Research Article

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

Vrinda Marigowda[1], Arun Sasidharan[2], Varsha Vijayan[1], Fayaz Khan Pathan[1], Aravind G[1], Karthik Raghavendran[1]

1. Neurostellar Private Limited, India; 2. Centre for Consciousness Studies, Dept of Neurophysiology, National Institute of Mental Health and Neuro Sciences, India

Consumer wearables for mental state assessment face scepticism regarding data quality. This study evaluated whether a consumer-grade EEG/PPG wearable headband (Neurostellar's wearable device) can reliably capture mental states in real-world settings. We compared the wearable device's EEG (AF7, AF8) and PPG-derived heart rate variability (HRV) signals with a laboratory-standard EEG/ECG device across 31 resting sessions (15 adults), assessing univariate correlations across 18 EEG, 12 HRV features. Participants then completed four cognitive and affective tasks while using the wearable device, providing subjective feedback. Novel multimodal features, combining short-window individual metrics with long-window complexity from multi-feature cluster sequences, were utilised in machine learning (Random Forest, Gradient Boosting) to classify tasks and predict changes in subjective relaxation and focus. Neurostellar's wearable device showed moderate-to-strong correlations with lab measurements for numerous EEG (eyes-closed: 12; eyes-open: 14) and HRV (eyes-closed: 9; eyes-open: 6) features in >50% of recordings. The advanced multimodal features significantly classified all tasks (Random Forest: z=13.08-13.11; Gradient Boosting: z=13.00-13.16, all p<0.001) and predicted combined relaxation/focus ratings (Random Forest: z=2.37; Gradient Boosting: z=1.79, all p<0.05) versus chance. These findings suggest that consumer wearables, employing robust multimodal feature engineering, can effectively capture psychophysiological changes and track mental states, such as focus and relaxation, in naturalistic environments, offering potential for real-world neuroscience applications. Limitations include a modest sample size and restricted demographic diversity, which warrant further research.

Correspondence: papers@team.qeios.com — Qeios will forward to the authors

Introduction

In today's highly stimulating digital environment, maintaining sustained attention and cognitive control has become increasingly difficult. The constant influx of information, particularly through smartphones and social media, has led to reduced focus, increased cognitive load, and growing inability to engage in deep, uninterrupted work[1][2][3]. Burnout and stress are widespread, yet many individuals lack the tools or awareness to properly address these challenges or regulate their mental states[4]. While psychometric scales can capture aspects of these mental states[5], they often fail to reflect the moment-to-moment nuances of lived experience[6][7][8][9]. Post-event interviews can provide additional detail, but tracking mental states over time remains a significant challenge[10][11]. Objective imaging techniques like EEG, ERP, and fMRI offer valuable insights but are mainly limited to laboratory settings and are not easily implemented in naturalistic, everyday environments.

To address these issues, wearable devices that quantify, track mental states, and provide objective data-driven insights can be highly effective[12][13]. They empower users to identify specific areas of difficulty and develop tailored interventions aligned with their personal goals[14][12]. This is particularly relevant for individuals aiming to enhance productivity, improve cognitive performance, and establish sustainable mental training routines[15][16]. The wearable EEG (Electroencephalography) device should ideally capture brain activity with signal quality comparable to gold-standard, lab-standard EEG systems[17]. Combining EEG with PPG (photoplethysmography) enables a multimodal approach, capturing physiological changes that reflect mental states such as focus, cognitive load, and relaxation[18][19][20].

Most available wearables assess these mental states by measuring standard EEG frequency bands (alpha, beta, theta, gamma, and delta)[21][22][17] and often use a one-size-fits-all model[23]. However, due to substantial inter-individual variability, these generalised patterns may not be applicable across all users. Therefore, there is a need for a tool or algorithm that can assess these states while accounting for subject-specific variations[24]. The applicability of consumer-grade EEG/PPG wearables for mental state assessment is often questioned due to concerns regarding data quality, including noise from dry electrodes, user-induced artifacts, limited spatial resolution, and basic mobile filtering techniques. Moreover, most validation studies are frequently conducted in controlled settings, potentially overlooking the complexities of real-world data, thereby contributing to reservations concerning proprietary mental state metrics. These metrics are typically derived from compromised data through the application of undisclosed algorithms. Nonetheless, the potential of real-world wearable data remains significant. Rigorous multimodal frameworks that address inter-individual variability and mitigate noise can enable more accurate representation of complex mental states by integrating brain and physiological signals. Given the variability in brain and physiological patterns across individuals[25][26], we also need multivariate measures that capture relations unique to an individual or state, adaptively targeting core cognitive processes such as focus and relaxation for a more holistic and objective assessment of mental states[27][19].

To improve the robustness of multimodal features, this study uses complexity measures from multi-feature cluster sequences, inspired by microstate-based analysis[28][29]. By defining clusters based on multi-feature relationships, the overall pattern can remain robust even if one or two features within a short segment are briefly affected. The cluster assignment thus relies on the Gestalt of the feature set rather than the precise value of every single feature, combining the essence of both EEG and HRV features in a lower dimension. Moreover, the complexity of these cluster sequences (e.g., Lempel-Ziv complexity, entropy of the sequence) over a larger time window (like 45s) is more dependent on the patterns and transitions between states (clusters) rather than the absolute fidelity of each state's feature vector. A brief artefact might introduce a "noisy" cluster label, but it might not drastically alter the overall predictability or randomness of the sequence if the surrounding labels are consistent with the underlying mental state. While abstracting to cluster sequences can reduce noise, it might also discard some fine-grained information present in the raw features. So, combining both short-window individual multi-model features and long-window complexity of multi-feature cluster sequences could be a more comprehensive and robust approach for the noisy real-world data from wearable devices.

In this study, we used a wearable device (Neurostellar, India), which captures two frontal EEG and PPG signals. This will help us design a multimodal framework, integrating both brain and physiological signals (such as heart rate variability and respiration). This study aims to answer two research questions: (1) 'Are the EEG and PPG signals recorded using the wearable device headband valid and reliable compared to a lab-standard device even in less-than-ideal settings?', and (2) 'Can wearable device-derived multimodal neurophysiological markers (derived from the EEG and PPG signals) be useful to evaluate mental states which varies in terms of focus and relaxation?'

Firstly, we expect the signal quality of this wearable device (no skin preparation and dry electrodes) to be noisier than the lab-standard system (standard skin preparation and wet research-grade electrodes), but to show at least moderate-level concordance. Secondly, we expect the multimodal algorithm to be robust enough to capture significant multivariate changes associated with mental state transitions. Participants will undergo baseline assessments, followed by cognitively engaging tasks such as chess and an adaptive Sternberg task. These will be followed by relaxation sessions involving music listening and video watching. EEG and PPG data collected during these sessions will be used to validate the algorithm's performance and reliability. Ultimately, the device can serve as a companion in the user's cognitive training journey, facilitating targeted improvement, tracking measurable progress, and supporting long-term self-regulation across a range of real-world, non-clinical applications.

Methods

Sample size calculation

For the between-device signal feature correlation, considering a moderate correlation of r=0.5 with an error rate of 5% and statistical power of 80% for a two-tailed assessment, the required sample size was 29. For evaluating the usefulness of the wearable device-derived measures using F1 scores of ML algorithms, considering a moderate effect size of g=0.3 for above chance change in a binomial one-sample test with an error rate of 5% and statistical power of 80% for a one-tailed assessment, the required sample size was 18. So we decided to acquire at least 30 recordings from 10 subjects with 3-day recordings per subject. This would be similar to the recording profile that Neurostellar's wearable device will have to collect in the real world, i.e., a mixture of multi-subject and multi-day data. The sample size calculation was done using G*Power software (version 3.1).

This study was performed in accordance with the Declaration of Helsinki. This human study was approved by the Company's Human Ethics Committee (NIEC001/06/2025). All adult participants provided written informed consent to participate in this study.

Participants comprised both males and females aged between 18 and 50 years, with normal, or normal to corrected vision and hearing, with no history of neurological or psychiatric disorders, were recruited based on predefined inclusion and exclusion criteria using a snowball sampling method, following the completion of written informed consent. Subjects were screened using the General Health Questionnaire-28 (GHQ-28) to assess their general well-being, and handedness was evaluated using the Edinburgh Handedness Inventory (EHI). During recruitment, all 10 participants did not complete 3 day recordings and we had to recruit additional subjects for single day sessions, achieving 31 recordings from 15 participants. The demographics of the participants are mentioned in Table 1 and the details regarding number of recordings are mentioned in Table 2.

