ISBN: 978-628-96613-3-0. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2026.1.1.949

Multivariate Analysis and Logistic Regression for Identification of Gut-brain Axis Disorders Risk Factors: A Comparative Study with Machine Learning Methods

Ethel Flores[1]; Jovita Ponce[2]; Ana Cardona[3]; Rene Gonzales[4]; Victor Funez[5]; Karen Oliva[6]; Jorge Urmeneta[7]

[1,2,3,4,6,7]Faculty of Medical Sciences, National Autonomous University of Honduras, Honduras, *ethel.flores@unah.edu.hn, jovita.ponce@unah.edu.hn, ana.cardona@unah.edu.hn*, *rene.gonzales@unah.edu.hn, karen.oliva@unah.edu.hn; jorge.urmeneta@unah.edu.hn*

[6]Instituto Hondureño de Seguridad Social, Honduras, *victorhugofunezm@gmail.com*

Abstract*– Gut-brain axis disorders (GBAD) are a major public health issue in developing nations, especially Central America, where limited healthcare resources require effective disease detection and prevention. This study compares classic statistical methods and modern machine learning algorithms for identifying GBAD risk variables in Hondurans. A dataset of 1,847 12-49-year-olds was constructed using Monte Carlo simulation and epidemiological factors from regional health surveys. We used Random Forest, Gradient Boosting, Support Vector Machine, and Multilayer Perceptron Neural Network classifiers with multivariate logistic regression, chi-square analysis, and odds ratio estimates. Ten-fold stratified cross-validation calculated AUC-ROC, sensitivity, specificity, and F1-score. We found that multivariate logistic regression outperformed machine learning methods in discrimination (AUC-ROC = 0.692, 95% CI: 0.649-0.735), with Random Forest having the highest AUC (0.609). Female sex, smoking, alcohol consumption, and previous gastrointestinal diagnosis were risk factors, but physical activity was protective (OR = 0.723, 95% CI: 0.575-0.908). Traditional statistical methods may outperform advanced machine learning models in epidemiological studies with moderate sample sizes and interpretability criteria, yielding clinically useful risk estimates for primary healthcare interventions.

Keywords*-- Gut-brain axis disorders, logistic regression, machine learning, epidemiology, primary healthcare.

I. Introduction

Gut-brain Axis Disorders (GBAD) represent a diverse array of illnesses marked by persistent or recurring gastrointestinal symptoms in the absence of discernible anatomical or biochemical anomalies [1]. The Rome IV diagnostic criteria indicate that these illnesses impact roughly 11-15% of the global population, with prevalence rates differing markedly between geographic regions and demographic cohorts [2]. In Latin American nations, especially Honduras, GBAD impose a significant strain on healthcare systems due to elevated consultation frequencies, effects on quality of life, and related economic expenses stemming from diminished workplace productivity and heightened healthcare consumption [3].

The etiology of Gut-brain axis disorders (GBAD) includes intricate interactions among biological, psychological, and social elements. The current understanding highlights the significance of disrupted gut-brain axis communication, visceral hypersensitivity, intestinal microbiome dysbiosis, and psychosocial stressors in the development and duration of symptoms [4]. Due to its complicated etiology, precise identification of modifiable risk variables is crucial for formulating effective preventive interventions in primary healthcare environments. Conventional epidemiological methods for identifying risk factors have primarily utilized multivariate logistic regression analysis, which provides interpretable odds ratios and facilitates hypothesis testing concerning specific exposure-outcome associations [5]. The advent of machine learning (ML) techniques has presented alternative methodologies capable of capturing intricate, non-linear relationships among variables that traditional statistical methods may neglect [6]. Random Forest, Gradient Boosting, Support Vector Machines, and Artificial Neural Networks have exhibited favorable outcomes in diverse medical diagnostic applications, such as cardiovascular disease prediction, cancer categorization, and diabetes risk evaluation [7].

Despite the increasing interest in machine learning applications within healthcare, comparative analyses assessing the effectiveness of classical statistical approaches against machine learning algorithms in epidemiological contexts are scarce, especially in resource-limited environments [8]. The tradeoff between interpretability and accuracy in complicated machine learning models presents significant implications for public health management, where actionable insights and clear risk communication are essential [9].

