

Detecting Adolescent Emotional Distress from Daily Smartphone Behavioral Patterns Using Long Short-Term Memory Networks

Martín. Echeverría¹, Valentina. Rincón^{2*}

¹ Department of Psychology, University of the Republic, Montevideo, Uruguay

² Department of Organizational Psychology, Universidad de los Andes, Bogotá, Colombia

* Corresponding author email address: v.rincon@uniandes.edu.co

E d i t o r	R e v i e w e r s
John S. Carlson  Distinguished Professor of the Department of Educational Psychology, Michigan State University, East Lansing, MI, United carlsoj@msu.edu	Reviewer 1: Parvaneh Mohammadkhani  Professor, Department of Clinical Psychology, University of Rehabilitation Sciences and Social Health, Tehran, Iran. Email: Pa.mohammadkhani@uswr.ac.ir Reviewer 2: Mahdi Khanjani  Associate Professor, Department of Psychology, Allameh Tabataba'i University, Tehran, Iran. Email: khanjani_m@atu.ac.ir

1. Round 1

1.1. Reviewer 1

Reviewer:

In the Study Design and Participants section, the sentence “Participants were excluded if they... provided fewer than 70% of the minimum required daily behavioral records” requires additional methodological justification. Please explain why 70% was selected as the data-completeness threshold and whether this criterion was determined a priori. It would also be useful to report how many participants were excluded for insufficient sensing coverage and whether excluded participants differed systematically from included participants in age, sex, school type, or baseline emotional symptoms. Differential attrition related to distress or smartphone-use behavior could introduce meaningful selection bias.

In the Data Analysis section, the sentence “Daily observations were organized into sequential windows of seven consecutive days, with the smartphone behavioral features from these days constituting the input sequence to the LSTM network and emotional-distress status on the final day of each sequence serving as the primary prediction target” requires clarification regarding contemporaneous prediction. If smartphone features from the seventh day are included when predicting distress on that same day, the model is performing detection rather than forecasting. This is compatible with the article title, but the manuscript should state this explicitly. If the authors intend to claim early warning or prediction before distress occurs, the target should instead be distress on day 8 using behavioral features from days 1–7.

Authors uploaded the revised manuscript.

1.2. Reviewer 2

Reviewer:

In the Data Collection Tools section, the paragraph beginning “Data were collected using a custom smartphone sensing application...” lists numerous smartphone-derived variables, but the technical acquisition procedures remain insufficiently reproducible. Please describe the operating-system permissions required, sampling frequency for location and motion data, how background restrictions were handled, whether data were stored locally before transmission, how often synchronization occurred, and whether battery-preservation algorithms affected sensing density. For a digital-phenotyping study, these implementation details are essential because differences in sensing architecture can substantially alter the completeness and reliability of behavioral features.

In the same paragraph, the sentence “precise geographical coordinates were transformed into mobility features before analysis and were not retained in the final analytical dataset” is appropriate from a privacy perspective, but the feature-engineering procedure should be reported more precisely. Please define how “radius of gyration,” “distinct location clusters,” “mobility entropy,” and “time spent at the most frequently visited location” were computed. The clustering algorithm, distance threshold, minimum dwell time, treatment of GPS error, and handling of missing location intervals should be specified so that the mobility features can be replicated in future studies.

In the Data Collection Tools section, the paragraph describing ecological momentary assessment states that “A daily emotional distress score was calculated as the mean of the completed items,” using sadness, anxiety/worry, irritability, stress, emotional exhaustion, and difficulty coping. Please provide evidence for the psychometric validity of this composite in the current sample. At minimum, report internal consistency or multilevel reliability and explain whether the six indicators formed a unidimensional construct. If this daily measure was created specifically for the present study rather than drawn directly from a validated EMA scale, that should be stated explicitly, and its development, content validity, and pilot testing should be described.

The sentence “For the primary classification analysis, each daily observation was categorized as high emotional distress when the daily mean score was 2.00 or higher” requires stronger justification. The cutoff appears analytically consequential because it determines class prevalence and all subsequent model-performance metrics. Please explain whether the value of 2.00 corresponds to a clinically or empirically established threshold, was based on the response-scale midpoint, or was selected pragmatically. The sensitivity analyses using alternative thresholds are useful, but the primary threshold must nevertheless be conceptually justified before model training rather than appearing data-driven.

In the Data Analysis section, the paragraph stating that “short periods of missing sensor data were imputed within the training data using the median value of the corresponding participant” raises a potential temporal-leakage concern. Please clarify whether participant-specific medians were calculated using only observations temporally available within the training partition and whether future observations within a participant’s training period could contribute to imputing earlier missing values. For sequential prediction, any use of future information to reconstruct past features may produce optimistic estimates. A causal or forward-only imputation strategy may be preferable if the intended application involves prospective detection.

Authors uploaded the revised manuscript.

2. Revised

Editor’s decision after revisions: Accepted.

Editor in Chief’s decision: Accepted.