

Prediction of Depression Risk Among Adolescents Using Random Forest Based on Psychological, Family, and School Factors

Arben. Krasniqi¹, Rachel. Gupta², Sophie. Krier³

¹ Department of Psychology, University of Prishtina, Prishtina, Kosovo

² Department of Applied Psychology, University of Saskatchewan, Saskatoon, Canada

³ Department of Behavioral Sciences, University of Luxembourg, Esch-sur-Alzette, Luxembourg

* Corresponding author email address: rachel.gupta@usask.ca

Editor

Sergii Boltivets^{id}
Chief Researcher of the Department
of Scientific Support of Social
Formation of Youth. Mykhailo
Drahomanov University, Ukraine
sboltivets@ukr.net

Reviewers

Reviewer 1: Faranak Saboonchi^{id}
Assistant Professor, Department of Psychology, Payam Noor University, Tehran,
Iran. Email: faranaksaboonchi@pnu.ac.ir
Reviewer 2: Sara Nejatifar^{id}
Department of Psychology and Education of People with Special Needs, Faculty of
Educational Sciences and Psychology, University of Isfahan, Isfahan, Iran.
Email: s.nejatifar@edu.ui.ac.ir

1. Round 1

1.1. Reviewer 1

Reviewer:

In the final part of the Introduction, the sentence “Random Forest offers a methodologically suitable framework because it can accommodate nonlinear associations, complex interactions, mixed predictor types, and correlated variables” is broadly correct but insufficiently comparative. The authors should clarify why Random Forest was selected over alternative machine-learning methods such as gradient boosting, XGBoost, support vector machines, penalized logistic regression, or neural networks. Because the article is explicitly framed as a machine-learning prediction study, readers need to know whether Random Forest was prespecified for theoretical or practical reasons or selected after model comparison. If no comparison was undertaken, this should be acknowledged explicitly rather than implying that Random Forest was necessarily the optimal algorithm.

The training/test split requires clarification because the manuscript reports an 80/20 stratified division resulting in 490 training participants and 122 test participants. The authors should specify the random seed used to ensure reproducibility and explain whether the test set was completely untouched until all preprocessing, variable selection, hyperparameter tuning, and threshold selection were finalized. This is critical because repeated inspection of the test set during model development would invalidate the claimed independent test performance.

The hyperparameter-selection procedure in Table 2 is not sufficiently reproducible. The manuscript states that “hyperparameters...were optimized using randomized and grid-based searches,” but does not provide the search space, number of candidate configurations, optimization metric, or whether randomized search preceded grid refinement. The authors should report the ranges evaluated for $n_{\text{estimators}}$, maximum depth, minimum samples split, minimum samples leaf, maximum features, and class weighting. In addition, nested cross-validation should be considered if model selection and performance estimation rely on the same folds, as non-nested tuning can yield optimistic estimates.

The reporting of classification performance in Table 3 should be expanded to include uncertainty estimates. Accuracy = .877, sensitivity = .850, specificity = .890, precision = .791, F1 = .829, and ROC-AUC = .923 are informative point estimates, but confidence intervals are essential given the relatively small independent test sample of 122 adolescents, particularly because only 40 participants were positive cases. Bootstrap 95% confidence intervals should be provided for ROC-AUC and ideally for sensitivity, specificity, accuracy, precision, and F1. Without interval estimates, the apparent precision of the model performance may be overstated.

Authors uploaded the revised manuscript.

1.2. Reviewer 2

Reviewer:

The Methods paragraph reporting that “a total of 658 adolescents were initially approached, of whom 627 returned questionnaires” requires a more transparent participant-flow description. The authors should specify the number of schools, school districts, classrooms, and students approached at each recruitment stage, the response rate at the school and individual levels, and the exact reasons for exclusion of the 15 participants removed from analysis. A participant flow diagram would also be useful. Without such information, it is difficult to evaluate possible selection bias or the representativeness of the final sample of 612 adolescents.

In the Study Design and Participants section, the statement that participants were recruited from “Ontario, British Columbia, and Alberta using a multistage cluster sampling procedure” requires considerably more methodological detail. Please report how provinces, school districts, schools, and classrooms were selected, whether sampling probabilities were equal, whether school type or urban/rural status was considered, and whether clustering was accounted for analytically. Students nested within schools are not statistically independent, and school-level clustering may affect model performance estimates. The authors should justify why clustering was ignored or incorporate group-aware validation, such as school-level grouped cross-validation, to prevent adolescents from the same school appearing in both training and validation folds.

The Methods state that adolescents aged 13–18 years were included, whereas the reported grade distribution ranges from Grade 8 to Grade 12. The authors should clarify the correspondence between age and grade and indicate whether any participants were enrolled in atypical grades for their age. More importantly, age-related heterogeneity should be considered analytically because early and late adolescence differ substantially in developmental characteristics, social relationships, autonomy, and depression prevalence. At minimum, the authors should report whether age was included as a predictor and consider age-stratified performance analyses.

The outcome definition requires further justification. The Methods state that “a PHQ-A score of 10 or higher [was] operationalized as elevated depression risk.” Please provide a specific validation reference supporting this threshold in adolescent populations and clarify whether the PHQ-A was being used as a screening instrument rather than a diagnostic measure. The manuscript should consistently use terminology such as “elevated depression risk” or “elevated depressive symptoms” rather than implying that the model predicts clinically diagnosed depression. This distinction is especially important in the title, discussion, and implications sections.

The Data Collection Tools section reports several established measures but does not provide reliability estimates from the present sample. Internal consistency should be reported for each multi-item instrument and subscale, preferably using McDonald’s omega in addition to or instead of Cronbach’s alpha where appropriate. For example, the manuscript should provide reliability estimates for PHQ-A, Rosenberg Self-Esteem Scale, cognitive reappraisal, expressive suppression, Family

APGAR, family support, friend support, and school connectedness. Psychometric evidence from previous studies does not demonstrate adequate reliability in this specific Canadian adolescent sample.

The measurement of school connectedness is insufficiently specified. The text refers generically to “a school-connectedness measure” and describes typical items but does not clearly identify the instrument, original authors, number of items, response options, scoring range, or validation evidence. This is not reproducible. The authors should provide the exact name and citation of the scale used, including whether it was the CDC School Connectedness Scale, Psychological Sense of School Membership Scale, or another instrument. Similar precision is required for the peer-victimization variable, for which the source measure and scoring procedure are currently unclear.

The Data Analysis section states that missing values were imputed using “the median for continuous variables and the most frequent category for categorical variables,” but the manuscript does not report the extent or pattern of missingness. Please report the percentage of missing data for each variable, the number of participants affected, and whether missingness was examined for systematic patterns. It should also be made explicit that imputation parameters were estimated separately within each training fold during cross-validation rather than globally before cross-validation; otherwise, information leakage could inflate internal validation performance.

Authors uploaded the revised manuscript.

2. Revised

Editor’s decision after revisions: Accepted.

Editor in Chief’s decision: Accepted.