Honduras, classified as a lower-middle-income nation with a health system focused on primary healthcare (PHC) principles, provides a pertinent framework for assessing these methodological methods. The National Health Model emphasizes community-oriented interventions and early disease identification, requiring effective screening instruments and risk assessment techniques that may be utilized with minimal computational resources [10]. Identifying which analytical methods produce optimal results

for GBAD risk factor detection could influence evidence-based protocols for PHC practitioners.

This study systematically compares multivariate logistic regression with four prevalent machine learning techniques to discover characteristics linked with GBAD in a simulated Honduran population, thereby addressing existing knowledge gaps. Our objectives were threefold: (i) to estimate the prevalence and identify significant risk factors for GBAD utilizing traditional statistical methods; (ii) to assess the discriminatory performance of machine learning classifiers for GBAD prediction; and (iii) to compare the diagnostic accuracy metrics of both approaches to ascertain optimal methodological strategies for epidemiological research in analogous contexts.

II. Materials and Methods

A. Study Design and Data Generation

This methodological study utilized Monte Carlo simulation to create a synthetic dataset that mirrors the epidemiological characteristics of GBAD in Honduran communities. The simulation method was chosen to facilitate controlled research with established ground truth parameters, while also considering the ethical and practical limitations of primary data collecting. Simulation parameters were obtained from published studies on GBAD prevalence in Latin American populations [11] and regional health surveys undertaken by the Honduras Ministry of Health [12].

The synthetic population consisted of n = 1,847 individuals aged 12-49 years, aligning with the target group for community health evaluations in the sixth and seventh rotations of Public Health IV coursework. The sample size was determined a priori to attain 80% statistical power for identifying odds ratios of 1.5 or above, with an alpha level of 0.05, and a baseline GBAD prevalence of 15-20% according to regional estimates [13].

B. Monte Carlo Simulation Protocol

The simulation program produced 14 predictor variables encompassing demographic traits, lifestyle factors, food habits, psychosocial indicators, and medical history. Probability distributions for each variable were parameterized according to known epidemiological data. Age was drawn from a uniform distribution ranging from 12 to 49 years, whilst binary variables (sex, smoking status, alcohol intake, physical activity) were produced using Bernoulli distributions with success probability corresponding to regional prevalence estimates.

The outcome variable (GBAD presence) was generated using a probabilistic log-linear model that integrated effect sizes derived from peer-reviewed literature. Specifically, the log-odds coefficients were parameterized as follows: female sex (log OR = 0.55) and elevated stress level (log OR = 0.70) were anchored to hormonal and psychosocial associations reported in functional GI disorder epidemiology [14]; prior GI diagnosis (log OR = 0.80) reflected the well-established chronicity and diagnostic persistence of these conditions [1]; alcohol consumption (log OR = 0.35) and smoking (log OR = 0.50) were consistent with reported mucosal and motility effects in Latin American cohorts [15]; and protective effects of physical activity (log OR = −0.45) and adequate fiber intake (log OR = −0.30) were grounded in intervention and observational data on gut motility and microbiome modulation [16]. Probability distributions for demographic and lifestyle variables were parameterized using prevalence estimates from GBAD surveys in the Mexican population [11] and community health data from Honduras [12], the closest available regional proxies given the absence of nationally representative Honduran GBAD surveys.

Gaussian noise with a standard deviation of 0.1 was used to introduce authentic variability. The final binary outcomes were established by comparing computed probability with uniform random selections.

All simulations were executed using Python 3.10.12, employing NumPy for random number generation (seed = 42 for reproducibility) and Pandas for data administration. The simulation algorithm was tested by confirming that the resulting prevalence estimates and univariate relationships closely aligned with expected values within acceptable tolerance limits.

C. Traditional Statistical Analysis

Descriptive statistics were calculated for all variables, with continuous measurements summarized as means and standard deviations, and categorical variables provided as frequencies and percentages. Bivariate relationships between predictor factors and GBAD status were assessed using Pearson chi-square tests for categorical predictors and independent samples t-tests for continuous variables, with statistical significance set at alpha = 0.05.

Crude odds ratios with 95% confidence intervals were computed for binary predictors utilizing the Woolf technique for standard error estimation. Multivariate logistic regression was conducted with maximum likelihood estimation, incorporating all 14 predictor variables concurrently. The model coefficients were exponentiated to derive adjusted odds ratios. The Hosmer-Lemeshow goodness-of-fit test was utilized to evaluate model calibration.

