

Artificial Neural Network Prediction of Psychological Well-Being Among Women Using Emotional, Cognitive, and Social Features

Justas. Rimkus¹, Jessica. Moore^{2*}

¹ Department of Health Psychology, Lithuanian University of Health Sciences, Kaunas, Lithuania

² Department of Media Psychology, Toronto Metropolitan University, Toronto, Canada

* Corresponding author email address: jessica.moore@torontomu.ca

E d i t o r	R e v i e w e r s
Donatella Di Corrado  Associate Professor, Department of psychology, Kore University of Enna, Enna, Italy donatella.dicorrado@unikore.it	Reviewer 1: Mohammad Masoud Dayaria  Assistant Professor, Department of Counseling, Khomeinishahr Branch, Islamic Azad University.Komeinishar/Isfahan,Iran. dayariyan@iaukhsh.ac.ir Reviewer 2: Nadereh Saadat  Department of Couple and Family therapy, Alliant International University, California, United States of America. mdaneshpour@alliant.edu

1. Round 1

1.1. Reviewer 1

Reviewer:

In the same Study Design and Participants paragraph, the sentence “Participants were recruited from community settings, universities, workplaces, women’s organizations, and online community networks across several Canadian provinces...” is too broad to permit reproducibility. The authors should specify the recruitment period, the exact recruitment platforms or institutional channels, whether recruitment was centralized or province-specific, whether advertisements were paid or unpaid, and whether quotas or stratification procedures were used to prevent overrepresentation of particular subgroups. These details are particularly important because the final sample was disproportionately concentrated in Ontario, which may affect the generalizability of the predictive model.

The sentence “Multiple network initializations were evaluated because neural network performance can vary depending on randomly assigned starting weights, and the model demonstrating the most stable validation performance was retained” raises an important methodological concern. The authors should report how many independent initializations were performed and define “most stable validation performance” quantitatively. Selecting the single best initialization may produce an optimistically biased performance estimate. A stronger approach would report mean and standard deviation of test performance

across repeated runs or repeated train-validation-test partitions, thereby demonstrating that the results are not dependent on one favorable random initialization.

Table 3 reports training, validation, testing, and “overall sample” predictive performance. The “overall sample” row should be reconsidered because combining observations that played different roles in model development can produce a performance estimate with unclear inferential meaning. The independent test-set metrics should be treated as the principal estimate of generalization performance. If the authors retain the overall metric, they should explicitly state that it is descriptive only and should not be interpreted as independent predictive validation. The test-set RMSE, MAE, and R^2 should receive primary emphasis throughout the Results, Abstract, and Discussion.

The manuscript reports that the test sample produced “ $R^2 = .71$ ” and a correlation of “.85 between observed and predicted values.” The authors should clarify how R^2 was computed. Squaring the Pearson correlation coefficient between observed and predicted outcomes is not necessarily equivalent to the coefficient of determination calculated from prediction residuals, especially when predictions are biased or poorly calibrated. The manuscript should define R^2 mathematically or specify the software function used and should additionally consider reporting calibration intercept and slope to demonstrate whether predicted values are systematically shifted or compressed relative to observed scores.

Authors revised the manuscript and uploaded the document.

1.2. Reviewer 2

Reviewer:

The Methods section states that participants were “adult women residing in Canada,” but the operational definition of the target population is insufficiently clear. The authors should clarify whether eligibility was based on self-identified gender, sex assigned at birth, or both, and whether transgender women were eligible. This is not merely a demographic reporting issue; it is essential for defining the inferential population. The manuscript should also specify whether Canadian citizenship was required or whether permanent residents, temporary residents, international students, and other adult women living in Canada were eligible.

In the Data Collection Tools section, the paragraph beginning with “Psychological well-being, the primary outcome variable, was assessed using Ryff’s Psychological Well-Being Scale” lacks the exact version of the instrument. Ryff’s measure exists in multiple forms with substantially different item counts, including 18-, 42-, 54-, and 84-item versions. The authors should report the precise version used, number of items, response scale, possible total-score range, subscales, reverse-scored items, scoring procedure, and internal consistency obtained in the present sample. Without these details, the reported mean psychological well-being score of 71.64 cannot be meaningfully interpreted or compared with previous studies.

The paragraph describing emotional intelligence states that it “was assessed using a standardized self-report measure” but does not identify the instrument by name, developer, publication year, item count, response format, or scoring method. This is a major reproducibility problem. The manuscript should name the exact emotional intelligence instrument and provide sufficient psychometric information, including whether a total score or subscale scores were used as ANN inputs. If a total score was entered into the network, the authors should justify this choice, particularly if the selected measure was originally designed as a multidimensional construct.

A similar issue occurs in the statement “Dispositional mindfulness was assessed with a validated mindfulness questionnaire” and in the description of social connectedness. These measures are insufficiently specified. The authors should explicitly identify every questionnaire used, including creators, year, number of items, subscales, Likert response format, score range, direction of interpretation, and evidence of reliability and validity. A Methods section intended for replication cannot refer generically to “a validated mindfulness questionnaire” or “a standardized social connectedness measure.”

The Data Analysis paragraph states that continuous variables were standardized “using z-score standardization or equivalent normalization procedures.” The phrase “or equivalent normalization procedures” introduces ambiguity regarding a critical preprocessing step. The authors should specify exactly which transformation was used for each input variable. More importantly, normalization parameters should be estimated exclusively from the training dataset and then applied unchanged

to validation and test datasets. If standardization was performed before data partitioning, this would constitute data leakage and could lead to optimistic estimates of out-of-sample performance.

The manuscript reports that the data were randomly divided into 300 training cases, 64 validation cases, and 64 testing cases. The paragraph describing this split should explain the randomization procedure and provide the random seed used, particularly because the relatively modest test sample can produce performance estimates that are sensitive to a particular split. The authors should also indicate whether the partitioning was stratified according to the distribution of psychological well-being or important demographic characteristics. Reproducibility of the reported $R^2 = .71$ in the test sample depends heavily on documentation of this procedure.

In Table 2, the final network architecture is reported as two hidden layers containing 12 and 6 neurons, but the manuscript only states that “several candidate network architectures were examined.” The model-selection process requires substantially greater detail. The authors should report the full hyperparameter search space, including candidate numbers of layers and neurons, activation functions, optimizers, learning rates, batch sizes, regularization procedures, and stopping criteria. They should also explain whether these configurations were selected manually, through grid search, random search, Bayesian optimization, or another method. Otherwise, the reported architecture appears post hoc and cannot be independently reproduced.

Authors revised the manuscript and uploaded the document.

2. Revised

Editor's decision: Accepted.

Editor in Chief's decision: Accepted.