Qeios

Peer Review

Review of: "Evaluating Neurostellar's Wearable Device Along with Its Robust Multimodal Approach for Real-Time Mental State Assessment: A Validation Study"

Birju Patel[1]

1. Independent researcher

This study presents a comprehensive validation of a consumer-grade EEG/PPG wearable headband for mental state assessment in naturalistic settings. The authors employ a thoughtful two-phase design comparing the device against laboratory-standard equipment and testing whether multimodal features can classify cognitive/affective states and predict subjective ratings. The findings demonstrate moderate-to-strong correlations with lab measurements across numerous features and successful differentiation of mental states using robust machine learning approaches.

1. Well-Motivated Multimodal Framework

The authors present a compelling rationale for combining EEG and HRV signals to capture mental states more robustly than unimodal approaches. The integration of short-window individual metrics with long-window complexity measures from multi-feature cluster sequences is innovative and well-grounded in the microstate literature. This approach sensibly addresses the inherent noise challenges of real-world wearable data.

2. Rigorous Statistical Methodology

The use of permutation testing (1000 permutations) for both classification and regression analyses provides robust statistical inference that does not rely on distributional assumptions. The authors

doi.org/10.32388/M7K40N

appropriately employed stratified cross-validation to maintain subject integrity, and the use of robust statistical measures from the WRS2 library demonstrates attention to handling outliers appropriately.

3. Ecologically Valid Design

The study admirably prioritizes ecological validity by conducting recordings in office and home environments rather than controlled laboratory settings. Participants sat in comfortable positions of their choosing, and the protocol included self-selected music and videos, enhancing the relevance of findings to real-world consumer use cases.

4. Comprehensive Feature Engineering

The extraction of 18 EEG features and 12 HRV features, spanning frequency-domain, time-domain, and nonlinear measures, provides a thorough characterization of the physiological signals. The novel cluster sequence complexity measures add an additional layer of analysis that may prove valuable for future wearable research.

5. Transparent Reporting

The authors provide detailed methodological descriptions, including preprocessing parameters, feature extraction procedures, and machine learning configurations. The acknowledgment of conflicts of interest and limitations demonstrates scientific integrity.

Minor Suggestions for Improvement

1. Clarification of Correlation Criteria

The validation criterion of r>0.5 correlations in ">50% of recordings" could benefit from additional context. The authors might briefly discuss how this threshold compares to validation standards in similar consumer wearable studies, which would help readers appreciate the device's performance relative to the field.

2. Additional Detail on Artifact Handling

While the preprocessing pipeline is well-described, a brief mention of how many epochs or segments were excluded due to poor signal quality would strengthen the methods section. This information would help readers understand data retention rates in naturalistic recordings.

3. Expanded Discussion of Regression Effect Sizes

The R^2 values of 0.16-0.17 for the RelaxedFocus prediction, while statistically significant, could be contextualized further. A brief discussion comparing these effect sizes to similar prediction studies in the psychophysiology literature would help readers interpret the practical significance.

4. Future Directions for the Cluster Sequence Approach

The novel complexity measures derived from multi-feature cluster sequences represent an interesting methodological contribution. The authors might consider briefly discussing how this approach could be validated against established measures in future work, which would strengthen the foundation for this promising technique.

5. Minor Presentation Suggestions

- Figure 2 is information-dense; consider adding a brief caption clarification for the color coding differences between EEG and HRV clusters
- Table 2 formatting could be slightly adjusted for improved readability
- Consider reporting exact p-values where feasible rather than $p<0.001$

Technical Soundness

The study demonstrates appropriate sample size calculation using G*Power, achieving the target of 30+ recordings for the correlation analyses. The machine learning pipeline employs well-established algorithms (Random Forest, Gradient Boosting) with appropriate hyperparameter tuning via grid search. The use of both classifiers provides convergent evidence for the robustness of findings.

The feature selection rationale, drawing from sleep staging algorithms and mind-wandering detection literature, is well-justified. The decision to use 6 clusters based on interpretability, consistency, and data representation criteria reflects thoughtful methodological choices.

Significance and Impact

This work makes a valuable contribution to the growing literature on consumer wearable validation for mental state assessment. The demonstration that multimodal features can significantly classify different

activity states and predict combined relaxation-focus ratings addresses an important gap in translating laboratory neuroscience methods to everyday settings.

The honest acknowledgment that individual Relax ratings were not significantly predicted, while RelaxedFocus was, provides nuanced insights into the psychophysiological signatures that wearables can and cannot capture. This kind of balanced reporting advances the field more than overclaiming would.

Overall Assessment

This is a well-designed validation study that makes meaningful contributions to consumer neurotechnology research. The multimodal approach is innovative, the statistical methods are rigorous, and the ecological validity enhances translational relevance. The limitations acknowledged by the authors (modest sample size, restricted demographic diversity) represent appropriate directions for future research rather than fundamental flaws in the current work.

The findings provide encouraging evidence that consumer-grade wearables, when paired with sophisticated feature engineering and appropriate analytical frameworks, can capture meaningful psychophysiological variations in naturalistic settings. This work lays important groundwork for accessible mental state monitoring applications.

Recommendation: Accept with minor revisions. The suggested clarifications would strengthen an already solid manuscript but are not essential for the core contributions to stand.

Declarations

Potential competing interests: No potential competing interests to declare.