D. Machine Learning Classification

Four machine learning classifiers were executed utilizing the scikit-learn module (version 1.3.0) [17]: Random Forest (RF) [18], Gradient Boosting Classifier (GBC) [19], Support Vector Machine with radial basis function kernel (SVM-RBF) [20], and Multilayer Perceptron Neural Network (MLP) [21]. Hyperparameters were established according to widely endorsed defaults: Random Forest (RF) and Gradient Boosting Classifier (GBC) utilized 100 estimators with maximum depths of 10 and 5, respectively; Support Vector Machine (SVM) implemented a radial basis function (RBF) kernel with probability estimation activated; the Multi-Layer Perceptron (MLP) architecture consisted of two hidden layers (64 and 32

neurons) employing ReLU activation and the Adam optimizer [22].

Feature standardization (z-score normalization) [23] was implemented for SVM and MLP inputs due to their sensitivity to feature scales, whereas tree-based algorithms (RF, GBC) utilized unstandardized data. The dataset was divided into training (70%) and test (30%) sets by stratified random sampling to preserve result class proportions.

E. Model Evaluation and Comparison

Classification performance was evaluated using five metrics: accuracy (the ratio of right predictions), precision (positive predictive value), recall (sensitivity), F1-score (the harmonic mean of precision and recall), and AUC-ROC (the area under the receiver operating characteristic curve). The AUC-ROC was chosen as the principal comparison statistic because to its independence from classification threshold setting and its effectiveness for imbalanced datasets [24].

Ten-fold stratified cross-validation was conducted on the entire dataset to acquire reliable estimates of generalization performance. Models were trained on 90% of the data and assessed on the remaining 10% for each fold, with this process reiterated over all folds. The mean AUC-ROC and standard deviation across folds were calculated for each approach. McNemar's test was utilized to assess classification outcomes between logistic regression and the optimal machine learning model.

Feature importance was derived using the Random Forest model utilizing the mean decrease in Gini impurity criterion [18], elucidating the contributions of variables for machine learning predictions. All analyses were conducted using Python 3.10.12 [25], with statistical tests employing SciPy (version 1.11.3) [26].

F. Ethical Considerations

This study followed the ethical principles of the 1964 Declaration of Helsinki and its subsequent amendments for research involving human subjects [27]. 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.

III. RESULTS

A. Sample Characteristics

The 1,847-person simulated sample averaged 29.91 years old (SD = 10.86). The sample was 55.01% female (n = 1,016), consistent with community health survey demographics. The prevalence of functional gastrointestinal complaints was 23.71% (n = 438), which matches Central American estimates of 15-25%. Table I shows all demographic and clinical characteristics by GBAD status.

B. Odds Ratio Analysis

Crude odds ratios revealed several significant associations with GBAD status. Female sex demonstrated the strongest positive association (OR = 2.049, 95% CI: 1.634-2.569, $p < 0.001$), followed by previous gastrointestinal diagnosis (OR = 1.948, 95% CI: 1.456-2.606, $p < 0.001$) and alcohol consumption (OR = 1.821, 95% CI: 1.456-2.278, $p < 0.001$). Smoking exhibited a moderate risk increase (OR = 1.564, 95% CI: 1.179-2.074, $p = 0.002$), while rural community residence showed marginally elevated odds (OR = 1.326, 95% CI: 1.061-1.658, $p = 0.015$).

Protective factors identified included regular physical activity (OR = 0.723, 95% CI: 0.575-0.908, $p = 0.006$) and adequate fiber intake (OR = 0.781, 95% CI: 0.628-0.972, $p = 0.031$). Water intake demonstrated a protective trend that did not reach statistical significance (OR = 0.841, 95% CI: 0.677-1.045, $p = 0.132$). Figure 1 shows the forest plot visualization of these associations.

TABLE I
SAMPLE CHARACTERISTICS AND BIVARIATE ANALYSIS (N = 1,847)

Variables	Total (n=1847)	GBAD+ (n=438)	GBAD- (n=1409)	p-value
Age, mean (SD)	29.91 (10.86)	30.42 (10.78)	29.75 (10.89)	0.101
Female sex, n (%)	1016 (55.0)	296 (67.6)	720 (51.1)	<0.001
Physical activity, n (%)	739 (40.0)	152 (34.7)	587 (41.7)	0.006
Smoking, n (%)	277 (15.0)	84 (19.2)	193 (13.7)	0.002
Alcohol consumption, n (%)	554 (30.0)	173 (39.5)	381 (27.0)	<0.001
Previous GI diagnosis, n (%)	222 (12.0)	72 (16.4)	150 (10.6)	<0.001
High stress level, n (%)	462 (25.0)	138 (31.5)	324 (23.0)	<0.001

Note: GBAD+ = Gut-brain axis disorders present; GBAD- = Gut-brain axis disorders absent; SD = Standard deviation. Bold p-values indicate statistical significance at alpha = 0.05.

