KMAN PUBLISHING INSTITUTE
Journal Website

Article history:
Received 12 April 2026
Revised 04 June 2026
Accepted 13 June 2026
Initial Publish 22 June 2026
Final Publish 01 July 2026

Journal of Personality and Psychosomatic Research

Open Peer-Review Report

E-ISSN: 3041-8542

Multimodal Machine-Learning Prediction of Eating-Related Psychosomatic Risk Based on Impulsivity, Interoceptive Deficits, Emotional Eating, and Body Mass Index

Cynthia. Maldonado[1], Estefania. Villarreal-Garza[1], Jesica. Muniz-Laura[2*]

[1] Institute of Psychological Research, Universidad Veracruzana, Av. Dr. Luis Castelazo Ayala s/n, Col. Industrial Ánimas, 91190, Xalapa, VER, Mexico
[2] National Institute of Psychiatry Ramón de la Fuente Muñiz, Sub-directorate of Clinical Research, Mexico City, Mexico

*** Corresponding author email address**: nesicalaura1983@gmail.com

1. Round 1

1.1. Reviewer 1

Reviewer:

In the Data Collection Tools section, the manuscript refers to “the Spanish-language version of the Barratt Impulsiveness Scale–11,” “the Interoceptive Deficits subscale of the Eating Disorder Inventory–3,” and the “Spanish-language Dutch Eating Behavior Questionnaire,” but does not identify the specific validated versions used in Mexico. Please report the translators or validation studies, response options, scoring ranges, reverse-scored items, and evidence of reliability and construct validity in Mexican or Latin American populations. If a Mexican validation was unavailable, the authors should describe the cultural-adaptation process in greater detail and report whether factorial validity and measurement invariance were tested in the present sample.

The outcome definition requires stronger conceptual and empirical justification. The manuscript states that participants were classified as elevated risk when they "simultaneously obtained an Eating Attitudes Test–26 score of 20 or higher and a Patient Health Questionnaire–15 score of 10 or higher." This conjunctive rule creates a novel composite endpoint, but no evidence is presented that the combination validly represents "eating-related psychosomatic risk." Please explain why an AND rule was selected rather than an OR rule, weighted composite, latent-variable score, or clinician-confirmed classification. Sensitivity analyses using alternative thresholds and outcome definitions should be reported to determine whether model performance depends heavily on this operational decision.

Table 2 reports point estimates and a 95% confidence interval only for ROC–AUC, whereas sensitivity, specificity, predictive values, F1 score, balanced accuracy, PR–AUC, and Brier score are presented without uncertainty. Given that the test set contained only 34 elevated-risk cases, these estimates may be unstable. Please provide bootstrap confidence intervals for all primary performance measures. The manuscript should also report the test-set confusion matrix for each major model, or at least for the benchmark and final model, to make the derivation of sensitivity, specificity, positive predictive value, and negative predictive value transparent.

The claim that the early-fusion model "significantly outperformed conventional logistic regression, $p = .008$, and elastic-net logistic regression, $p = .029$" requires specification of the statistical test and correction for multiple model comparisons. Please identify whether DeLong's test or a bootstrap procedure was used, confirm that predictions were paired on the same test participants, and report confidence intervals for the difference in ROC–AUC. Because several algorithms were compared, the authors should address multiplicity or clearly identify these comparisons as exploratory. It would also be informative to compare PR–AUC and calibration, not only ROC–AUC, because the outcome prevalence was approximately 21%.

Authors revised the manuscript and uploaded the document.

1.2. Reviewer 2

Reviewer:

The Methods paragraph describing the continuous risk index states that the standardized Eating Attitudes Test–26 and Patient Health Questionnaire–15 scores were averaged. This assumes equal weighting and comparable construct relevance, although the two instruments measure distinct domains. Please provide a psychometric rationale for combining them, report the correlation between the component scores, and consider confirmatory factor analysis or latent-variable modeling to establish whether they support a common higher-order construct. At minimum, the manuscript should clarify that this index is an exploratory severity measure rather than a validated psychosomatic-risk scale.

The predictor construction may introduce redundancy and increase overfitting. The manuscript states that "Both the total score and the three dimensional scores were retained as candidate model features," and that item-level responses were additionally included in some models. Total scores, subscale scores, and item scores are mathematically dependent and highly collinear. Please specify exactly which representation was used in each algorithm and avoid entering totals, subscales, and their constituent items simultaneously unless a compelling rationale and regularization strategy are provided. A supplementary feature map should list every predictor, its coding, and the model in which it was used.

The preprocessing paragraph states that remaining missing continuous values were imputed using the median within each training fold. However, the extent and pattern of missingness are not reported. Please provide the percentage of missing data for every variable, the number of participants with any missing values, and evidence regarding whether data were plausibly missing completely at random, missing at random, or missing not at random. Median imputation may attenuate variance and distort associations, particularly in psychometric item data. Multiple imputation, iterative imputation, or model-native missing-value handling should be considered in sensitivity analyses.

The manuscript should provide substantially greater detail regarding the early-fusion neural network. The sentence "the psychometric inputs were processed through two fully connected layers with rectified linear activation, batch normalization, and dropout regularization" is insufficient for replication. Please report the exact number of input variables, units in each layer, activation functions, dropout probabilities, weight initialization, optimizer, learning rate, batch size, maximum epochs, early-

JPPR
E-ISSN: 3041-8542

stopping criterion, patience parameter, loss function, class weights, and number of training repetitions. The final selected hyperparameters should be reported, not only the search ranges.

The nested cross-validation procedure is conceptually appropriate, but the relationship between nested cross-validation, final model fitting, and the independent test set requires clarification. The paragraph stating that "The outer loop consisted of five folds" should explain whether outer-fold performance was used only for algorithm selection, whether hyperparameters were reoptimized on the full development dataset, and how the final model was refitted before test evaluation. Please report mean and standard deviation of performance across the outer folds for every candidate model. Presenting only test-set results does not reveal model stability during development.

Threshold selection requires further methodological precision. The manuscript states that the threshold was selected "to maximize the Youden index while maintaining a sensitivity of at least .80 whenever possible." These are potentially competing criteria, and "whenever possible" is not reproducible. Please specify the exact decision rule, selected threshold for each model, and whether the rule was applied within each cross-validation fold or after pooling development predictions. Because the intended use is screening, the authors should consider prespecifying a clinically acceptable false-negative rate and reporting decision-curve analysis across plausible thresholds.

Authors revised the manuscript and uploaded the document.

2. Revised

Editor's decision: Accepted.
Editor in Chief's decision: Accepted.