

Classification of Adolescents With High Versus Low Emotion-Regulation Difficulties Using k-Nearest Neighbors and Psychological Indicators

Andrej. Kováčik¹, Emily. Chen^{2*}

¹ Department of Cognitive Psychology, Comenius University, Bratislava, Slovakia

² Department of Educational Psychology, University of Victoria, Victoria, Canada

* Corresponding author email address: echen@uvic.ca

E d i t o r	R e v i e w e r s
John S. Carlson ^{id} Distinguished Professor of the Department of Educational Psychology, Michigan State University, East Lansing, MI, United States carloj@msu.edu	Reviewer 1: Parvaneh Mohammadkhan ^{id} Professor, Department of Clinical Psychology, University of Rehabilitation Sciences and Social Health, Tehran, Iran. Email: Pa.mohammadkhani@uswr.ac.ir Reviewer 2: Seyed Ali Darban ^{id} Assistant Professor, Department of Psychology and Counseling, South Tehran Branch, Islamic Azad University, Tehran, Iran. Email: Ali.darbani@iaui.ac.ir

1. Round 1

1.1. Reviewer 1

Reviewer:

The paragraph describing the Rosenberg Self-Esteem Scale, PHQ-9, GAD-7, PSS-10, and Brief Resilience Scale should clarify whether the exact original English-language versions were used and whether any adaptations were made for adolescent comprehension. Because the participants were 13–18 years old, the authors should state whether all instruments have been specifically validated in adolescent populations comparable to the present sample. For the PHQ-9 in particular, it should be clarified whether the ninth item concerning self-harm or suicidal ideation triggered any safeguarding protocol, referral procedure, or clinical follow-up, because administration of this item to minors creates an ethical obligation that should be explicitly addressed.

The Study Design and Participants section notes that data were collected in “supervised school settings through a secure electronic survey platform,” but important procedural details are absent. The authors should indicate whether teachers were present, whether research staff supervised completion, whether students could complete questionnaires privately, the approximate duration of the assessment battery, the order of instruments, and whether order was fixed or randomized. These

conditions can affect self-disclosure, particularly for depression, anxiety, self-esteem, and emotional difficulties, and should therefore be described to permit assessment of possible measurement-context effects.

The cross-validation results in Table 2 are presented to three decimal places but without fold-to-fold variability, despite the Methods section stating that “cross-validation performance estimates were summarized using their mean and standard deviation across folds.” This inconsistency should be corrected. The table should report mean $\pm$ SD, or 95% confidence intervals, for the primary cross-validation metrics. The difference between the best model at $k = 9$ and nearby configurations such as $k = 7$ or $k = 11$ is very small; without variability estimates, it is impossible to determine whether the apparent superiority of $k = 9$ reflects a stable advantage or ordinary sampling fluctuation across folds.

The manuscript identifies the model at $k = 9$ as optimal because its F1 score was .858 compared with .854 at $k = 11$ and .853 at $k = 7$. Given these very small differences, the authors should consider applying a parsimonious selection principle, such as choosing the simplest model whose performance lies within one standard error of the best-performing model. The manuscript should also state whether hyperparameter selection was based strictly on a predefined primary metric. Using several metrics simultaneously after inspecting results may introduce implicit researcher degrees of freedom, whereas a prespecified selection criterion would improve reproducibility.

Authors uploaded the revised manuscript.

1.2. Reviewer 2

Reviewer:

The Data Analysis paragraph states that “missing predictor values, when present at a limited level, were imputed using the median calculated from the training data,” yet the manuscript does not report the extent or pattern of missingness. The authors should provide the percentage of missing observations for each predictor, the number of participants affected, and the rationale for median imputation. It should also be stated whether missingness was examined for systematic associations with group membership or demographics. If missingness exceeded trivial levels, multiple imputation or a sensitivity analysis should be considered because single-value median imputation may attenuate variance and alter local distance structures in KNN.

The preprocessing description appropriately states that “standardization parameters were estimated exclusively from the training partition and subsequently applied to the validation and test data,” but the analysis should make clearer whether imputation, standardization, and hyperparameter tuning were embedded within every cross-validation fold through a single analytical pipeline. The later statement that the “workflow was implemented as a preprocessing-and-classification pipeline” suggests this was done correctly; however, this crucial anti-leakage procedure should be stated once clearly and explicitly. Given that KNN is highly sensitive to scaling, any leakage during preprocessing could materially inflate performance estimates.

The manuscript reports that the dataset was divided into “an 80% training set and a 20% independent test set,” corresponding to 547 and 137 adolescents. Although this is acceptable, performance estimates based on only 137 test observations may be imprecise. Table 3 should therefore include 95% confidence intervals for accuracy, sensitivity, specificity, precision, F1 score, balanced accuracy, and particularly ROC-AUC. Bootstrap confidence intervals would be appropriate. Reporting point estimates alone, such as ROC-AUC = .923, risks conveying greater precision than is justified by the test-sample size.

Table 1 reports independent-samples t tests for self-esteem, depression, anxiety, perceived stress, and resilience, but the assumptions underlying these comparisons are not reported. The manuscript should state whether distributions, influential observations, and variance homogeneity were examined. More importantly, because the groups were constructed directly from the DERS median rather than independently sampled clinical categories, inferential p values for psychological differences should be interpreted cautiously. Given the large effect sizes, confidence intervals for Cohen’s d would be more informative than emphasizing statistical significance, especially because all comparisons yielded $p < .001$.

The demographic paragraph reports age, sex, grade distribution, and DERS group membership, but it does not evaluate whether demographic variables differed between the high- and low-difficulty groups. The authors should test whether age, sex/gender, and grade were associated with the classification outcome, particularly because prior literature cited in the manuscript suggests age and gender differences in adolescent emotional and behavioral functioning. If meaningful

demographic imbalance exists, the authors should consider whether these variables should be incorporated into the predictive model or examined in stratified analyses to assess whether performance is comparable across demographic subgroups.

In Table 2, the manuscript states that candidate values of k from 3 to 25 were examined and that “the optimal configuration was obtained with $k = 9$ and distance-weighted voting.” The hyperparameter search appears relatively narrow. The authors should clarify why Euclidean distance was fixed rather than tuned against alternatives such as Manhattan or Minkowski distance and whether additional weighting specifications were considered. Because distance metrics fundamentally determine neighborhood composition in KNN, optimizing only k and voting strategy may leave meaningful model performance unexplored. A more complete hyperparameter grid or explicit methodological rationale for fixing Euclidean distance is recommended.

Authors uploaded the revised manuscript.

2. Revised

Editor’s decision after revisions: Accepted.

Editor in Chief’s decision: Accepted.