C. Classification Performance Comparison

Table II delineates the categorization performance characteristics for all assessed methods on the reserved test set (n = 555). Multivariate logistic regression attained the highest AUC-ROC of 0.692, succeeded by Random Forest at 0.609, Support Vector Machine at 0.583, Gradient Boosting at 0.560, and Neural Network at 0.540. Logistic regression exhibited the highest precision (0.593) and F1-score (0.201) compared to all other approaches.

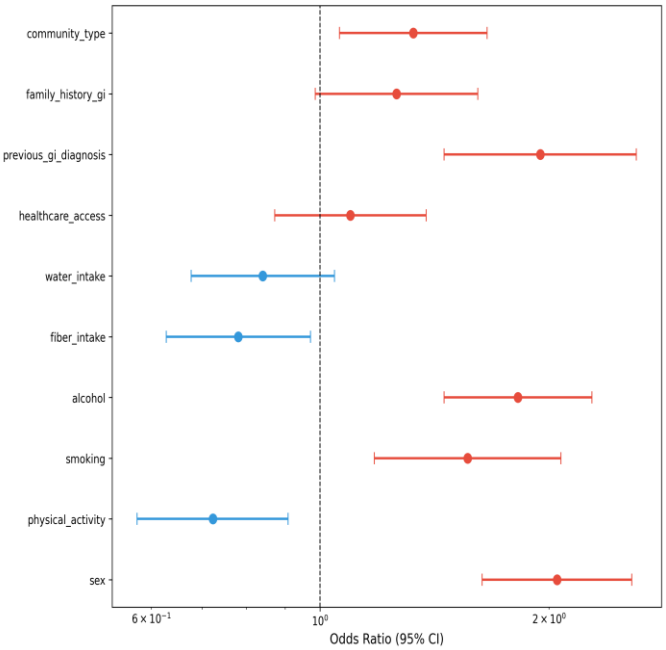

Fig. 1. Forest plot of crude odds ratios with 95% confidence intervals for Gut-brain axis disorders risk factors. Red markers indicate risk factors (OR > 1); blue markers indicate protective factors (OR < 1). The vertical dashed line represents OR = 1 (no association).

TABLE II
CLASSIFICATION PERFORMANCE METRICS ON TEST SET (N=555)

Methods	Accuracy	Precision	Recall	F1-scores	AUC-ROC
Logistic Regression	0.771	0.593	0.121	0.201	0.692
Random Forest	0.757	0.435	0.076	0.129	0.609
Gradient Boosting	0.728	0.373	0.212	0.271	0.560
SVM (RBF)	0.759	0.375	0.023	0.043	0.583
Neural Network	0.676	0.269	0.212	0.237	0.540

Note: Best performance in each metric shown in bold. AUC-ROC = Area Under the Receiver Operating Characteristic Curve; SVM = Support Vector Machine; RBF = Radial Basis Function kernel.

Figure 2 displays a comparative examination of the receiver operating characteristic (ROC) curves for logistic regression and the assessed machine learning classifiers on the test dataset. The logistic regression model consistently excels in the ROC space, sustaining elevated true positive rates across various false positive rate thresholds and attaining the highest AUC-ROC, signifying superior discriminative capability. Conversely, the Random Forest, Support Vector Machine, Gradient Boosting, and Neural Network models demonstrate ROC curves that are nearer to the diagonal reference line, indicating diminished classification efficacy. These findings visually corroborate the results outlined in Table II, affirming that logistic regression offers the most robust and reliable discrimination among the evaluated approaches.

