g. Before deployment, the questionnaire was pilot tested for understanding and internal consistency.

C. Variable Operationalization

The primary outcome variable, GBAD presence, was defined as a binary composite indicator equal to 1 if the respondent reported any of the following: abdominal pain in the last 30 days, heartburn in the last 2 weeks, abdominal inflammation after meals. This operationalization aligns with the Rome IV symptomatic criteria for functional gastrointestinal disorders [1]. Independent variables were categorized as follows: Sex (0=Male, 1=Female); Age Group (0=12-25 years, 1=26-35 years, 2=36-49 years); Education (0=Low/Incomplete, 1=High/Complete secondary or university); Physical Activity (0=Less than 3 times/week, 1=At least 3 times/week); Smoking (0=No, 1=Yes); Alcohol Consumption (0=No, 1=Yes); Fiber Intake (0=No daily fiber, 1=Daily fiber consumption); Water Intake (0=Less than 8 glasses/day, 1=At least 8 glasses/day); and Prior Medical Diagnosis (0=No, 1=Yes).

D. Bayesian Network Construction

Medical literature on GBAD risk factors guided the BN structure's design. Causal assumptions for the directed acyclic graph: Demographic factors (age, sex) directly affect GBAD probability; behavioral risk factors (smoking, alcohol, physical activity, diet) directly affect GBAD; prior medical diagnosis is associated with higher GBAD probability; education level influences health behaviors (physical activity, fiber intake); age increases the likelihood of a prior diagnosis; and smoking and alcohol consumption are correlated risk behaviors. While being frugal, this structure reflects the biopsychosocial paradigm of functional gastrointestinal diseases [15].

Conditional probability tables were estimated using Maximum Likelihood Estimation (MLE) from the observed data frequencies. For a binary child node Y with parents X_1, X_2, ..., X_k, the CPT entry $P(Y=1|X_1=x_1, X_2=x_2, ..., X_k=x_k)$ was computed as the proportion of observations where Y=1 among all observations with the specified parent configuration. This approach provides intuitive, data-driven probability estimates while maintaining computational efficiency [16].

E. Model Validation

The prediction performance of the model was assessed by stratified 5-fold cross-validation. In each fold, the Bayesian Network was trained on 80% of the data (training set) and predictions were made for the remaining 20% (test set). Probabilistic inference was conducted via Variable Elimination, calculating P(GBAD=1|Evidence) for each test observation. The classification utilized a probability threshold of 0.5. Performance measurements encompassed accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC). The AUC-ROC was chosen as the principal evaluation statistics because of its resilience to class imbalance and independence from thresholds [17].

F. What-If Intervention Analysis

The BN framework enabled simulation of hypothetical intervention scenarios through conditional probability queries. Four scenarios were evaluated: Scenario 1 (elimination of tobacco and alcohol use), Scenario 2 (promotion of physical activity and fiber intake), Scenario 3 (comprehensive healthy lifestyle combining all protective factors), and Scenario 4 (high-risk profile with all adverse behaviors). For each scenario, P(GBAD=1|Intervention) was computed and compared to the baseline population prevalence. Relative risk reduction (RRR) was calculated as: [18].

$$RRR = (P\_baseline - P\_intervention) / (P\_baseline \times 100\%) \quad (1)$$

Variable relevance was assessed by Mutual Information (MI), which quantifies the statistical reliance between each predictor and the result. Mutual Information was computed via the k-nearest neighbors' algorithm with discrete attributes [19]. Elevated MI values signify enhanced predictive significance. Furthermore, chi-square tests of independence were conducted to evaluate the statistical significance of bivariate relationships, with odds ratios (OR) and 95% confidence intervals calculated for clarity.

G. Statistical Software

All analyses were conducted using Python 3.12 [20] with the following libraries: pandas 2.0 for data manipulation, NumPy 1.26 for numerical computations, SciPy 1.11 for statistical tests, and scikit-learn 1.3 for machine learning algorithms and validation procedures [21]-[24]. Statistical significance was set at $\alpha = 0.05$.

H. Ethical Considerations

This study followed the ethical principles of the 1964 Declaration of Helsinki and its subsequent amendments for research involving human subjects [25]. All participants provided informed consent prior to inclusion. The study was subsequently submitted to the Ethics Committee for Scientific Research in Health, College of Nutritionists and Dietitians of Honduras, which granted its approval on October 2025; under the number CEICS-CNDH-002-10-2025. This is required, because we used data from humans to validate our simulation.