Sl No	Variables	Details
1	Age (Mean ± SD)	33.2 ± 8.74 years
2	Gender	Male = 8; Female = 7
3	Education (Years)	18.84 ± 3.9
4	GHQ-28 Score (Mean ± SD)	1.92 ± 2.3
5	Handedness	All right-handed

Table 1. Demographic Details of Participants (N = 15)

Study Design

The study comprised two distinct phases for each participant. Initially, subjects underwent simultaneous recordings using both the Neurostellar's wearable device and a laboratory-standard portable EEG/ECG device during an extended resting-state period of ten minutes (Both eyes open and eyes closed conditions, five minutes each). Subsequently, within a short interval, participants were monitored solely with the Neurostellar's wearable device across four distinct conditions, which were designed to elicit focus, relaxation, and mental effort. These were three-minute conditions with continuous eye opening: a modified Sternberg's working memory task, listening to self-selected soothing music, watching self-selected captivating videos, and playing a computer chess match (if proficient). Each condition was preceded and followed by a two-minute resting-state recording with alternating eye closure and opening (see Figure 1 and Table 2). Each participant completed recordings under all aforementioned conditions within a single day, with this protocol being replicated over two additional days. Consequently, up to three recordings for each specific condition were obtained from each participant. All the recordings were done either in a Regular office room or a home, and not in a controlled lab environment. Subjects sat in the place and position of their comfort, and the environment was made as natural as possible. At the end of each three-minute condition, participants provided their perceived levels of focus and relaxation on a 10-point Visual Analogue Scale (VAS). They were instructed that a scale of 5 would mean their baseline rating before the start of that condition recording; thus, the VAS rating would give us a rating relative to their immediate baseline state.

doi.org/10.32388/HDMEQ5

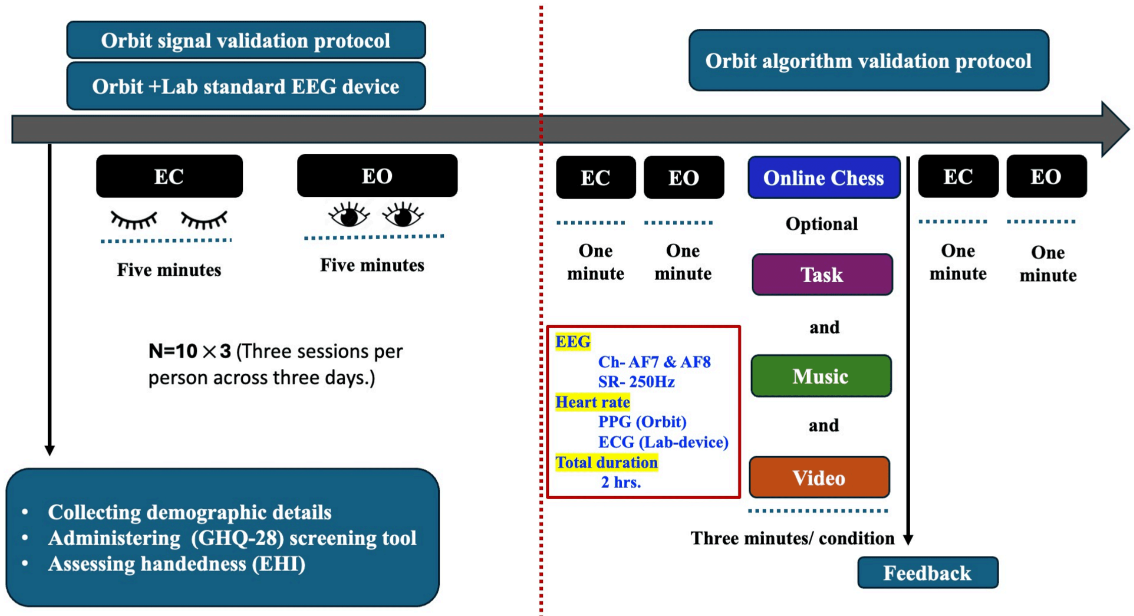


Figure 1. Study Design

Subjects wore the Neurostellar's wearable device on their own, and the investigator only intervened when hair came between the electrode contact and the skin. No skin preparation was done for the Neurostellar's wearable device, which used flat dry electrodes (conductive polymer with Ag-AgCl coating). The Lab device (portable battery-operated amplifier; xAMP-L10, Axxonet, India) captured two EEG channels close to the Neurostellar's wearable device EEG sites (AF7 & AF8), with ground and reference electrodes also in corresponding locations (midline frontal). Electrolyte paste was used, along with sintered Ag-AgCl ring electrodes. For the placement of the Lab device EEG electrodes, the area was cleaned with 70% isopropyl alcohol using gauze or cotton pads to remove dirt, sweat, oil, and cosmetics from the skin, mimicking the lab standard procedure. An additional ECG electrode (bottom electrode with ECG gel pad) was placed over the left shoulder. The Lab device was placed in a pouch in front of the abdomen (hung by a neck strap), and electrode cables ran down the neck before connecting to the amplifier. Subjects were in a sitting posture. Both devices recorded EEG at a sampling rate of 250 Hz; ECG was recorded at 250Hz by lab device and PPG was recorded at 62.5Hz by Neurostellar's wearable device. Lab Device streamed wirelessly (BLE) to a Laptop with proprietary acquisition software, which also provided time-locked audio instructions and task markers. Neurostellar's wearable device data was wirelessly streamed to another laptop using custom Python code and saved into a CSV file. To later sync the recordings between the devices, subjects were asked to make a left-right eye movement repeated three times at the beginning of each recording.

Condition	No of Recordings	Device Used
Rest (5 mins Eyes Closed & 5 mins Eyes Open)	**31** (8 subjects x 3 days + 7 subjects x 1 day)	Neurostellar's wearable device and Lab standard Device simultaneously
Playing **Chess** (preRest + Chess + postRest)	**17** (4 subjects x 3 days + 5 subjects x 1 day)	Only Neurostellar's wearable device
Performing WM **Task** (preRest + Task + postRest)	**30** (8 subjects x 3 days + 6 subjects x 1 day)	Only Neurostellar's wearable device
Listening to **Music** (preRest + Music + postRest)	**30** (8 subjects x 3 days + 6 subjects x 1 day)	Only Neurostellar's wearable device
Watching **Video** (preRest + Video + postRest)	**30** (8 subjects x 3 days + 6 subjects x 1 day)	Only Neurostellar's wearable device

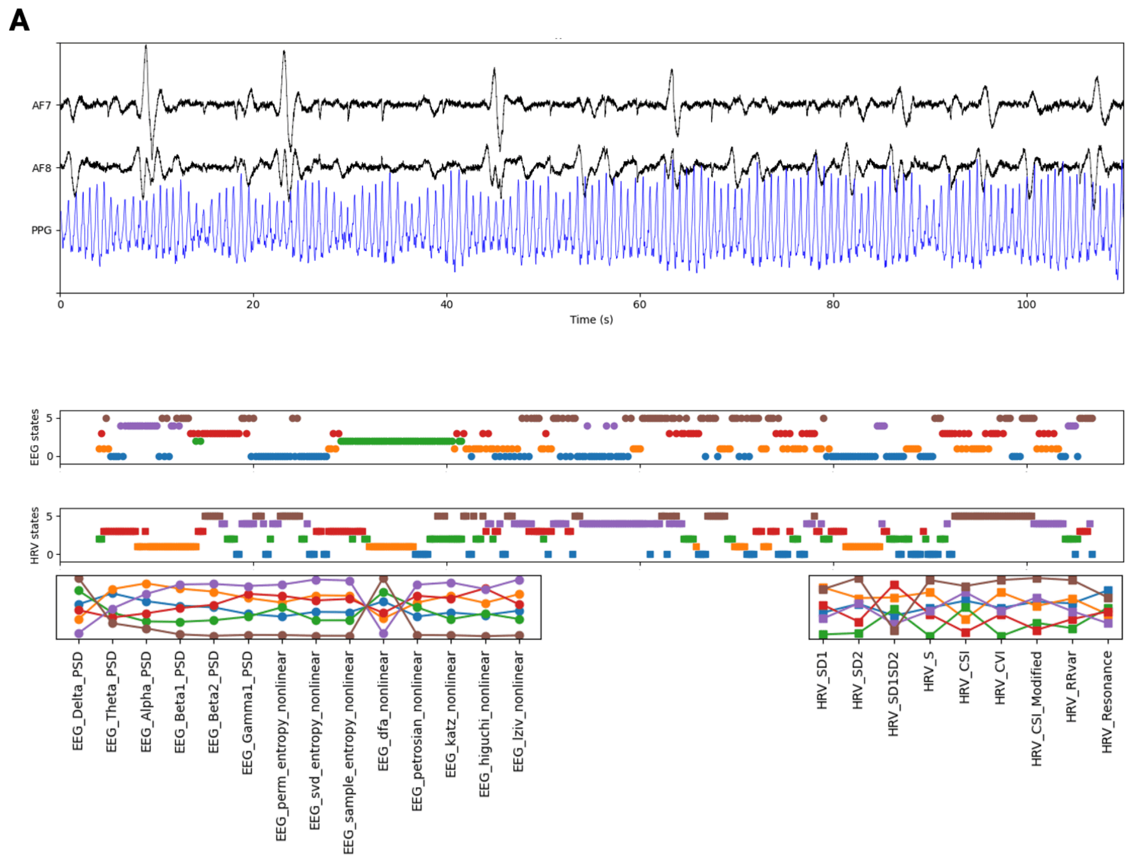
Table 2. Number of recordings

EEG and PPG Analysis

The EEG data (AF7 and AF8 channels) undergo basic preprocessing, including 1-40 Hz band-pass filtering (using a finite impulse response (FIR) filter) and 50 Hz notch filtering, via the "mne" library[30]. The PPG data preprocessing included 1-10 Hz band-pass filtering (FIR filter), 50 Hz notch filtering, a 3-sample moving median filter, and minmax normalisation. Both datasets are segmented into 16-second overlapping epochs with a 0.25-second sliding.