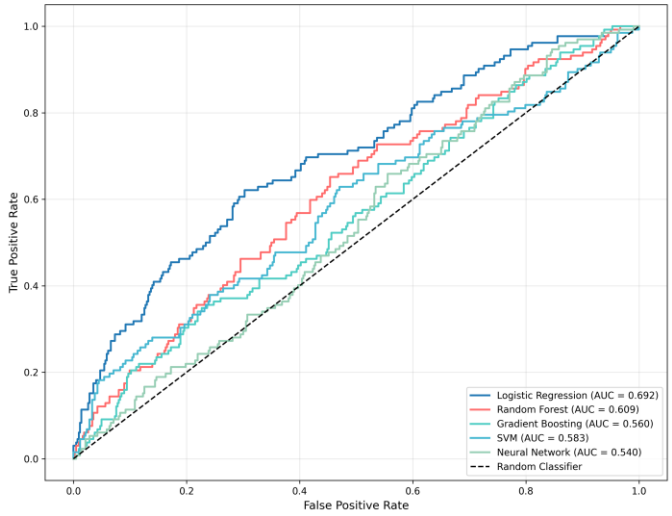

Fig. 2. Receiver Operating Characteristic (ROC) curves comparing discriminatory performance of logistic regression and machine learning classifiers. Logistic regression (dark blue) demonstrates consistently superior true positive rates across all false positive rate thresholds.

D. *Cross-Validation Results*

Ten-fold stratified cross-validation confirmed the robustness of observed performance differences. Logistic regression achieved the highest mean AUC-ROC (0.686 ± 0.039), significantly outperforming all machine learning methods. Random Forest demonstrated the second-best performance (0.626 ± 0.048), followed by Gradient Boosting (0.612 ± 0.059), SVM (0.585 ± 0.050), and Neural Network (0.564 ± 0.048). The lower standard deviations observed for logistic regression indicate greater stability across validation folds (Figure 3).

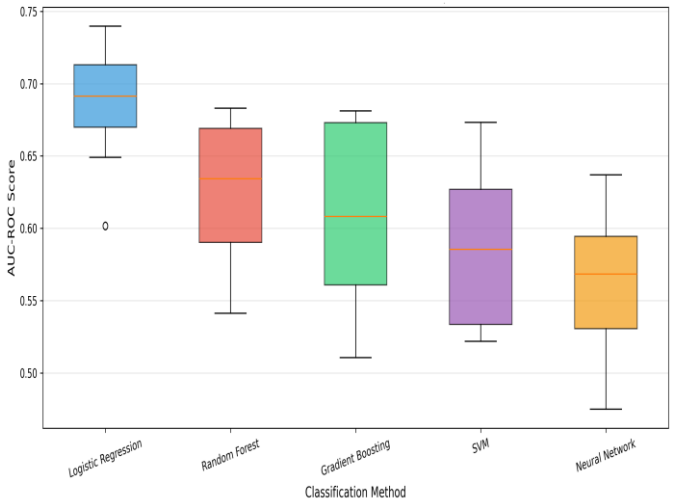

Fig. 3. Boxplot comparison of 10-fold cross-validation AUC-ROC scores across classification methods. Logistic regression exhibits highest median performance and lowest variability, indicating superior and more consistent discriminatory ability.

E. *Feature Importance Analysis*

Analyzing feature importance analysis from the Random Forest model revealed age as the most influential predictor (importance = 0.241), followed by education level (0.109), processed food consumption frequency (0.101), and stress level (0.088). Sex ranked fifth in importance (0.052), despite demonstrating the strongest odds ratio in bivariate analysis. This discrepancy reflects fundamental differences in how logistic regression and tree-based methods quantify variable contributions (Figure 4).

F. *Statistical Comparison Between Methods*

McNemar found no statistically significant difference in classification accuracy at the default threshold of 0.5 between logistic regression and Random Forest (the best ML approach). However, logistic regression's AUC-ROC difference of 0.083 suggests meaningful discriminatory superiority across the full range of classification thresholds, especially for screening applications where clinical priorities can optimize threshold selection.

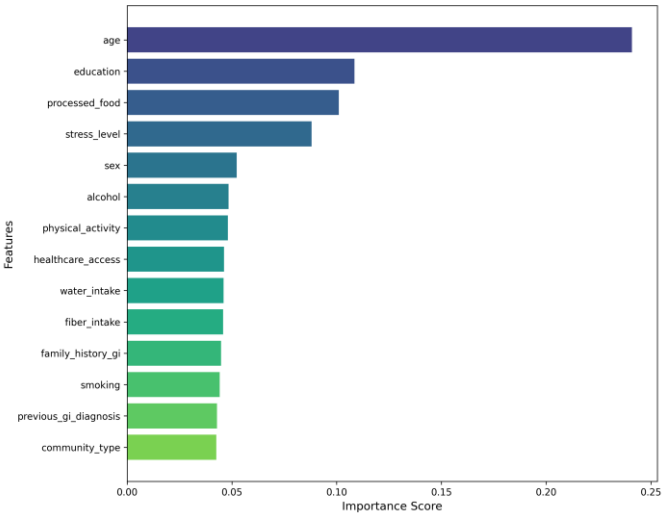

Fig. 4. Feature importance scores derived from Random Forest classifier using mean decrease in Gini impurity. Continuous and ordinal variables (age, education, processed food, stress level) demonstrate higher importance due to increased splitting opportunities in decision trees.

