**Title**

Revisiting the Connection between Microbiota and Depressive Disorder- A Comparative Evaluation and Reproduction of Machine Learning Approaches

**Abstract**

Metagenomic analyses have increasingly implicated the gut microbiota in the etiology of major depressive disorder (MDD). Recent investigations have applied machine learning to classify individuals by depressive status based on metagenomic signatures, with one study comparing modest logistic regression performance to a high-performing convolutional neural network (YOLOv8) model. However, that study’s YOLOv8 results were obtained after training on a substantially augmented, synthetic dataset, whereas the logistic regression model was evaluated on a smaller real dataset without extensive parameter optimization.

In this work, the same original dataset was re-examined using a range of methods, including logistic regression, random forest, and multiple support vector machine (SVM) variants. Models were evaluated with thorough parameter tuning, cross-validation, and, when applicable, training on the synthetic dataset. Under these conditions, logistic regression performed notably better than previously reported, achieving a cross-validated accuracy of 0.74 and a ROC AUC of 0.81 on the original data. When trained on the synthetic dataset, logistic regression, random forest, and SVM models all reached near-perfect classification accuracy, thus approximating the YOLOv8 performance reported in the original study. These findings suggest that previously underappreciated classical models can achieve competitive results when given similar data conditions and appropriate optimization. The results underscore the importance of consistent evaluation frameworks and tuning methodologies to ensure fair comparisons among different machine learning approaches in the context of gut microbiome-based MDD classification.

**Introduction**

**Methods**

**Results**

**Discussion**

Comparing Results to Original Findings:

Discuss how your tuned logistic regression results (accuracy 0.74, ROC AUC 0.81) diverge from the originally reported performance (accuracy ~0.55).

Reflect on the improvements due to cross-validation, hyperparameter tuning, and bagging.

Impact of Synthetic Data Augmentation:

Examine how the introduction of a large synthetic dataset improved all models.

Highlight that logistic regression, random forest, and SVM achieved near-perfect classification, challenging the notion that YOLOv8’s success was unique to the CNN architecture.

Model Complexity Versus Performance:

Discuss that simpler models can match or exceed YOLOv8’s performance if given similar data conditions and proper tuning.

Explore reasons why classical ML models benefit from additional data and parameter optimization (e.g., better generalization, stable decision boundaries).

Interpretation of Features:

Consider how some models (e.g., linear models, random forest) allow better insight into feature importance.

Highlight that understanding which species-enzyme pairs contribute most significantly to classification could inform biological interpretations and potential biomarkers.

Methodological Considerations:

Address the importance of consistent validation strategies, dataset size, and hyperparameter search.

Note limitations: small initial sample size, reliance on synthetic data, and the need for external validation sets.

Broader Implications:

Place these findings in the context of gut microbiome research and mental health diagnostics.

Suggest that high-performance classification isn’t limited to deep learning methods when other factors are controlled.

**Conclusion**

**References**

1. Kovtun AS, Averina OV, Angelova IY, et al. *Alterations of the Composition and Neurometabolic Profile of Human Gut Microbiota in Major Depressive Disorder*. Biomedicines. 2022;10(9 :2162. Published 2022 Sep 2. <https://doi.org/10.3390/biomedicines10092162>

2. Angelova IY, Kovtun AS, Averina OV, Koshenko TA, Danilenko VN. *Unveiling the Connection between Microbiota and Depressive Disorder through Machine Learning.* International Journal of Molecular Sciences. 2023; 24(22):16459. <https://doi.org/10.3390/ijms242216459>

**Acknowledgements**

**Supplementary Materials**