From EEG epochs, the first principal component analysis (PCA) component was chosen (this ensured a single time series representing at least 75% of the two-channel data segment), and EEG features were extracted using functions from multiple Python libraries like "scipy" and "antropy". The 18 EEG features extracted include power spectral density (welch approach; 4 second window with 50% overlap) in five frequency bands (Delta:1-4Hz, Theta:4-8Hz, Alpha:8-12, Beta:12-30Hz and Gamma:30-40Hz), the power ratios (Theta/Alpha, Theta/Beta, Beta/Alpha and Gamma/Alpha) and non-linear features (Permutation Entropy, Single Value Decomposition Entropy, Sample Entropy, Detrended Fluctuation Analysis, Petrosian Entropy, Katz Fractal Dimension, Higuchi Fractal Dimension and Lempel-Ziv Complexity). This multifeature selection was initially inspired from that used in sleep staging in a popular sleep stage scoring algorithm using single channel EEG[31] and further expanded with multiple nonlinear measures based on literature citing their relevance in detecting mental states like focus and mindwandering[32], especially when used in combinations[33]. From the PPG epochs, pulse peak detection[34], correction of erroneous peak placements based on outliers in peak-to-peak differences[35], quality assessment[36], deriving respiratory rate from heart rate[37][38], and extraction of HRV features (here we are using pulse rate variability as a surrogate of conventional HRV) were all done using the "neurokit2" Python library[39]. The 12 HRV features extracted include frequency-domain features (low frequency power [0.04-0.15Hz], high frequency power [0.15-0.4Hz], low frequency / high frequency power ratio and total HRV power), poincaré plot geometry-based non-linear features (like SD1, SD2, SD1/SD2 ratio, ellipse of SD1 and SD2, Cardio Sympathetic Index and Cardio Vagal Index[40][41]) and other statistical measures (like average heart rate, median average deviance of instantaneous heart rate and median average deviance of instantaneous respiratory rate). These short-time window features were used for signal quality correlations between the devices.

Additionally, we extracted a few novel higher-order multi-variate measures (complexity of quasi-stable multi-variate pattern sequences) from the above features derived across multiple consecutive time-windows (every 45s). This analysis was done using a subset of EEG and HRV features (14 EEG and 9 HRV features), wherein 6 clusters of feature combinations were determined that can explain the entire data (Figure 2B), using k-means clustering (an ensemble consensus of 50 k-means models was used) using the "scikit-learn" library. After initial exploration using methods like the Elbow plot, the number of clusters was refined to 6 through a trial-and-error process, wherein the feature combinations showed meaningful differences, showed consistency across repetitions and subjects, and all clusters represented a good portion of the data (>10%). These 6 cluster patterns were very consistent across different recordings and subjects, and hence a grand mean cluster (clustered from all data) was backfitted to generate cluster sequences (see Figure 2A, bottom panel for grand mean cluster pattern). The sequence of the EEG and HRV cluster dynamics was then subjected to complexity analysis using multiple standard measures (like Higuchi Fractal dimension, Lempel-Ziv Complexity, and Permutation Entropy), as well as a novel measure called Effort-To-Compress (ETC). ETC is defined as the number of iterations needed for a loss-less compression algorithm called Non-Sequential Recursive Pair Substitution (NSRPS) to transform the input sequence to a constant sequence[42]. ETC has been found to be more reliable to capture complexity of short and noisy time series, than some of the other measures[43][44]. The complexity of the sequence was calculated across the whole 3-minute recording to get the overall score, and also in short sliding windows (45s window with 1s sliding) to track the trend within a setting (Figure 2C). This sequence-based approach can be robust to poor signal quality or artefacts, as it relies on relations between features and not on the intensity of any single feature. For both EEG and HRV cluster sequences, this generated 12 features, which, when added to the individual features, made 45 multi-modal features available for every 45s segment.



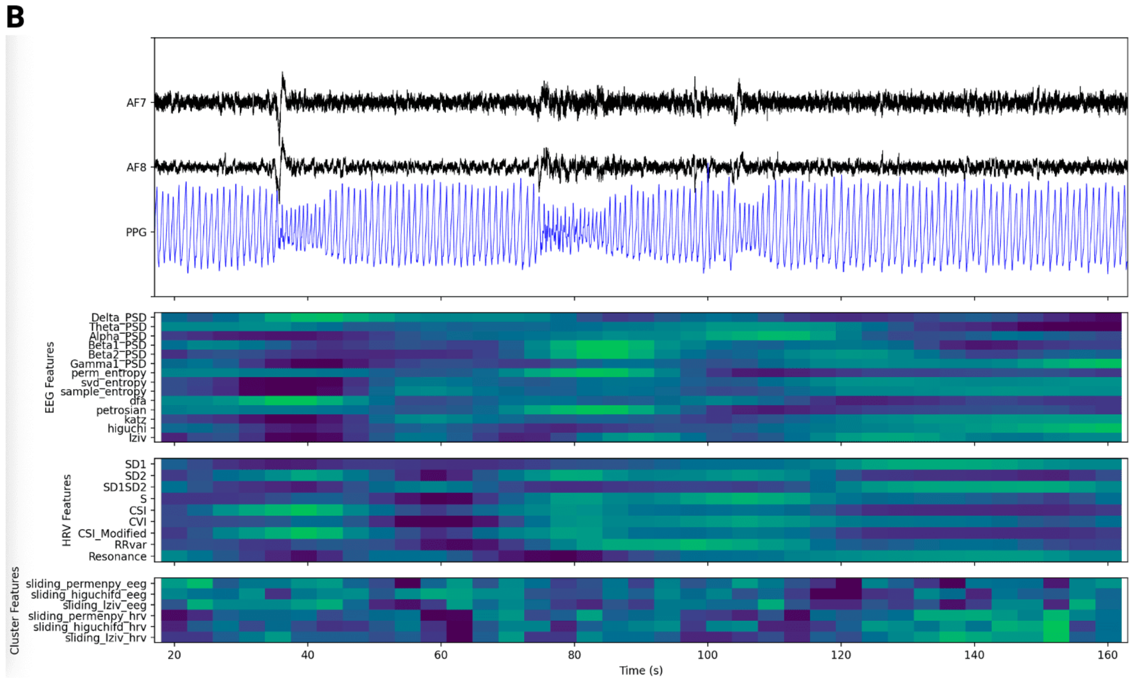


Figure 2. Multimodal Feature extraction from EEG and PPG data. (A) A 110s EEG and PPG segment is converted to quasi-stable multi-variate cluster sequences of the six EEG feature clusters (second panel) and six HRV feature clusters (third panel) based on distinct patterns of 14 EEG features (bottom left panel) and 9 HRV features (bottom right panel). Each colour represents a different cluster (Note that the colouring of EEG and HRV clusters is different). (B) Temporal profiles of the different multimodal features are shown from a data segment. Note that the large amplitude spikes in EEG waveforms or signal drop in PPG waveforms have a differential impact on the different features (including the novel cluster sequences), highlighting the robust nature of this approach. Feature values were z-scored for display (dark colour: low value; light colour: high value).

Machine Learning Analysis

For both classification and regression analysis, we used Random Forest and Gradient Boost models, as they are inherently explainable, non-linear, less impacted by value scaling, and consistently perform well in popular machine learning tasks. The 'sklearn' library was used. The above-mentioned 45 multi-modal features were used in the analysis.

Firstly, as a classification problem, we assessed the ability of the features to predict the condition labels (precondition rest versus that condition; e.g., pre-task rest vs Task). For this, a multivariate permutation testing approach was employed for each participant under each experimental condition. So, for every unique participant and condition combination, the feature set (X) consisted of the 45 features, and the target variable (y) was the two condition labels (precondition rest and the condition). Before model training, infinite values

doi.org/10.32388/HDMEQ5

were replaced with Not-a-Number (NaN), and all NaN values were subsequently imputed using the mean of the respective feature column for that participant-condition dataset. A Random Forest Classifier and a Gradient Boost Classifier model with default hyperparameters was used as the predictive model. The predictive performance of the model was evaluated using a 5-fold stratified cross-validation scheme (with data shuffling). The F1-weighted score was used as the primary performance metric, chosen to account for potential class imbalances in the subjective ratings. To determine the statistical significance of the observed F1-weighted score, a permutation test (n_permutations=1000) was conducted. In this procedure, the target variable (y) was randomly shuffled 1000 times, and the predictive model was trained and evaluated (using the same 5-fold stratified cross-validation) on each shuffled dataset. This generated a null distribution of F1-weighted scores. The p-value was calculated as the proportion of permutation scores that were greater than or equal to the actual (unshuffled) F1-weighted score. An effect size for the classification performance was calculated as the difference between the actual F1-weighted score and the mean of the F1-weighted scores obtained from the permuted datasets. This raw effect size was then standardised by dividing it by the standard deviation of the permutation scores, yielding a z-score. This entire procedure—data selection, preprocessing, model training, cross-validation, permutation testing, and effect size calculation—was repeated independently for each participant and each experimental condition, using both models.

