

Forecasting Daily Stress in Adolescents from Mood Variability, Sleep Quality, Social Interaction, and Smartphone Use Using Long Short-Term Memory Networks

Katrín. Thórisdóttir¹, Thomas. Campbell^{2*}, Elena. Trajkovska³

¹ Department of Clinical Psychology, University of Iceland, Reykjavík, Iceland

² Department of Experimental Psychology, University of New Brunswick, Fredericton, Canada

³ Department of Psychology, Ss. Cyril and Methodius University, Skopje, North Macedonia

* Corresponding author email address: tcampbel@unb.ca

E d i t o r	R e v i e w e r s
Sergii Boltivets Chief Researcher of the Department of Scientific Support of Social Formation of Youth. Mykhailo Drahomanov University, Ukraine sboltivets@ukr.net	Reviewer 1: Sara Nejatifar 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 Ali Darbani Assistant Professor, Department of Psychology and Counseling, South Tehran Branch, Islamic Azad University, Tehran, Iran. Email: Ali.darbani@iau.ac.ir

1. Round 1

1.1. Reviewer 1

Reviewer:

In the paragraph beginning “Long short-term memory networks are a form of recurrent neural network designed to process sequential data,” the technical justification for selecting an LSTM is incomplete. The authors should explain why an LSTM was preferred over alternative sequence models such as gated recurrent units, temporal convolutional networks, autoregressive mixed-effects models, or transformer-based architectures. This justification should address the sample size, sequence length, number of predictors, expected temporal dependencies, and risk of model overparameterization. Without this information, the choice of architecture may appear method-driven rather than based on the specific statistical properties of the data.

In the Study Design and Participants section, the manuscript reports that “the study included exactly 312 adolescents between 13 and 18 years of age who were recruited from public and private secondary schools, community youth centers, and adolescent health services.” The authors should provide the sampling frame and recruitment flow in greater detail, including the number of institutions approached, the number that agreed to participate, the number of adolescents invited, the number screened,

reasons for ineligibility, and reasons for withdrawal. The current description does not allow readers to evaluate selection bias or determine whether recruitment from adolescent health services produced overrepresentation of participants with psychological difficulties.

In the Data Analysis section, the manuscript states that “short gaps in time-varying continuous predictors were imputed using participant-specific interpolation,” followed by rolling medians and training-set medians. This approach may be problematic for forecasting because interpolation using values occurring after a missing observation can introduce future information into a predictor sequence. The authors should clarify whether interpolation was conducted only within historical predictor windows and whether any value from the forecast target day or a later day was used. A strictly causal imputation strategy should be applied, such as forward filling, historical rolling estimates, or model-based imputation restricted to information available before the forecast time.

In the paragraph describing participant-level data partitioning, 218 participants were allocated to training, 47 to validation, and 47 to testing. The authors should explain how the stratified allocation was implemented when simultaneously balancing age, gender, province, socioeconomic status, and baseline stress. With several stratification variables and relatively small validation and test samples, exact balance may be difficult to achieve. The manuscript should report the distribution of these variables across partitions, preferably in supplementary material, and provide the random seed used for allocation so that the analysis can be reproduced.

In the description of the sliding-window procedure, the manuscript reports that seven consecutive days were used to predict next-day stress. However, the inclusion of “previous-day stress” as the strongest predictor is introduced later in the Findings and is not sufficiently emphasized in the Methods. The predictor list should explicitly state that lagged stress scores were included in each sequence and clarify whether all seven prior stress scores or only the immediately preceding score were entered. This information is essential because past values of the outcome can account for a large proportion of predictive performance and materially affect the interpretation of the incremental value of mood, sleep, social, and smartphone variables.

Authors uploaded the revised manuscript.

1.2. Reviewer 2

Reviewer:

In the same section, the manuscript states that adolescents were recruited from Ontario, British Columbia, Alberta, and Quebec, but it does not explain how language was handled in Quebec. The eligibility criteria required sufficient English-language proficiency, which may have excluded francophone adolescents and reduced the representativeness of the Quebec subsample. The authors should report whether French-language study materials were available, how language proficiency was assessed, and whether recruitment differed by school language. If only English-language participation was possible, this limitation should be acknowledged explicitly because it affects the interpretation of the sample as broadly Canadian.

In the Data Collection Tools section, the manuscript states that “daily perceived stress was measured every evening using a brief four-item daily perceived stress measure.” The origin and psychometric status of this measure are not sufficiently documented. The authors should identify whether the items were adapted from an established scale, provide the exact item wording, explain the adaptation process for daily use, and report evidence of content validity. Because the findings depend on repeated measurement of a brief outcome, the manuscript should also report multilevel reliability coefficients, including within-person and between-person reliability, rather than merely stating that these were evaluated.

In the paragraph describing mood assessment, participants rated six affective states four times per day, and daily mood variability was calculated using within-day standard deviations and root mean square successive differences. The authors should specify exactly which mood-variability index entered the final LSTM model and whether multiple highly related indices were included simultaneously. Including overall affective instability, positive-affect variability, negative-affect variability, daily positive affect, and daily negative affect could introduce substantial redundancy. The manuscript should report correlations or variance inflation diagnostics among engineered mood features and explain how feature selection or regularization was used to prevent overrepresentation of the mood domain.

In the sleep-assessment paragraph, the authors report the use of a smartphone-adapted daily sleep diary and the Pittsburgh Sleep Quality Index at baseline. The manuscript should identify the source or standard on which the daily diary was based and explain whether sleep variables were derived according to accepted consensus definitions. For example, sleep efficiency requires precise definitions of time in bed and total sleep time, and the reported calculation should account for sleep-onset latency and wake after sleep onset. The authors should also clarify whether baseline Pittsburgh Sleep Quality Index scores were included in the model and, if so, whether they improved prediction beyond daily sleep indicators.

In the social-interaction assessment paragraph, the manuscript combines self-reported face-to-face interaction, perceived social support, loneliness, call metadata, messaging frequency, and communication-related application sessions. These indicators represent distinct constructs and may have different directions of association with stress. The authors should provide a clearer taxonomy separating interaction quantity, interaction modality, perceived interaction quality, support, and social disconnection. They should also specify whether communication metadata were equally available on iOS and Android devices, because operating-system restrictions could create systematic measurement differences that may be learned by the model rather than reflecting genuine psychosocial processes.

In the smartphone-use paragraph, the manuscript states that application activity was classified into “social networking, communication, entertainment, gaming, education, productivity, and general browsing.” The classification procedure requires substantially more detail. The authors should explain whether applications were categorized manually, through operating-system taxonomies, or using an external database; how multifunctional applications were classified; whether categories were verified across devices; and how uncategorized applications were handled. Reproducibility is limited unless the feature-extraction and classification rules are fully described, preferably in supplementary materials.

Authors uploaded the revised manuscript.

2. Revised

Editor’s decision after revisions: Accepted.

Editor in Chief’s decision: Accepted.