

# Predicting Adolescent Depression From Digital Behavior, Sleep, and Social Connectedness Using CatBoost With SHAP Explainability

Zuzana. Kováčová<sup>1</sup>, Kevin. Patel<sup>2\*</sup>, Joseph. Mkwawa<sup>3</sup>

<sup>1</sup> Department of Experimental Psychology, Comenius University, Bratislava, Slovakia

<sup>2</sup> Department of Quantitative Psychology, York University, Toronto, Canada

<sup>3</sup> Department of Health Psychology, University of Dar es Salaam, Dar es Salaam, Tanzania

\* Corresponding author email address: kpatel@yorku.ca

## E d i t o r

Mehdi Rostami<sup>id</sup>  
Department of Psychology and  
Counseling, KMAN Research  
Institute, Richmond Hill, Ontario,  
Canada.  
Email: dr.mrostami@kmanresce.ca

## R e v i e w e r s

**Reviewer 1:** Sara Nejatifar <sup>id</sup>  
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  
**Reviewer 2:** Seyed Hadi Seyed Alitabar <sup>id</sup>  
Department of Psychology and Counseling, KMAN Research Institute, Richmond  
Hill, Ontario, Canada Email: hadialitabar@kmanresce.ca

## 1. Round 1

### 1.1. Reviewer 1

Reviewer:

The paragraph beginning “The distinction between digital exposure and digital behavior is especially important when examining depression because smartphones and wearable devices can simultaneously function as potential risk contexts and as measurement instruments” introduces digital phenotyping in considerable detail. However, the actual study relies primarily on self-reported digital-use indicators rather than passively sensed smartphone metadata. This distinction should be made explicit to prevent readers from assuming that the present dataset constitutes full digital phenotyping in the technical sense. The manuscript should state that the study draws conceptually from the digital-phenotyping literature but operationalizes digital behavior using structured self-report measures rather than continuous passive sensing.

In the Methods section, the paragraph beginning “This study employed a cross-sectional, predictive analytical design...” states that 1,284 adolescents were initially recruited and 1,238 were retained. The exclusion of 46 participants should be described with exact frequencies for each exclusion reason rather than only listing “substantial missing data, incomplete depression assessments, implausible response patterns, or failure to complete the principal sections of the survey.” A transparent attrition accounting would improve reproducibility and allow readers to assess potential selection bias. If possible,

provide the number excluded for missing PHQ-9A data, missing predictor data, implausible records, and incomplete surveys separately.

The Study Design and Participants paragraph reports recruitment across Ontario, Quebec, British Columbia, Alberta, and Manitoba, yet the manuscript does not explain how sites or recruitment channels were distributed across these provinces. The phrase “multistage convenience and school-based sampling strategy” requires additional detail. Please specify how schools or youth networks were approached, whether recruitment targets were set by province, whether public and private schools were proportionally represented, and whether recruitment occurred online, in classrooms, or through institutional mailing lists. Without these details, it is difficult to judge the external validity of the sample or whether provincial composition reflects actual population distributions.

The statement “Where participants had access to device-based screen-time records on their smartphones, they were encouraged to consult the previous seven-day screen-time summary” introduces potential measurement heterogeneity. Some participants apparently used device-recorded values, whereas others relied on recall. This difference may produce differential measurement error and should be handled explicitly. The manuscript should report what proportion of participants consulted device-based records and whether sensitivity analyses were performed comparing respondents with device-assisted versus recall-only estimates. Otherwise, screen-time measures cannot be assumed to have equivalent measurement precision across the sample.

In the Data Analysis section, the manuscript states that “limited missing values among predictors were handled using procedures compatible with the CatBoost algorithm and, where necessary, appropriate imputation methods determined according to variable type.” This description is too vague for reproducible machine-learning research. The authors should report the percentage of missingness for each predictor, the exact imputation procedure used for continuous and categorical variables, and whether imputation was conducted separately within training folds to avoid information leakage. If CatBoost’s native missing-value handling was used for some variables, this should be specified rather than combining it ambiguously with imputation.

Authors uploaded the revised manuscript.

## 1.2. Reviewer 2

Reviewer:

In the Study Design and Participants section, the sentence “Eligibility criteria included being between 13 and 18 years of age, currently residing in Canada, being able to read and understand English or French sufficiently to complete the questionnaires...” raises an important methodological issue that is not subsequently addressed. The manuscript should clarify whether validated English and French versions of all psychometric instruments were used and whether measurement equivalence across language versions was assessed or assumed. Because the study includes adolescents from Quebec and other bilingual contexts, language-related measurement differences could influence depression, sleep, and connectedness scores and should therefore be documented.

The Data Collection Tools section identifies the PHQ-9A as the primary outcome and states that “the total PHQ-9A score was used as the principal continuous outcome in the predictive analysis.” This decision is defensible, but the manuscript should justify why regression on continuous symptom severity was selected instead of binary or ordinal classification of clinically meaningful depressive symptom categories. It would also be useful to report the internal consistency of the PHQ-9A in the current sample, preferably Cronbach’s alpha and/or McDonald’s omega, rather than relying solely on the established validity of the instrument.

The description of the digital-use questionnaire is one of the weakest reproducibility points in the Methods. The paragraph beginning “Digital behavior was assessed through a structured digital-use questionnaire developed to capture both the amount and the pattern of adolescents’ engagement...” indicates that a study-specific instrument was used, but no information is provided regarding item development, pilot testing, content validity, test-retest reliability, or response formats. Because several of the most important predictors originated from this questionnaire, the authors should either provide a formal validation

description or acknowledge that these variables were derived from a non-standardized instrument. At minimum, the exact item wording, response scales, recall period, and scoring procedures should be provided in an appendix or supplementary file.

Authors uploaded the revised manuscript.

## **2. Revised**

Editor's decision after revisions: Accepted.

Editor in Chief's decision: Accepted.