Next, as a regression problem, we assessed the ability of the features to predict the various subjective ratings ('Focus', 'Relax', and their composite rating 'Relaxed Focus'). As with the classification problem, a multivariate permutation testing approach was independently employed for each subjective rating, but combining data from multiple subjects. For each subjective rating (the target variable, y), the corresponding 45 features were used as predictor variables (X). The dataset for each rating was partitioned into a training set (80%) and a testing set (20%) using a stratified split, ensuring all data from one subject stays in one set (train or test). No explicit feature scaling was applied to the training or testing data before the main model training and evaluation, as Random Forest models are generally insensitive to monotonic transformations of features. A Random Forest Regressor or Gradient Boosting Regressor model was employed for predicting each subjective rating. Hyperparameter tuning was conducted on the training data using a grid search approach with 5-fold cross-validation, again using a stratified approach ensuring subject integrity. The hyperparameters explored included n_estimators ([50, 100, 200]) and max_depth ([None, 10, 20]). The model yielding the highest R-squared (R^2) score during cross-validation was selected as the best model. The performance of the optimised model for each subjective rating was assessed on the held-out test set. Evaluation metrics included Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2. Feature importances were also extracted from the trained best model to identify the relative contribution of each objective measure to the prediction. The optimised model for each subjective rating was saved using joblib for persistence. To determine the statistical

doi.org/10.32388/HDMEQ5

significance of each optimised model, a permutation test (n_permutations=1000) was performed. This test was conducted on the full dataset (i.e., before train-test splitting) for each subjective rating to assess the overall relationship between the objective measures and the specific rating. The permutation test utilised a 5-fold cross-validation strategy (KFold with shuffling, stratified splitting) and R^2 as the scoring metric. The p-value was calculated as the proportion of R^2 scores from permuted datasets that were greater than or equal to the R^2 score obtained from the original (unpermuted) data. An effect size for the permutation test was calculated as the difference between the R^2 score on the original data and the mean R^2 score from the 1000 permuted datasets. This difference was also standardised by dividing it by the standard deviation of the permutation R^2 scores, providing a z-score-like measure of effect size.

Statistical Analysis

Each of the EEG features, HRV features, and the complexity of cluster sequences was averaged for each setting/session and subjected to statistical analysis. Statistical analysis was done in R-software using Robust statistical measures (takes care of outliers in the data and is immune to data distribution assumptions) from the "WLS2" library. One-sample test (one-sample percentile bootstrap test using the one-step M-estimator) was done using the 'onesampb' function, and Robust ANOVA followed by trimmed mean (10% trim) based post-hoc comparisons were performed. A p-value threshold of <0.05 was used for determining statistical significance, after correcting for multiple comparisons using the Holm method.

Results

Does Neurostellar's wearable device biosignal quality meet lab standards in real-world use?

Despite variations in amplifier profiles, electrode types, and skin preparation, concurrent recordings obtained with a laboratory standard device and the Neurostellar's wearable device demonstrated substantial similarity (Figure 3A & 3B). However, instances were observed where Neurostellar's wearable device recordings exhibited increased noise levels (Figure 3C & 3D) or laboratory standard device recordings were characterised by greater noise (Figure 3E & 3F).

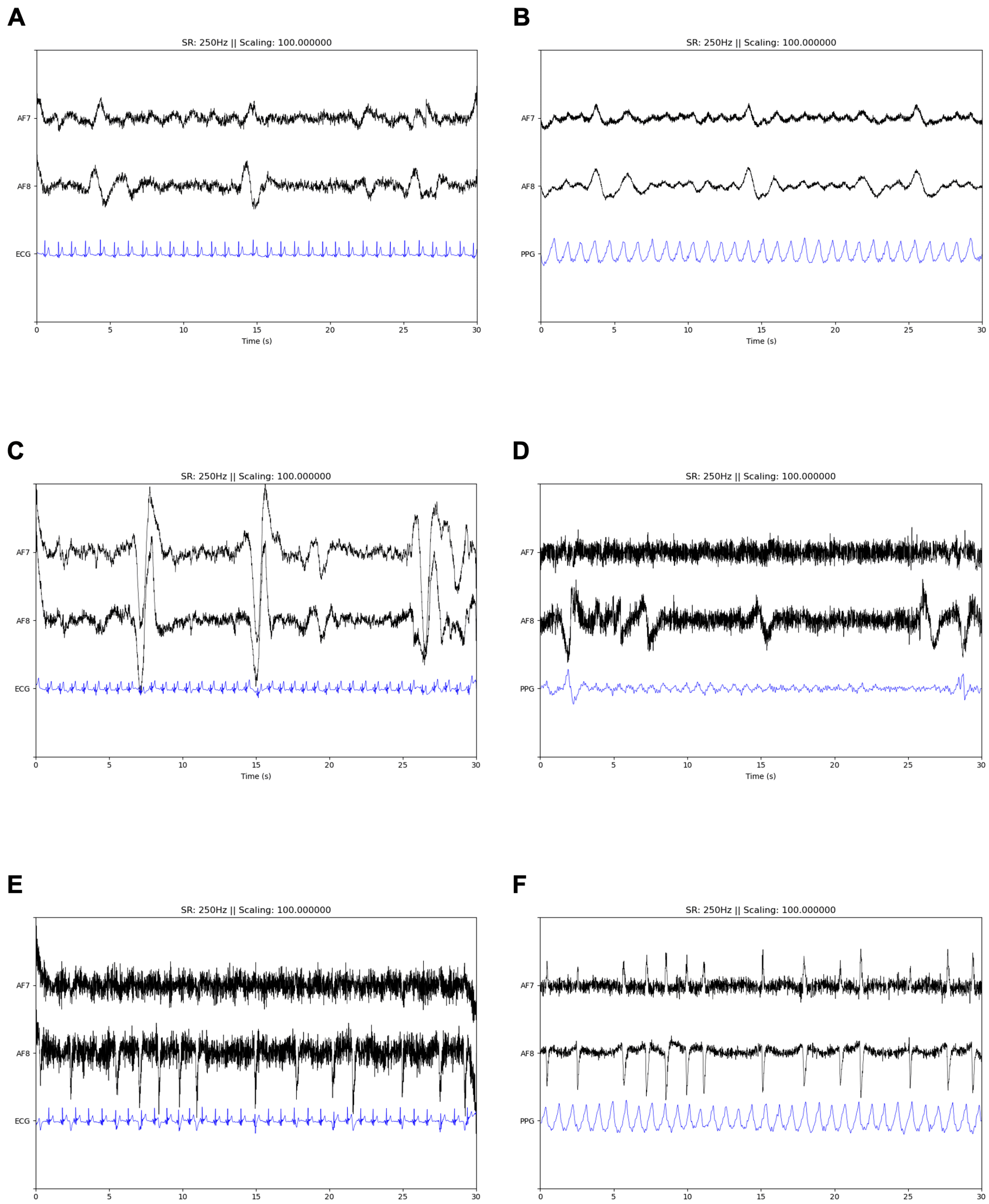
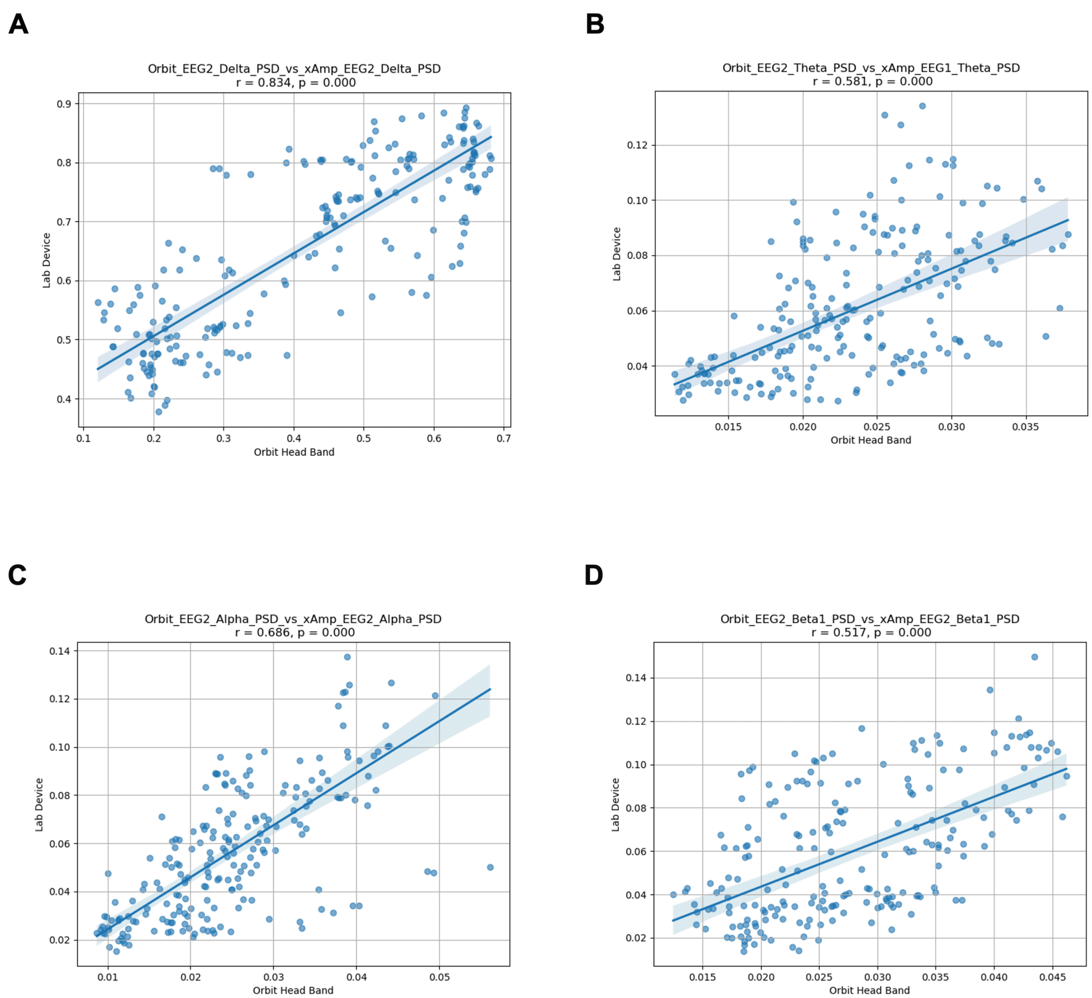