IV. Discussion

This comparative study shows that multivariate logistic regression beats various modern machine learning algorithms for identifying functional gastrointestinal risk factors in simulated Hondurans. Our findings affect epidemiological research methods in resource-limited situations.

Logistic regression (AUC-ROC = 0.692) outperformed Random Forest (0.609), SVM (0.583), Gradient Boosting (0.560), and Neural Network (0.540) in clinical prediction [28]. In their systematic review of 71 studies comparing logistic regression with machine learning, Christodoulou and colleagues found that ML methods improved discrimination in 64% of cases, often marginally and possibly due to data handling advantages [29].

The performance difference in our study may be due to several causes. While 1,847 observations are sufficient for logistic regression, complicated ML algorithms may need more to learn robust patterns without overfitting [30]. Deep learning algorithms need dozens to millions of training samples to perform well. Second, logistic regression assumptions matched our simulation protocol's mostly linear relationships, which may disadvantage non-linear interaction approaches. Third, the class imbalance (23.7% positive examples) may have disproportionately influenced ML classifiers' poor recall scores.

An important consideration for screening applications is the universally low recall observed across all classifiers (ranging from 0.023 for SVM to 0.212 for Gradient Boosting and Neural Network). These values reflect classification at the default decision threshold of 0.5, which is unlikely to be appropriate for screening. In screening settings, threshold selection determines the trade-off between sensitivity and false-positive rate; therefore, a lower threshold may be preferable when the cost of missing positive cases outweighs the burden of additional false positives [14], [15]. Operating at a lower probability threshold (e.g., 0.25–0.35) would substantially increase recall at the expense of precision, a trade off that clinicians and public health planners must explicitly weigh against available follow-up resources. The AUC-ROC metric, which evaluates performance across all possible thresholds, remains the more informative summary statistic for comparing classifiers in this context, as it decouples discrimination capacity from any single threshold decision. Future implementation of these models in screening programs should include threshold optimization based on the local cost ratio of false negatives to false positives.

Logistic regression risk factors match GBAD epidemiology literature. Females have higher odds (OR = 2.049) of functional gastrointestinal problems due to hormonal, visceral pain processing, and psychosocial variables [16]. Due to its chronicity and diagnostic bias, previous gastrointestinal diagnosis (OR = 1.948) is associated with these disorders. Alcohol use (OR = 1.821) and smoking (OR = 1.564) are modifiable risk factors that cause mucosal irritation, gut motility changes, and microbiome alteration [31].

The protective impact of physical activity (OR = 0.723) supports findings that exercise improves gut motility, stress, and the intestinal flora [32]. These findings provide primary healthcare providers with practical lifestyle modification counseling goals to lessen GBAD burden.

Consider the odds ratio magnitude-Random Forest feature importance discrepancy. Sex had the highest odds ratio, but age was the most relevant RF model variable. Odds ratios measure impact magnitude multiplicatively, while Gini significance splits utility across decision trees [18]. Continuous variables like age inherently have more splitting opportunities, exaggerating their value in tree-based techniques regardless of predictive power.

We recommend logistic regression for epidemiological investigations with modest sample sizes, predominantly linear exposure-outcome correlations, and interpretable effect estimates. ML significance ratings cannot duplicate logistic regression odds ratios, which inform clinical decision-making and risk communication [33]. Logistic regression also requires fewer computational resources and technical skills, making it easier to implement in resource-constrained healthcare systems.