III. Results

A. Sample Characteristics

A total of 1,838 participants met the inclusion criteria and provided complete data. The mean age was 28.3 years (SD = 10.9), with ages ranging from 12 to 49 years. The sample was predominantly female (65.2%, n=1,198). Regarding education, 29.4% completed secondary or higher education. Geographic distribution included Intibucá (31.6%), Colón (29.8%), Comayagua (26.6%), Olancho (8.7%), El Paraíso (2.9%), and Atlántida (0.4%). Table I presents detailed sample characteristics.

TABLE I
SAMPLE CHARACTERISTICS (N=1,838)

Variable	n	%
Sex		
Female	1,198	65.2
Male	640	34.8
Age Group		
12-25 years	748	40.7
26-35 years	533	29.0
36-49 years	557	30.3
GBAD Present	1,030	56.0
Current Smoker	189	10.3
Alcohol Consumer	300	16.3
Regular Physical Activity	878	47.8
Daily Fiber Intake	1,191	64.8
Adequate Water Intake	786	42.8
Prior GI Diagnosis	388	21.1

Note: GI = Gastrointestinal; GBAD = Gut-brain Axis Disorders.

B. Optimal Cluster Number Determination

The overall prevalence of GBAD in the study population was 56.0% (n=1,030). Table II presents the conditional probabilities of GBAD given each risk factor, along with chi-square test statistics and odds ratios. Prior medical diagnosis exhibited the strongest association (OR=5.57, 95% CI: 4.17-7.45, p<0.001), followed by smoking (OR=3.88, 95% CI: 2.66-5.68, p<0.001) and alcohol consumption (OR=2.97, 95% CI: 2.23-3.94, p<0.001). Regular physical activity (OR=0.63, 95% CI: 0.52-0.76, p<0.001) and daily fiber intake (OR=0.69, 95% CI: 0.56-0.83, p<0.001) demonstrated protective effects.

C. Model Validation Performance

Table III encapsulates the outcomes of the 5-fold cross-validation. The BN model attained an AUC-ROC of 0.686 (±0.028), signifying a satisfactory discriminative capacity for a model reliant exclusively on survey-derived risk factors, without biomarkers. The model exhibited high precision (0.780 ± 0.024), indicating dependable identification of real GBAD cases among anticipated positives. Recall was moderate (0.489 ± 0.016), indicating conservative predictions that prioritize specificity.

TABLE II
CONDITIONAL PROBABILITIES AND ASSOCIATION MEASURES

Factor	P(GBAD=1 |0)	P(GBAD=1 |1)	χ^2	OR (95% CI)
Sex (Female)	0.539	0.572	1.68	1.14 (0.94-1.38)
Smoking	0.531	0.815	54.21*	3.88 (2.66-5.68)
Alcohol	0.521	0.763	58.96*	2.97 (2.23-3.94)
Physical Activity	0.615	0.501	23.50*	0.63 (0.52-0.76)
Fiber Intake	0.620	0.528	13.93*	0.69 (0.56-0.83)
Water Intake	0.588	0.523	7.57*	0.77 (0.64-0.92)
Prior Diagnosis	0.486	0.840	154.88*	5.57 (4.17-7.45)

*p < 0.001; OR = Odds Ratio; CI = Confidence Interval.

TABLE III
CROSS-VALIDATION PERFORMANCE METRICS

Metric	Mean	Std. Dev.
Accuracy	0.635	0.025
AUC-ROC	0.686	0.028
Precision	0.780	0.024
Recall	0.489	0.016
F1-Score	0.600	0.009

Note: Results from stratified 5-fold cross-validation.

D. What-If Intervention Analysis

Table IV shows what-if intervention results. Baseline GBAD probability was 0.560. Eliminating smoke and alcohol (Scenario 1) lowered GBAD likelihood to 0.507, a 9.5% relative risk reduction. Physical activity and fiber consumption (Scenario 2) reduced P(GBAD) by 14.3% to 0.480. The most effective healthy living intervention (Scenario 3) reduced GBAD probability to 0.381, 32.0% relative risk. Scenario 4's high-risk profile increased GBAD probability by 65.7% to 0.929.

E. Variable Importance Analysis

Mutual Information (MI) analysis identified prior medical diagnosis as the most significant predictor (MI=0.047), succeeded by alcohol intake (MI=0.017) and smoking (MI=0.016). The age group (MI=0.010) and physical activity (MI=0.007) had modest predictive significance. Sex (MI=0.0005) and educational attainment (MI=0.0007) had

negligible contributions to GBAD prediction. These findings support the conditional probability analysis and emphasize modifiable behavioral elements as primary intervention targets.