Figure 3. 30s snapshots showing time-locked EEG and ECG/PPG preprocessed data from Lab standard device (left) and Neurostellar's wearable device (right), from a representative subject. Some recordings were comparable between the two devices (A & B). Some were noisier in the Neurostellar's wearable device (C & D). Some were noisier in the Lab standard device (E & F).

 doi.org/10.32388/HDMEQ5

Although visual inspection of signal recordings revealed a high degree of concordance between devices, the strength of correlations varied across different extracted features. This observation was consistent for both electroencephalogram (EEG) features, as illustrated in Figure 4, and heart rate variability (HRV) features, as depicted in Figure 5.



 doi.org/10.32388/HDMEQ5

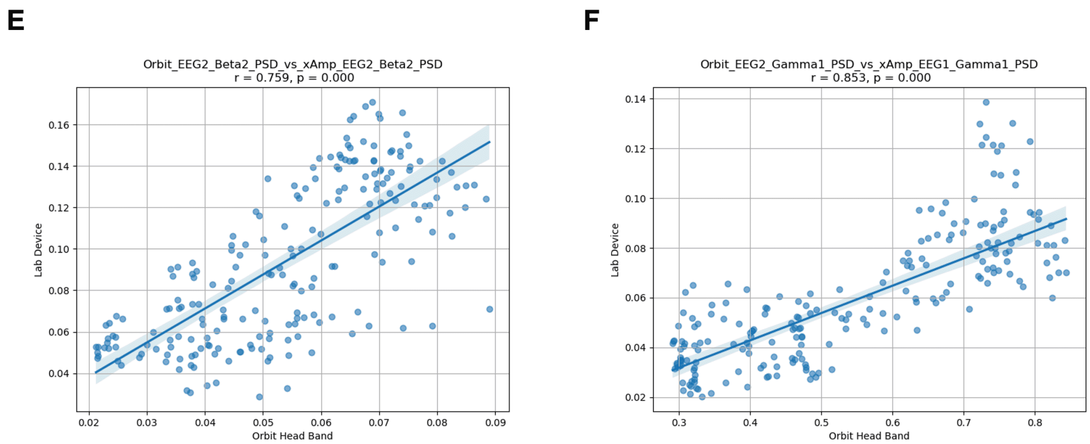
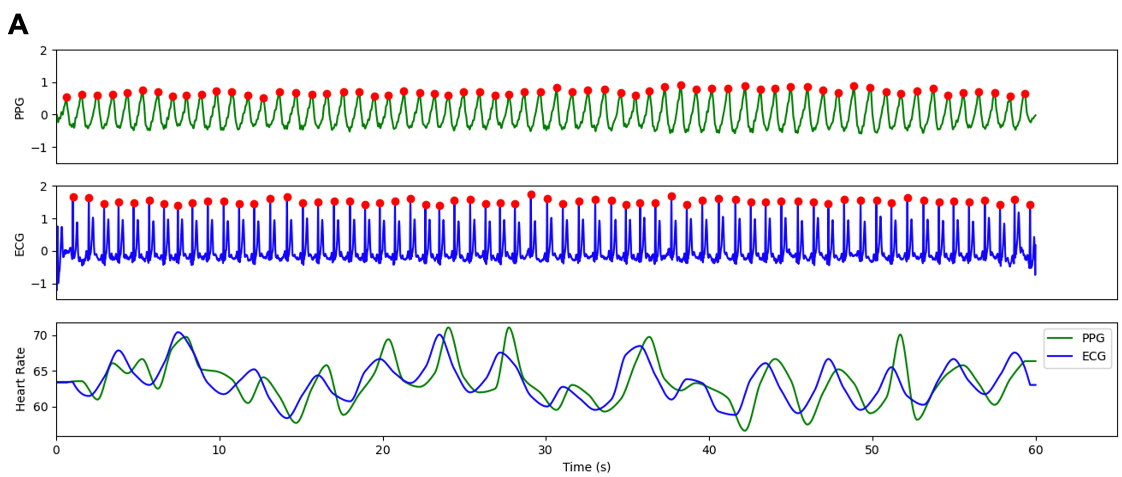


Figure 4. EEG Feature comparison from a single recording with comparable signal quality: Correlation plots of 8s EEG features between data from the Lab standard device and the Neurostellar's wearable device, for (A) Delta power, (B) Theta power, (C) Alpha power, (D) Beta1 power, (E) Beta2 power and (F) Gamma1 power.



doi.org/10.32388/HDMEQ5

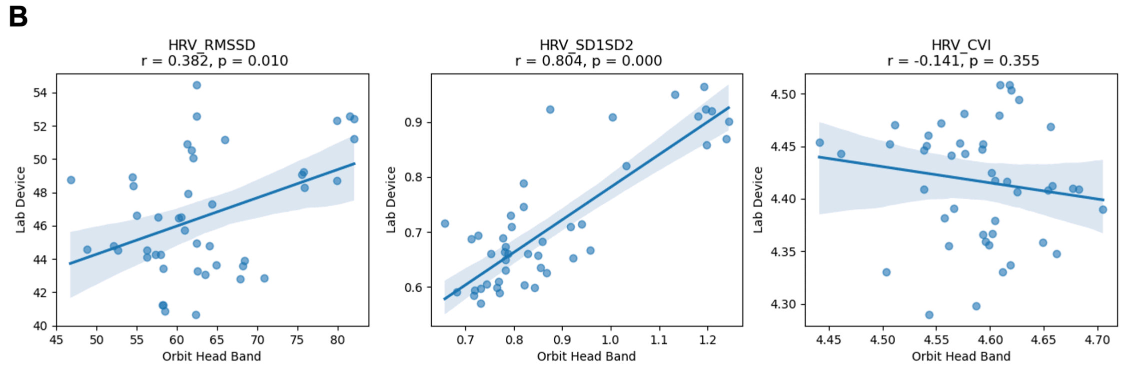


Figure 5. HRV Feature comparison from a single recording with comparable signal quality: (A) Peak detections (red dots) on a 60s data from PPG from the Neurostellar's wearable device (top panel; green waveform) and simultaneously acquired ECG from the lab device (middle panel; blue waveform). The instantaneous heart rates in beats-per-minute (bottom panel) are derived from PPG (green) and ECG (blue). (B) Correlation plots of 16s HRV features (RMSSD, SD1/SD2 ratio, and CVI) from the Lab standard device and the Neurostellar's wearable device are shown.

