

Article history:
Received 25 April 2026
Revised 11 June 2026
Accepted 18 June 2026
Published online 10 September 2026

Predicting Psychological Distress Among Adolescents Using a Multilayer Perceptron Neural Network Based on Academic, Family, and Social Factors

Ji-Eun. Han¹, Hsin-Yi. Chang^{2,4}, Ayesha. Siddiqui³

¹ Department of Counseling Psychology, Yonsei University, Seoul, South Korea

² Department of Cognitive Psychology, National Cheng Kung University, Tainan, Taiwan

³ Department of Applied Psychology, Quaid-i-Azam University, Islamabad, Pakistan

* Corresponding author email address: hsinyi.chang@ncku.edu.tw

E d i t o r	R e v i e w e r s
John S. Carlson ¹  Distinguished Professor of the Department of Educational Psychology, Michigan State University, East Lansing, MI, United carlosoj@msu.edu	Reviewer 1: Parvaneh Mohammadkhai ¹  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 ¹  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 model-development description requires stronger justification for the chosen architecture. The manuscript reports “two hidden layers, with 16 neurons in the first hidden layer and 8 neurons in the second hidden layer,” an Adam optimizer, learning rate of .001, batch size of 32, and 500 maximum epochs, but it is unclear whether these were prespecified or selected through hyperparameter optimization. Please describe the tuning procedure, search space, selection criterion, random seed, and whether tuning was confined to the training and validation data. If the architecture was selected after observing validation performance, this should be described explicitly. Ideally, nested cross-validation or another procedure that separates tuning from final performance estimation should be considered to reduce optimistic bias.

The sentence “the complete dataset of 612 adolescents was randomly partitioned into training, validation, and testing subsets using a 70:15:15 ratio” raises an important concern regarding the stability of the performance estimate. With only 92 participants in the final testing set, R^2 , MAE, and RMSE may vary materially depending on the random split. Please provide confidence intervals for these metrics, ideally based on bootstrap resampling of the independent test predictions. The authors

should also report whether the outcome distribution and key demographic characteristics were balanced across the training, validation, and test subsets. A single split should not be presented as definitive evidence of generalizability without uncertainty estimates.

The manuscript additionally states that “repeated model evaluation and cross-validation procedures were used during model development,” but no actual cross-validation design or results are reported. This creates an inconsistency between the Methods and Findings. Please specify the number of folds, number of repeats, whether folds were stratified or grouped, what metric guided model selection, and the mean and variability of cross-validated performance. If cross-validation was not actually used in the final analysis, the statement should be removed. If it was used, its findings should be presented so that readers can distinguish internal resampling performance from the final independent test-set performance.

Figure 1, titled “Observed and Predicted Psychological Distress Scores in the Independent Testing Dataset,” should be strengthened as a quantitative validation figure rather than treated primarily as a visual confirmation of R^2 . Please ensure that the plot is generated directly from the actual 192 test observations and not from illustrative or simulated points. The figure should include the identity line, an estimated calibration line, and preferably a 95% confidence band. The manuscript should also report the calibration intercept and slope or another quantitative calibration index. Because the current narrative mentions “regression toward the mean” at extreme K10 scores, this pattern should be formally assessed rather than inferred only from visual inspection.

Authors uploaded the revised manuscript.

1.2. Reviewer 2

Reviewer:

In Table 2, the manuscript reports Pearson correlations such as “educational stress total, $r = .68$, $p < .001$ ” and “family functioning, $r = -.53$, $p < .001$.” The authors should report 95% confidence intervals for correlation coefficients rather than relying only on p values. Additionally, because 13 correlations are tested simultaneously, please discuss control of multiplicity, such as a false-discovery-rate adjustment or a prespecified exploratory interpretation. While all reported p values are very small, adjustment remains important for methodological transparency. The authors should also clarify whether assumptions for Pearson correlation were evaluated and whether Spearman correlations yielded comparable results if distributions were non-normal or bounded.

Table 3 reports R^2 values of .76, .69, and .66 for training, validation, and testing data, respectively. These values appear strong for prediction of adolescent psychological distress from questionnaire-based variables and therefore warrant additional verification. Please report the baseline performance of a naïve comparator, such as predicting the training-set mean for all cases, and preferably compare the MLP with simpler models such as ordinary least squares regression, penalized regression, random forest, or gradient boosting. Without benchmark models, it is impossible to determine whether the neural network provides meaningful incremental predictive value or whether comparable accuracy could be obtained with a more interpretable conventional approach.

The statement beneath Table 3 that “the relatively small differences between training, validation, and testing indices indicated acceptable model generalizability rather than substantial memorization” is somewhat stronger than the evidence permits. Similar performance across internal splits indicates limited apparent overfitting within this dataset, but it does not establish external generalizability. Please revise this interpretation and distinguish clearly among internal validity, out-of-sample performance within the sampled population, and external validity in new schools, regions, or cohorts. External validation using an independent Taiwanese sample would be required before claims of broad generalizability are justified.

Table 4 uses permutation-based importance and states that “educational stress total emerged as the most influential predictor.” This analysis requires greater methodological detail and more cautious interpretation. Please state the number of permutations performed for each predictor, whether importance values were averaged over repeated permutations, and whether uncertainty intervals were calculated. Because many predictors are correlated or mathematically overlapping, standard permutation importance can underestimate or redistribute importance when one variable's information is recoverable from

another. Conditional permutation importance, grouped permutation, or SHAP-based analyses may offer more defensible interpretation. At minimum, the manuscript should acknowledge that relative importance is model-dependent and does not represent causal importance.

The relationship between Table 2 and Table 4 should be examined more analytically. For example, “despondency” has a strong correlation with distress ($r = .64$) and ranks second in permutation importance, whereas “family support” and “total perceived social support” are both included despite conceptual overlap. The Discussion currently treats these rankings as substantively distinct findings without considering predictor dependence. I recommend adding an analysis of intercorrelations among all model inputs and discussing how correlated features affect neural-network attribution. A correlation heat map or supplementary matrix would be helpful, but the key requirement is to avoid interpreting each importance score as if predictors contributed independently.

Authors uploaded the revised manuscript.

2. Revised

Editor’s decision after revisions: Accepted.

Editor in Chief’s decision: Accepted.