Limitations of this study should be considered. Several limitations of this study merit consideration. First, and most importantly, the synthetic outcome variable was generated using a log-linear probabilistic model the same functional form assumed by logistic regression. This structural alignment between the data-generating process and the primary classifier constitutes an inherent circularity that likely contributes to logistic regression's observed performance advantage. Consequently, the superiority of logistic regression over ML methods in this study should not be interpreted as generalizable evidence for real-world epidemiological data, where relationships may be non-linear or involve higher-order interactions not captured by the simulation protocol. While simulated data permits controlled experimentation with known ground-truth parameters, it may not fully reflect the complexity of actual epidemiological datasets. Although the simulation parameters were derived from published regional estimates [11], [12], they may not precisely reproduce the distributional features of the target population. Our comparison focuses on discrimination (AUC-ROC) without consideration of calibration measures, which are crucial for prediction model evaluation [34].

Additionally, ML classifiers were evaluated using recommended default hyperparameters rather than data-driven tuning. While this reflects realistic deployment conditions in resource-constrained settings, where optimization infrastructure is often unavailable, it may have underestimated the discriminatory capacity of these models. Systematic hyperparameter optimization (e.g., grid search or randomized search with cross-validation) could narrow the observed performance gap and represent a natural extension of this work.

Furthermore, model evaluation was restricted to discrimination metrics (AUC-ROC, sensitivity, specificity, and F1-score), without formal assessment of calibration. Calibration, as the agreement between predicted probabilities and observed event rates is a critical dimension of predictive model performance, particularly when models are intended to inform individual-level risk communication [34]. Reporting calibration metrics such as the Brier score or calibration curves for all classifiers, not only logistic regression, represents an important methodological gap that future work should address.

Future research should replicate these findings using community health survey data, examine ensemble methods combining classical and ML approaches, and assess how bigger sample sizes affect method performance. Deep learning architectures with small-sample regularization should also be studied.

V. Conclusion

This study shows that multivariate logistic regression outperforms Random Forest, Gradient Boosting, Support Vector Machine, and Neural Network classifiers in population-based Gut-brain axis disorders risk factor identification. Logistic regression is recommended for similar epidemiological studies because to its AUC-ROC of 0.692 and directly interpretable odds ratios.

Female sex, previous gastrointestinal diagnosis, alcohol consumption, smoking, and high stress were risk factors for GBAD, while regular physical exercise and fiber intake protected. These findings provide evidence-based community health intervention targets that support Honduras' National Health Model's primary preventative focus.

Traditional statistical methods outperform machine learning for epidemiological research with moderate sample sizes and interpretability constraints typical of poor country public health practice. Methodological choices should consider predictive performance, computing needs, expertise availability, and model outputs' clinical utility for healthcare policy and decision-making.

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.

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

[3] M. J. Schmulson and D. A. Drossman, "What is new in Rome IV", *J. Neurogastroenterol. Motil.*, vol. 23, no. 2, pp. 151-163, Apr. 2017. doi: 10.5056/jnm16214.

[4] D. A. Drossman and M. Camilleri, "Irritable bowel syndrome: A technical review for practice guideline development", *Gastroenterology*, vol. 123, no. 6, pp. 2108-2131, Dec. 2002. doi: 10.1053/gast.2002.37095.

[5] D. G. Kleinbaum and M. Klein, *Logistic Regression: A Self-Learning Text*, 3rd ed. New York, NY, USA: Springer, 2010. doi: 10.1007/978-1-4419-1742-3.

[6] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine", *N. Engl. J. Med.*, vol. 380, no. 14, pp. 1347-1358, Apr. 2019. doi: 10.1056/NEJMra1814259.