Analysis revealed the number of recordings exhibiting moderate to high correlation ($r > 0.5$ & $p < 0.05$) for each feature between devices, assessed separately for eyes closed and eyes open conditions (Figure 6). A minimum of 50% of recordings demonstrated such significant correlations in 12 of 18 EEG features and 9 of 12 HRV features during the eyes closed condition. Conversely, 14 of 18 EEG features and 6 of 12 HRV features displayed this in the eyes open condition. This implies that correlations also varied according to the condition during recording.

doi.org/10.32388/HDMEQ5

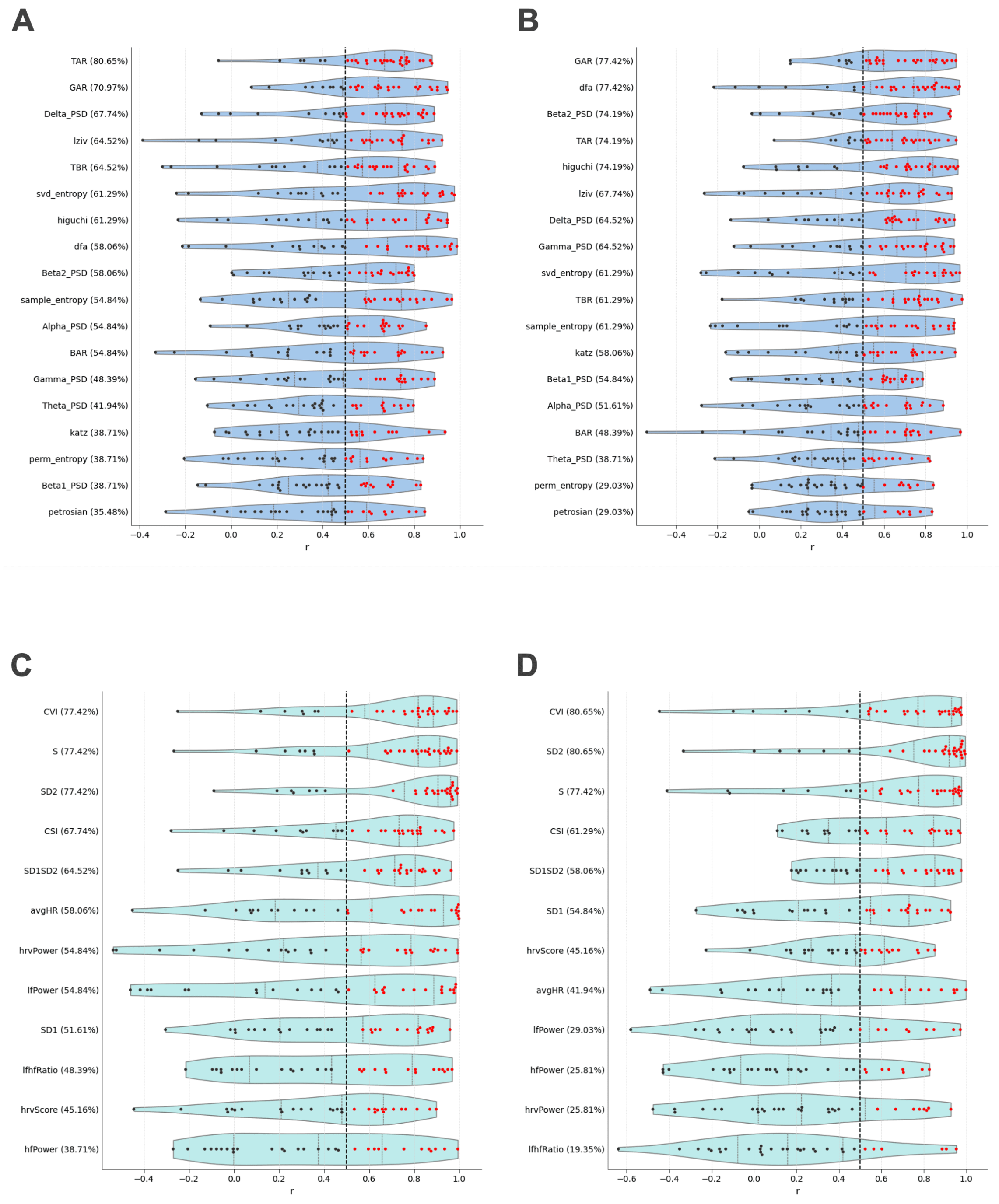


Figure 6. Recording-level correlation values for each features across all recording data from eyes closed (A & C) and eyes open conditions (B & D). EEG features are shown in the upper panels (A & B), and HRV parameters are in the lower panels (C & D). Each dot represents a single recording (n=31), with red dots representing recordings having moderate-to-high correlations (i.e., r>0.5 & p<0.05) and the percentage values in brackets show the proportion of such recordings, for each feature.

 doi.org/10.32388/HDMEQ5

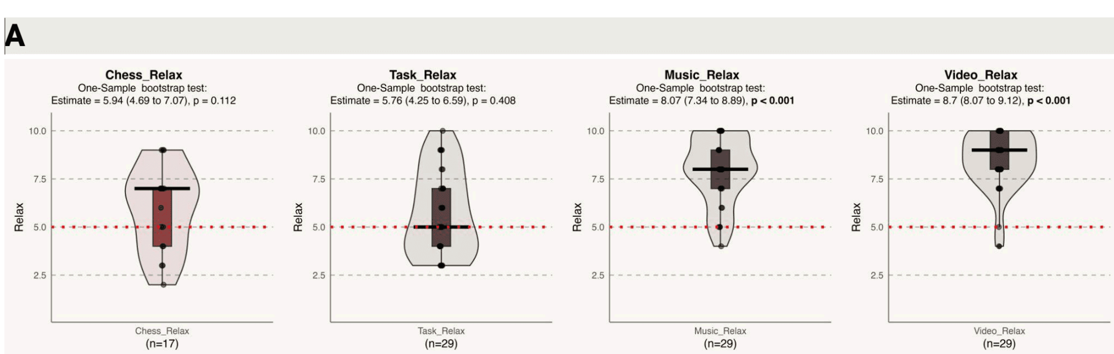
Does the four-condition protocol lead to varying degrees of focus and relaxation?

Participants utilised a 10-point Visual Analogue Scale (VAS) to evaluate their perceived levels of focus and relaxation immediately after the completion of different recording procedures. This quantitative method was implemented to validate the effectiveness of simultaneously collected electroencephalography (EEG) and heart rate variability (HRV) data, obtained from the Neurostellar's wearable device, in accurately reflecting these self-reported subjective states. On the VAS, ratings exceeding a score of 5 indicated a perceived increase in the respective state (focus or relaxation) compared to the participant's baseline level established before the commencement of each experimental condition. Conversely, ratings below 5 signified a perceived decrease relative to their baseline.

The analysis of the subjective relaxation ratings revealed statistically significant increases during both the Music listening session (t-statistic = 7.07, p-value < 0.001) and the Video viewing session (t-statistic = 8.70, p-value < 0.001) (Figure 7). These results suggest that both auditory and visual stimuli were effective in inducing a state of relaxation in the participants.

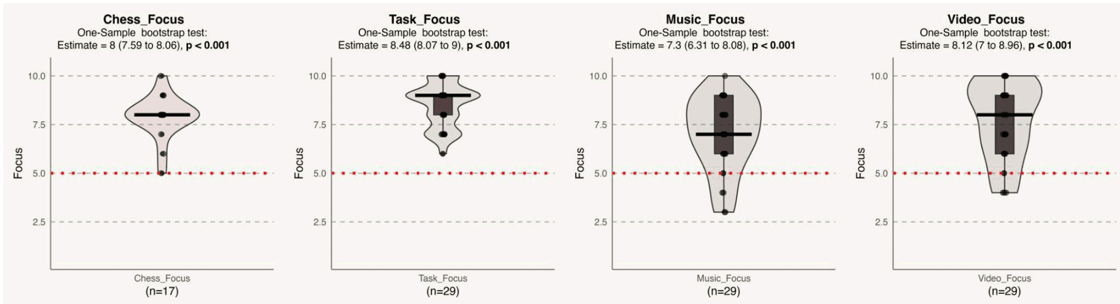
Subjective focus ratings also showed statistically significant increases across all experimental conditions (Chess: t-statistic = 8.00, p-value < 0.001; Task: t-statistic = 8.48, p-value < 0.001; Music: t-statistic = 7.30, p-value < 0.001; Video: t-statistic = 8.12, p-value < 0.001). Notably, the Task-oriented session elicited the highest focus ratings, while it was the lowest in the Music condition. This implies that while music enhanced focus to some extent, goal-directed tasks resulted in a greater perceived level of mental engagement.

Furthermore, participants rated a generally high score for overall comfort during the majority of the recording sessions (t-statistic = 8.25, p-value < 0.001). This indicates that the data acquisition process, involving the Neurostellar's wearable device and the experimental conditions, was well-tolerated and did not induce significant discomfort that might confound the physiological and subjective data. The consistency of high comfort ratings across most sessions suggests that the experimental setup was conducive to obtaining reliable and ecologically valid measurements of focus and relaxation.



doi.org/10.32388/HDMEQ5

B



C

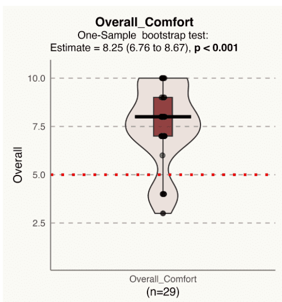


Figure 7. Box and violin plots illustrate subjective scores from VAS scales, collected after each condition: (A) Relax, (B) Focus or at the end of the recording: (C) Overall Comfort. The red dotted line at 5 indicates the baseline score, with scores above representing an increase and below a decrease from this baseline. A one-sample robust test was applied to each measure.

Do Neurostellar's wearable device EEG/HRV features effectively differentiate between various mental states?

An analysis of 45 EEG and HRV parameters from Neurostellar's wearable device data was conducted to differentiate between four experimental conditions (Task, Music, Video, and Chess) and baseline measures. Random Forest and Gradient Boosting machine learning algorithms were used. Model performance for each recording was rigorously evaluated using individual random permutation tests (1000 permutations with labels being permuted within each subject data) to determine significance beyond chance.