TABLE IV
WHAT-IF INTERVENTION SCENARIO ANALYSIS

Metric	P(GBAD=1)	Change (%)
Baseline (population prevalence)	0.560	Reference
Scenario 1: No smoking + No alcohol	0.507	-9.5
Scenario 2: Physical activity + Fiber intake	0.480	-14.3
Scenario 3: Comprehensive healthy lifestyle	0.381	-32.0
Scenario 4: High-risk profile	0.929	+65.7

Note: Comprehensive healthy lifestyle = No smoking, no alcohol, regular physical activity, daily fiber intake, adequate water intake.

TABLE V
VARIABLE IMPORTANCE RANKING BY MUTUAL INFORMATION

Rank	Variable	MI Score
1	Prior Medical Diagnosis	0.0466
2	Alcohol Consumption	0.0172
3	Smoking	0.0164
4	Age Group	0.0103
5	Physical Activity	0.0065
6	Fiber Intake	0.0039
7	Water Intake	0.0021
8	Education Level	0.0007
9	Sex	0.0005

Note: MI = Mutual Information; higher values indicate stronger predictive relevance.

IV. DISCUSSION

This study introduces the inaugural application of Bayesian Network modeling to delineate functional gastrointestinal complaints in Honduran communities. The documented prevalence of GBAD at 56.0% above global estimates of 11-15% [2], but direct comparisons are constrained by variations in case definitions and evaluation methodologies. The increased prevalence may indicate the comprehensive outcome definition that includes many symptom dimensions, together with socioeconomic and environmental factors particular to the study populations. This prevalence corresponds with regional data indicating a greater burden of GBAD in Latin American populations relative to North American or European cohorts [26].

The significant correlation between previous medical diagnosis and GBAD (OR=5.57) requires meticulous interpretation. This link presumably indicates diagnostic ascertainment bias, in which persons with enduring symptoms are more inclined to pursue and obtain a diagnosis. Nonetheless, it highlights the persistent and recurrent characteristics of functional gastrointestinal diseases, aligning with the Rome IV focus on the continuity of symptoms throughout time [1]. From a public health standpoint, individuals with previous diagnoses constitute a high-priority demographic for targeted treatments and continuous management.

The recognized correlations between behavioral variables and GBAD align with existing literature. Smoking is associated with modified gastrointestinal motility, mucosal inflammation, and compromised lower esophageal sphincter function [27]. Alcohol intake is recognized to alter gut microbiota composition and compromise intestinal barrier integrity [28]. In contrast, consistent physical activity has shown advantages for gastrointestinal function by improving motility and alleviating stress, and dietary fiber fosters a healthy gut flora and regular bowel motion [29], [30].

The BN model demonstrated satisfactory discriminative performance (AUC-ROC=0.686) despite its dependence on self-reported survey data in the absence of objective biomarkers. This performance level aligns with similar probabilistic models for symptom-based outcomes in primary care environments [31]. The high precision (0.780) indicates that the model consistently identifies persons at actual risk when a positive prediction occurs, a characteristic that is advantageous for resource-constrained healthcare environments where false positives result in substantial expenses. The intermediate recall (0.489) suggests possibility for enhancement, possibly by integrating supplementary variables like psychological distress metrics or nutritional information.

The what-if study yields meaningful insights into public health policy. The discovery that extensive lifestyle changes can decrease the likelihood of GBAD by 32.0% highlights the significant effect of multi-faceted therapies targeting tobacco, alcohol, physical activity, and diet concurrently. This is consistent with the World Health Organization's guidelines for integrated non-communicable disease prevention [32]. The significant risk rise in the high-risk scenario (65.7%) underscores the cumulative impact of several detrimental habits, emphasizing the necessity for early diagnosis and intervention for vulnerable individuals.

Numerous constraints merit attention. Cross-sectional architecture prevents causal inference, and the BN structure, although knowledge-driven, signifies one possible configuration among others. The symptom-based definition of GBAD may encompass both temporary illnesses and genuine functional impairments. Self-reported behavioral data are susceptible to social desirability and recollection biases. Geographic clustering within health regions may lead to unaccounted confounding variables. Subsequent research ought to utilize longitudinal designs, integrate structural learning algorithms for data-driven network identification, and corroborate findings in separate cohorts.

Notwithstanding these constraints, the study offers multiple contributions. This represents the inaugural community-level epidemiological characterization of GBAD in Honduras. The BN framework provides a clear, interpretable instrument that can be easily revised as new data emerges. The what-if analysis feature facilitates evidence-based scenario planning for public health officials. The

methodology is applicable to different community health contexts where comprehending conditional relationships among risk factors is crucial.