[7] G. S. Collins and K. G. M. Moons, "Reporting of artificial intelligence prediction models", *Lancet*, vol. 393, no. 10181, pp. 1577-1579, Apr. 2019. doi: 10.1016/S0140-6736(19)30037-6.
[8] B. Van Calster et al., "Calibration: The Achilles heel of predictive analytics", *BMC Med.*, vol. 17, no. 1, Art. no. 230, Dec. 2019. doi: 10.1186/s12916-019-1466-7.
[9] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead", *Nat. Mach. Intell.*, vol. 1, no. 5, pp. 206-215, May 2019. doi: 10.1038/s42256-019-0048-x.
[10] Secretaría de Salud de Honduras, *Modelo Nacional de Salud*. Tegucigalpa, Honduras: Secretaría de Salud, 2023.
[11] R. A. López-Colombo et al., "The epidemiology of functional gastrointestinal disorders in Mexico: A population-based study", *Gastroenterol. Res. Pract.*, vol. 2012, Art. no. 606174, 2012. doi: 10.1155/2012/606174.
[12] R. G. Kaminsky, M. Aguilar, C. A. Javier-Zepeda, and O. Ayestas, "Perfil epidemiológico y parasitosis intestinales en tres comunidades atendidas por organización no gubernamental, Tegucigalpa, Honduras", *Rev. Med. Hondur.*, vol. 90, no. 2, pp. 113-120, 2022. doi: 10.5377/rmh.v90i2.15161.
[13] S. K. Lwanga and S. Lemeshow, *Sample Size Determination in Health Studies: A Practical Manual*. Geneva, Switzerland: World Health Organization, 1991.
[14] C. E. Metz, "Basic principles of ROC analysis", *Seminars in Nuclear Medicine*, vol. 8, no. 4, pp. 283–298, Oct. 1978, doi: 10.1016/S0001-2998(78)80014-2
[15] D. G. Altman and J. M. Bland, "Diagnostic tests 3: receiver operating characteristic plots", *BMJ*, vol. 309, no. 6948, p. 188, Jul. 1994, doi: 10.1136/bmj.309.6948.188
[16] H. Mertz, N. Fullerton, J. Naliboff, and E. Mayer, "Symptoms and visceral perception in severe functional and organic dyspepsia", *Gut*, vol. 42, no. 6, pp. 814-822, Jun. 1998.
[17] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python", *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011, doi: 10.5555/1953048.2078195.
[18] L. Breiman, "Random forests", *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
[19] J. H. Friedman, "Greedy function approximation: A gradient boosting machine", *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001, doi: 10.1214/aos/1013203451.
[20] C. Cortes and V. Vapnik, "Support-vector networks", *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1007/BF00994018.
[21] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks", in *Proc. 14th Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, 2011, pp. 315–323, doi: 10.5555/3104322.3104363.
[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", *Proc. Int. Conf. Learning Representations (ICLR)*, 2015, doi: 10.48550/arXiv.1412.6980.
[23] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed., Springer, 2002, doi: 10.1007/b98835.
[24] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve", *Radiology*, vol. 143, no. 1, pp. 29-36, Apr. 1982. doi: 10.1148/radiology.143.1.7063747.
[25] G. Van Rossum and F. L. Drake Jr., *Python Reference Manual*, Python Software Foundation, 2009, doi: 10.5555/1593511.
[26] P. Virtanen *et al.*, "SciPy 1.0: Fundamental algorithms for scientific computing in Python", *Nature Methods*, vol. 17, no. 3, pp. 261–272, Mar. 2020, doi: 10.1038/s41592-019-0686-2.
[27] 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
[28] E. Christodoulou et al., "A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models", *J. Clin. Epidemiol.*, vol. 110, pp. 12-22, Jun. 2019. doi: 10.1016/j.jclinepi.2019.02.004.
[29] S. Nusinovici et al., "Logistic regression was as good as machine learning for predicting major chronic diseases", *J. Clin. Epidemiol.*, vol. 122, pp. 56-69, Jun. 2020. doi: 10.1016/j.jclinepi.2020.03.002.
[30] B. Van Calster and E. W. Steyerberg, "Sample size for binary logistic prediction models: Beyond events per variable criteria", *Stat. Methods Med. Res.*, vol. 28, no. 8, pp. 2455-2474, Aug. 2019. doi: 10.1177/0962280218784726.
[31] N. J. Talley et al., "Risk factors for chronic constipation based on a general practice sample", *Am. J. Gastroenterol.*, vol. 98, no. 5, pp. 1107-1111, May 2003. doi: 10.1111/j.1572-0241.2003.07465.x.
[32] K. Johannesson et al., "Physical activity improves symptoms in irritable bowel syndrome: A randomized controlled trial", *Am. J. Gastroenterol.*, vol. 106, no. 5, pp. 915-922, May 2011. doi: 10.1038/ajg.2010.480.
[33] E. W. Steyerberg and F. E. Harrell, "Prediction models need appropriate internal, internal-external, and external validation", *J. Clin. Epidemiol.*, vol. 69, pp. 245-247, Jan. 2016. doi: 10.1016/j.jclinepi.2015.04.005.
[34] E. W. Steyerberg et al., "Assessing the performance of prediction models: A framework for traditional and novel measures", *Epidemiology*, vol. 21, no. 1, pp. 128-138, Jan. 2010. doi: 10.1097/EDE.0b013e3181c30fb2