Both models demonstrated an F1-score of 1 in most recordings, indicating substantial effect sizes beyond chance across all conditions (Figure 8). Note that this analysis was done on each subject separately, and the purpose was to capture the change and not generalisability to another subject. Random Forest Classifier results: Task (f1-score difference = 0.492 ± 0.002, z = 13.11 ± 0.36, $p < 0.001$), Music (f1-score difference = 0.492

± 0.001, z = 13.10 ± 0.30, p < 0.001), Video (f1-score difference = 0.492 ± 0.002, z = 13.09 ± 0.44, p < 0.001), and Chess (f1-score difference = 0.492 ± 0.001, z = 13.08 ± 0.36, p < 0.001). Gradient Boosting Classifier results: Task (f1-score difference = 0.493 ± 0.005, z = 13.00 ± 0.37, p < 0.001), Music (f1-score difference = 0.493 ± 0.004, z = 13.07 ± 0.45, p < 0.001), Video (f1-score difference = 0.494 ± 0.006, z = 13.16 ± 0.38, p < 0.001), and Chess (f1-score difference = 0.493 ± 0.004, z = 13.08 ± 0.42, p < 0.001).

The consistently high z-scores and significant p-values (p < 0.001) provide strong statistical evidence for the models' ability to distinguish experimental conditions from baselines. These findings indicate that Neurostellar's wearable device data effectively captures meaningful multimodal feature variations during diverse activities in a naturalistic environment.

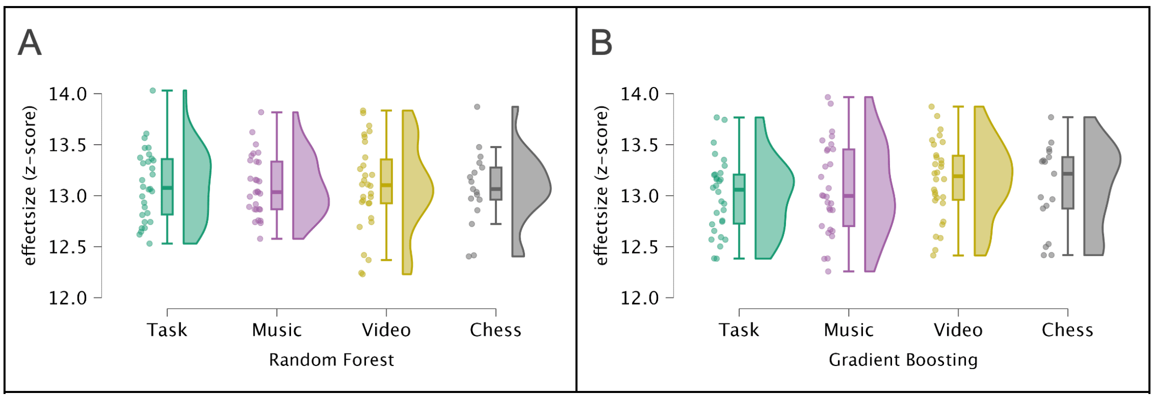


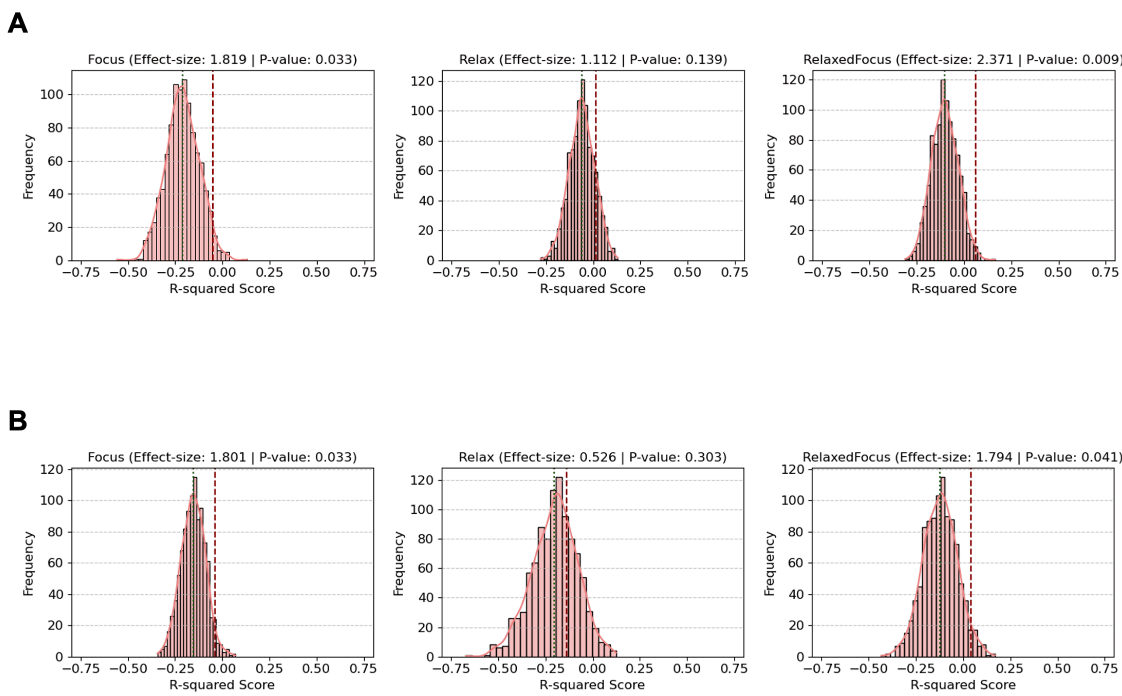
Figure 8: Classification accuracy based on multimodal EEG and HRV features between each condition (Performing Task, Listening to Music, Watching Video or Playing chess) and its pre-condition baseline, using Random Forest Classifier (A) and Gradient Boosting Classifier (B). Each dot represents the multivariate magnitude difference from a single recording computed as a z-score based on a permutation test (1000 permutations); all showed significant change (p<0.001). 106 recordings were used (Task=30, Music=29, Video=30, Chess=17).

Does Neurostellar's wearable device combined EEG and HRV data accurately indicate an individual's focus or relaxation levels?

Analysis of the combined EEG and HRV parameters from Neurostellar's wearable device data was further conducted to predict the subjective Focus, Relax, and RelaxedFocus (Focus x Relaxation) ratings across the four experimental conditions (Task, Music, Video, Chess) from the baseline condition. Random Forest and Gradient Boosting regression models were used. Model performance across the 102 recordings was computed a fs R^2 values, which were rigorously evaluated using random permutation tests (1000 permutations with labels being permuted) to determine significance beyond chance.

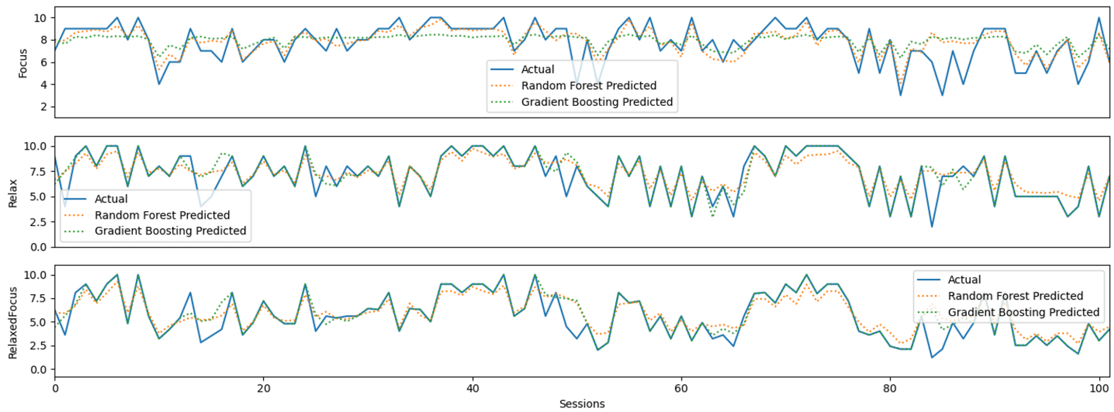
Both models demonstrated significant fit for some of the ratings, as shown in Figure 9. Random Forest Regressor results: Focus (R^2 change = 0.16, z = 1.82, p = 0.033), Relax (R^2 change = 0.07, z = 1.11, p = 0.139), and RelaxedFocus (R^2 change = 0.17, z = 2.37, p = 0.009). Gradient Boosting Regressor results: Focus (R^2 change = 0.12, z = 1.80, p = 0.033), Relax (R^2 change = 0.06, z = 0.53, p = 0.302), and RelaxedFocus (R^2 change = 0.16, z = 1.79, p = 0.041).

The high z-scores and significant p-values ($p < 0.05$) provide statistical evidence for the models' ability to predict the subjective rating of RelaxedFocus from baselines. These findings indicate that Neurostellar's wearable device data effectively captures meaningful multimodal feature variations for predicting a combined state of Relaxation and Focus.