V. Conclusion

A Bayesian Network model describing risk variables for functional gastrointestinal complaints in Honduran communities was created and verified in this study. The importance of these disorders for public health is highlighted by the high incidence of GBAD (56.0%) and the discovery of behavioral risk factors that can be changed. Physical activity and fiber intake showed protective relationships, although smoking, alcohol use, and previous medical diagnosis were the strongest predictors. By simulating intervention situations and achieving adequate prediction accuracy (AUC-ROC=0.686), the model demonstrated potential risk reductions of up to 32.0% with thorough lifestyle adjustment. For primary healthcare professionals and public health planners, the Bayesian Network concept offers an interpretable, practical tool for efficiently allocating resources and prioritizing interventions. To improve predictive accuracy and causal knowledge of functional gastrointestinal diseases in vulnerable populations, future research should include psychological aspects and longitudinal data.

Acknowledgment

We appreciate the support received from the rural communities and the Honduran Ministry of Public Health. This research was supported by the Institute for Research in Medical Sciences and Right to Health (ICIMEDES) of the Faculty of Medical Sciences (FCM) and the Directorate of Scientific, Humanistic, and Technological Research (DICIHT) of the National Autonomous University of Honduras (UNAH).

The authors thank the students of the Public Health IV course of UNAH for their contributions to the parent study design and data collection planning. We also acknowledge Prof. Marcio Madrid and Prof. Isaac Zablah as expert methodological and technical guidance and other aspects of this research. Likewise, we thank Martha Casco, MSc, from the Public Health Department.

References

[1] D. A. Drossman, "Functional Gastrointestinal Disorders: History, Pathophysiology, Clinical Features and Rome IV", *Gastroenterology*, vol. 150, no. 6, pp. 1262-1279, May 2016. doi: 10.1053/j.gastro.2016.02.032

[2] A. D. Sperber et al., "Worldwide Prevalence and Burden of Functional Gastrointestinal Disorders, Results of Rome Foundation Global Study", *Gastroenterology*, vol. 160, no. 1, pp. 99-114, Jan. 2021. doi: 10.1053/j.gastro.2020.04.014

[3] M. J. Schmulson and D. A. Drossman, "What Is New in Rome IV" *Journal of Neurogastroenterology and Motility*, vol. 23, no. 2, pp. 151-163, Apr. 2017. doi: 10.5056/jnm16214

[4] Y. Choe, "Obesity and Upper Gastrointestinal Diseases", *Korean J. Gastroenterol.*, vol. 83, no. 3, pp. 81–86, Mar. 2024, doi: 10.4166/kjg.2024.015.

[5] A. K. Singh, A. Gandotra, S. Kumar, A. Singh, R. Kochhar, and M. Manrai, "Ultra-processed foods: Implications for gastrointestinal health", *World J. Gastroenterol.*, vol. 31, no. 36, Art. no. 109143, Sep. 2025, doi: 10.3748/wjg.v31.i36.109143

[6] A. Da'Costa, J. Teke, J. E. Origbo, A. Osonuga, E. Egbon, and D. B. Olawade, "AI-driven triage in emergency departments: A review of benefits, challenges, and future directions", *International Journal of Medical Informatics*, vol. 197, art. no. 105838, May 2025, doi: 10.1016/j.ijmedinf.2025.105838

[7] H. Wu, W. Shi, A. Choudhary, et al., "Clinical decision making under uncertainty: a bootstrapped counterfactual inference approach", *BMC Medical Informatics and Decision Making*, vol. 24, art. no. 275, Sep. 2024, doi: 10.1186/s12911-024-02606-z

[8] S. Labkoff, B. Oladimeji, J. Kannry, et al., "Toward a responsible future: recommendations for AI-enabled clinical decision support", *Journal of the American Medical Informatics Association*, vol. 31, no. 11, pp. 2730–2739, Nov. 2024, doi: 10.1093/jamia/ocae209

[9] F. Mearin et al., "Bowel Disorders", *Gastroenterology*, vol. 150, no. 6, pp. 1393-1407, May 2016. doi: 10.1053/j.gastro.2016.02.031

[10] D. Koller and N. Friedman, Probabilistic Graphical Models: Principles and Techniques. Cambridge, MA, USA: MIT Press, 2009.

[11] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Francisco, CA, USA: Morgan Kaufmann, 1988.

