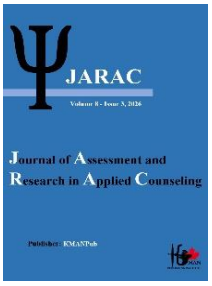
KMAN PUBLISHING INSTITUTE
Journal Website

Article history:
Received 16 March 2026
Revised 27 May 2026
Accepted 08 June 2026
Initial Publish 25 June 2026
Published online 01 July 2026

Journal of Assessment and Research in Applied Counseling

Open-Peer Review



E-ISSN: 3041-8518

Predicting Major Depressive Disorder Using Random Forest Models Based on Psychological, Behavioral, and Lifestyle Indicators

Liam. Wooderson[1], Milena. Fini[2*], Mariusz. Szypa[3]

[1] Department of Psychology, University of Toronto, Mississauga, ON L5L 1C6, Canada
[2] Department of Psychology, University of Quebec at Montreal, Montreal, QC, Canada
[3] School of Psychology, Australian Catholic University, Brisbane, Queensland, Australia

Editor	Reviewers
Azizuddin Khan Professor, Psychophysiology Laboratory, Department of Humanities and Social Sciences Indian Institute of Technology Bombay, Maharashtra, India aziz@hss.iitb.ac.in	**Reviewer 1:** Soumita Boral Department of Psychology, MDDM College, Babasaheb Bhimrao Ambedkar Bihar University, Muzaffarpur, Bihar, India. Email: Boral_Soumita@gmail.com **Reviewer 2:** Iheanyi Iwueze Department of Psychology, North-West University, Potchefstroom, South Africa. Email: taher. iheanyiiwueze@gmail.com

1. Round 1

1.1. Reviewer 1

Reviewer:

In the Introduction paragraph beginning “Traditional approaches to MDD assessment have relied primarily on clinical interviews, self-report questionnaires, and symptom-based diagnostic criteria,” the authors correctly identify limitations of conventional diagnostic methods. However, the paragraph should more clearly distinguish between clinical diagnosis and screening-based prediction. The study appears to classify probable MDD using PHQ-9 scores rather than structured diagnostic interviews; therefore, the introduction should avoid implying that the model predicts clinically confirmed MDD unless this is supported by diagnostic assessment.

In the Introduction paragraph discussing “psychological indicators,” the manuscript emphasizes negative affect, perceived stress, anxiety, and emotion regulation. This section would be stronger if the authors explicitly justify why each psychological variable was selected for the model. For example, the rationale for including DASS depression, DASS anxiety, DASS stress, perceived stress, and DERS-16 should be linked to distinct theoretical mechanisms rather than presenting them as a broad cluster of distress variables.

In the Introduction paragraph beginning “Behavioral and lifestyle indicators are also increasingly recognized as essential components of MDD risk and prognosis,” the manuscript presents an appropriate lifestyle psychiatry rationale. However, the

paragraph would benefit from a clearer transition to the specific lifestyle indicators included in this study. The authors should explain why screen time, physical activity, sleep duration, BMI, and alcohol consumption were selected, and whether diet, smoking, or substance use were ultimately retained in the predictive model.

In the Data Collection Tools paragraph describing lifestyle behaviors, the manuscript states that a "structured health behavior questionnaire" assessed smoking, alcohol, recreational drug use, dietary habits, caffeine consumption, screen time, social media use, working hours, and sleep duration. The authors should provide additional information on whether this questionnaire was validated, adapted from prior instruments, or developed for the present study. If it was researcher-made, the manuscript should report content validation, pilot testing, or reliability procedures.

In the Data Analysis section, the manuscript states that "missing values representing less than 5% of the dataset were addressed through multiple imputation using chained equations." The authors should clarify whether imputation was performed before or after train-test splitting. To avoid information leakage, imputation parameters should be learned only from the training set and then applied to the test set. This point is essential for the validity of machine learning performance estimates.

In the Data Analysis section, the manuscript states that "the dataset was randomly divided into training (80%) and testing (20%) subsets while preserving the proportional distribution of depressive and non-depressive cases through stratified sampling." This is appropriate, but the authors should report the exact number of depressed and non-depressed participants in both the training and test sets. This would make the evaluation more transparent and allow readers to assess whether class balance was maintained.

Authors revised and uploaded the document.

1.2. Reviewer 2

Reviewer:

In the final paragraph of the Introduction, the stated aim is clear: "The present study aimed to develop and evaluate a Random Forest classification model for predicting Major Depressive Disorder among Canadian adults using integrated psychological, behavioral, and lifestyle indicators." However, the aim should specify whether the objective was diagnostic classification, risk screening, or prediction of probable MDD based on symptom cut-off. This distinction is methodologically important because the study design is cross-sectional and therefore cannot predict future disorder onset.

In the Methods and Materials section, the Study Design and Participants paragraph states that "This study employed a cross-sectional, predictive machine learning design." This phrase is acceptable, but the authors should clarify that the model performs cross-sectional classification rather than prospective prediction. The term "predicting" may be retained in the title, but the methods should explicitly state that prediction refers to statistical classification of current probable MDD status.

In the Methods paragraph reporting recruitment, the manuscript states that participants were recruited "through community health centers, university mailing lists, workplace wellness programs, and online advertisements." The authors should provide more detail about the recruitment distribution across these sources. Without this information, it is difficult to assess selection bias, especially because online recruitment and university mailing lists may overrepresent younger, educated, digitally engaged participants.

In the Participants paragraph, the manuscript reports that "the final analytical sample consisted of 1,742 participants." This is a strength, but the authors should include a participant flow description in the text, specifying the number excluded for incomplete questionnaires, duplicate submissions, and failed attention checks. This would improve transparency and allow readers to evaluate whether exclusions may have introduced systematic bias.

In the Participants paragraph, the authors state that participants were recruited "across multiple provinces in Canada," but the exact provincial distribution is not reported. Since geographic diversity is used to support the representativeness of the sample, the manuscript should report the number or percentage of participants from major provinces or regions. This is particularly important if the authors claim applicability to Canadian adults broadly.

JARAC Assessment and Research in Applied Counseling
E-ISSN: 3041-8518

In the Data Collection Tools section, the manuscript states that "participants scoring 10 or higher were categorized as exhibiting clinically significant depressive symptoms indicative of probable Major Depressive Disorder." The authors should be careful with the wording "Major Depressive Disorder," because PHQ-9 ≥ 10 identifies probable depression rather than a formal DSM diagnosis. The outcome variable should be consistently described as "probable MDD" or "clinically significant depressive symptoms consistent with probable MDD."

In the Data Collection Tools section, the manuscript includes both PHQ-9 and DASS Depression. This creates a potential concern about criterion contamination because the outcome classification is based on depressive symptoms and one of the strongest predictors is another depressive symptom scale. The authors should address this explicitly by explaining why DASS Depression was included as a predictor despite conceptual overlap with the PHQ-9 outcome.

Authors revised and uploaded the document.

2. Revised

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

JARAC Assessment and Research in Applied Counseling
E-ISSN: 3041-8518