[12] P. J. F. Lucas, L. C. van der Gaag, and A. Abu-Hanna, "Bayesian networks in biomedicine and health-care", *Artificial Intelligence in Medicine*, vol. 30, no. 3, pp. 201-214, Mar. 2004. doi: 10.1016/j.artmed.2003.11.001

[13] O. Pourret, P. Naïm, and B. Marcot, Eds., Bayesian Networks: A Practical Guide to Applications. Chichester, UK: Wiley, 2008.

[14] Secretaría de Salud de Honduras, "Modelo Nacional de Salud," Tegucigalpa, Honduras, 2023. [Online]. https://extranet.who.int/countryplanningcycles/sites/default/files/planning_cycle_repository/honduras/modelo_nacional_de_salud_secretaria_de_salud_mayo_2013_version_21-05-13.pdf

[15] [1] D. A. Drossman, D. W. Powell, y J. T. Sessions Jr, "The irritable bowel syndrome", *Gastroenterology*, vol. 73, núm. 4, pp. 811–822, 1977, doi: 10.1016/s0016-5085(19)31790-1.

[16] R. E. Neapolitan, Learning Bayesian Networks. Upper Saddle River, NJ, USA: Pearson Prentice Hall, 2004.

[17] T. Fawcett, "An introduction to ROC analysis", *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861-874, Jun. 2006. doi: 10.1016/j.patrec.2005.10.010

[18] K. J. Rothman, S. Greenland, and T. L. Lash, Modern Epidemiology, 3rd ed. Philadelphia, PA, USA: Lippincott Williams & Wilkins, 2008.

[19] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information", *Physical Review E*, vol. 69, no. 6, p. 066138, Jun. 2004. doi: 10.1103/PhysRevE.69.066138

[20] Python Software Foundation, *Python Language Reference*, version 3.11. Wilmington, DE, USA, 2023. [Online]. Available: https://www.python.org

[21] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. Fernández del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke y T. E. Oliphant, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020, doi: 10.1038/s41586-020-2649-2

[22] The Pandas Development team, *pandas-dev/pandas: Pandas*, ver. 2.2.0, Zenodo, Jan. 20 2024, doi: 10.5281/zenodo.10537285

[23] W. McKinney, "Data structures for statistical computing in Python", in *Proc. 9th Python Sci. Conf. (SciPy)*, Austin, TX, USA, 2010, pp. 51–56.

[24] F. Pedregosa et al., "Scikit-learn: Machine learning in Python", *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011

[25] World Medical Association, "Declaration of Helsinki: Ethical principles for medical research involving human subjects", *JAMA*, vol. 310, no. 20, pp. 2191–2194, Nov. 2013, doi: 10.1001/jama.2013.281053

[26] A. López-Colombo, D. Morgan, D. Bravo-González, A. Montiel-Jarquín, S. Méndez-Martínez, y M. Schmulson, "The epidemiology of functional gastrointestinal disorders in Mexico: a population-based study",

Gastroenterol. Res. Pract., vol. 2012, p. 606174, 2012, doi: 10.1155/2012/606174.
[27]A. Berkowitz, S. M. Fiorino, and G. S. Ginsberg, "Tobacco use and gastrointestinal disorders", *Current Opinion in Gastroenterology*, vol. 34, no. 6, pp. 433-438, Nov. 2018. doi: 10.1097/MOG.0000000000000468
[28]E. Mutlu et al., "Intestinal dysbiosis: a possible mechanism of alcohol-induced endotoxemia and alcoholic steatohepatitis in rats", *Alcoholism: Clinical and Experimental Research*, vol. 33, no. 10, pp. 1836-1846, Oct. 2009. doi: 10.1111/j.1530-0277.2009.01022.x
[29]J. K. Johannesson et al., "Physical activity improves symptoms in irritable bowel syndrome: a randomized controlled trial", *American Journal of Gastroenterology*, vol. 106, no. 5, pp. 915-922, May 2011. doi: 10.1038/ajg.2010.480
[30]S. S. C. Rao et al., "Evaluation of gastrointestinal transit in clinical practice: position paper of the American and European Neurogastroenterology and Motility Societies", *Neurogastroenterology & Motility*, vol. 23, no. 1, pp. 8-23, Jan. 2011. doi: 10.1111/j.1365-2982.2010.01612.x
[31]Y.-N. Yan, J.-Q. Zeng, y X. Ding, "Artificial intelligence in functional gastrointestinal disorders: From precision diagnosis to preventive healthcare", *Artif. Intell. Gastroenterol*., vol. 7, núm. 1, 2026, doi: 10.35712/aig.v7.i1.112357.
[32]World Health Organization, "Global action plan for the prevention and control of noncommunicable diseases 2013-2020", Geneva, Switzerland, 2013