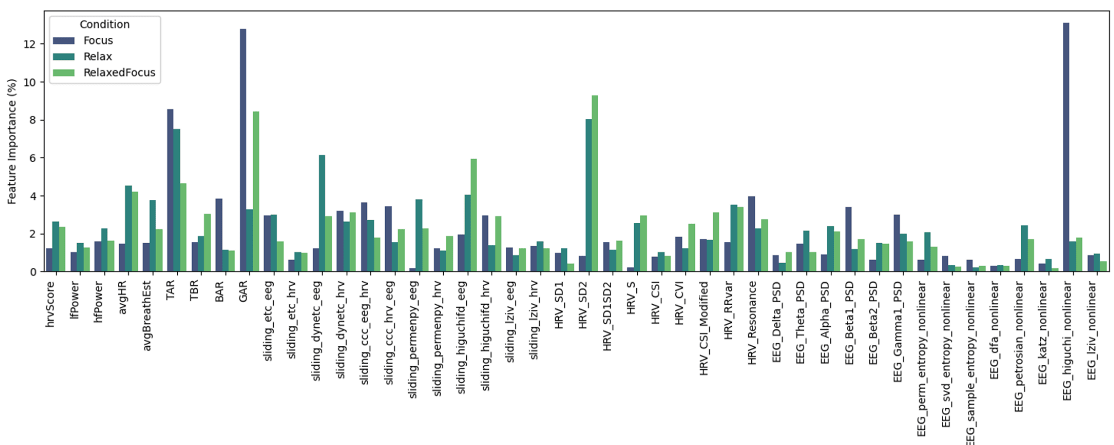


doi.org/10.32388/HDMEQ5

C



D



doi.org/10.32388/HDMEQ5

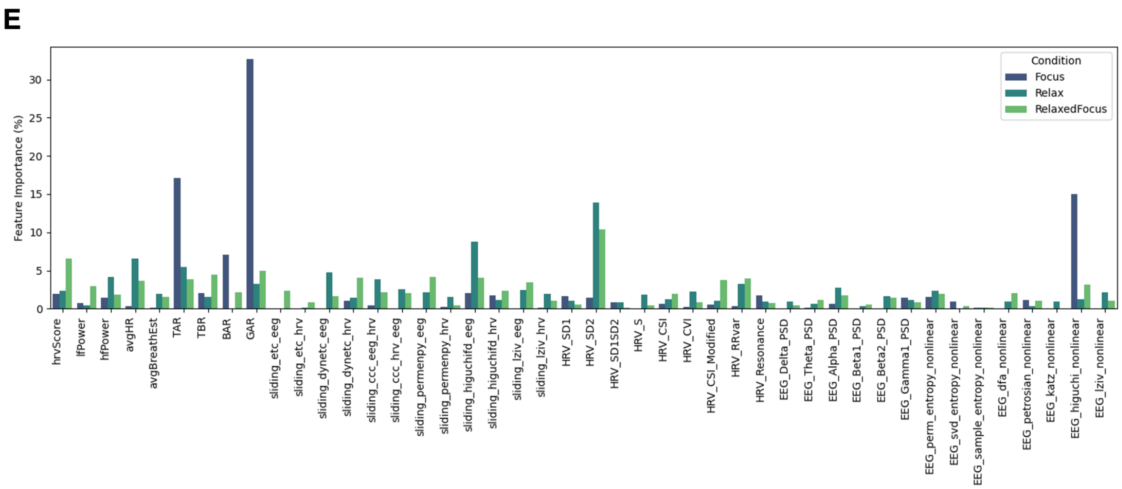


Figure 9. R-squared values showing the performance of the Random Forest (A) and Gradient Boosting (B) Regression models in predicting the subjective Focus, Relax and RelaxedFocus (Focus x Relaxation) ratings, based on multimodal EEG and HRV features from Neurostellar's wearable device. The histograms show the distribution of permutation R-squared values (1000 permutations), and the dotted green vertical line represents their mean (chance-level performance of the model). The dotted red vertical line represents the observed R-squared value (actual performance of the model). (C) Line plots showing the variation in Focus, Relax and RelaxedFocus ratings (blue) across the whole 102 recordings used, and how closely the predictions from Random Forest (orange dotted) and Gradient Boosting (green dotted) models match these. (D and E) Bar graphs showing the feature importances of the trained Random Forest and Gradient Boosting regressions models, respectively.

Discussion

This study aimed to evaluate the Neurostellar's wearable device EEG/PPG headband for real-world cognitive and mental state assessment, focusing on its biosignal quality compared to lab standards and its efficacy in differentiating and predicting subjective states using multimodal data analysis. The findings provide promising evidence for the utility of such consumer-grade devices in naturalistic settings (e.g., in a non-lab room environment unconstrained by cables and allowing non-rigorous natural body movements), albeit with important considerations.

Signal Quality of Neurostellar's wearable device in Real-World Use

The comparison between the Neurostellar's wearable device and a laboratory-standard device revealed a nuanced picture of signal quality. While concurrent recordings often showed substantial visual similarity (Figure 3A & 3B), indicative of the Neurostellar's wearable device capability to capture relevant biosignals, instances of increased noise were observed in both devices (Figure 3C-F). This highlights a critical point: noise is an inherent challenge in electrophysiological recordings, particularly in less controlled, real-world environments, and is not exclusive to consumer wearables[45]. The variability in noise underscores the importance of robust pre-processing and artefact handling in any real-world biosignal application.

The feature-level correlation analysis further elaborated on this. While a high degree of visual concordance was noted, the strength of correlations for extracted EEG and HRV features varied (Figure 4 & 5). Importantly, a significant number of EEG (5 in eyes-closed, 9 in eyes-open) and a majority of HRV features (10 in eyes-closed, 7 in eyes-open) exhibited moderate to high correlations ($r > 0.5$, $p < 0.05$) in at least 50% of recordings (Figure 6). This suggests that despite differences in hardware and application (e.g., dry vs. wet electrodes, amplifier characteristics), the Neurostellar's wearable device can reliably capture a considerable range of standard psychophysiological features, particularly for HRV, which showed strong concordance. The variation in correlation strength across features and conditions (eyes-open vs. eyes-closed) is expected, as different physiological states and features inherently possess different signal-to-noise ratios and susceptibility to artefacts. A prior study also reported a similar difference, demonstrating high correlation in frequency domain EEG features (mean=0.957) and acceptable correlation in time domain features (mean=0.580)[46]. A more recent study using in-ear dry electrodes even reported that the correlation with clinical-grade EEG varied based on conditions (0.76 during eyes closed rest and 0.39 during sleep) and overall level of movement and facial artefacts[47]. Therefore, the finding that over half the recordings yielded significant correlations for these features is encouraging for a consumer device designed for ease of use over stringent lab protocols.

Effectiveness of the Protocol in Eliciting Target Mental States

The subjective VAS ratings confirmed the efficacy of the four-condition protocol in modulating perceived levels of focus and relaxation. Significant increases in relaxation during Music and Video sessions, and in focus across all tasks (Chess, Task, Music, Video), validate that the chosen activities successfully induced the targeted subjective states. Notably, the Task condition elicited the highest focus, aligning with its demanding nature, while Music, often associated with relaxation, also enhanced focus, albeit to a lesser degree. This differentiation provides a solid basis for assessing the Neurostellar's wearable device ability to track these self-reported changes. Furthermore, the high overall comfort ratings ($t=8.25$, $p < 0.001$) are crucial, indicating that the Neurostellar's wearable device and experimental setup were well-tolerated, enhancing the ecological validity of the data by minimising discomfort as a confound.

Differentiating Mental States with Multimodal Features of Neurostellar's wearable device

The machine learning analyses demonstrated a strong capability of the wearable device's combined EEG and HRV features to differentiate between the four experimental conditions and baseline. Both Random Forest and Gradient Boosting classifiers achieved substantial effect sizes (all $z > 13.00$, $p < 0.001$; Figure 8) across all conditions. This is a key finding, indicating that the physiological changes captured by the Neurostellar's wearable device during diverse activities (ranging from cognitive tasks like Chess and the working memory Task to more passive engagement like Music and Video) are distinct enough to be reliably classified. It underscores the power of a multimodal approach, where combining brain activity and autonomic nervous system correlates (HRV) provides a richer, more robust representation of an individual's state than unimodal data alone. Studies have demonstrated the correlation between multimodal brain-computer interface metrics and HRV[48], and have shown that multimodal models incorporating both HRV and electrodermal activity offer more accurate estimates of sympathetic-driven arousal states than unimodal models[49]. The consistency across two different machine learning models further strengthens this conclusion.

Predicting Subjective Focus and Relaxation Levels

The study further ventured into predicting the intensity of self-reported focus and relaxation. The models demonstrated a statistically significant ability to predict a combined "RelaxedFocus" metric (Random Forest: $z = 2.37$, $p = 0.009$; Gradient Boosting: $z = 1.79$, $p = 0.041$), showing that the multimodal data of Neurostellar's wearable device can track this nuanced, combined state with considerable accuracy. The prediction of individual Relax ratings was not significant, while that of Focus was statistically significant but with smaller effect sizes. This suggests that while individual components of mental state can be predicted, a combined metric like "RelaxedFocus" might capture a more stable or physiologically distinct pattern reflected in the combined EEG/HRV features. The subjective experience of "focus" or "relaxation" in isolation may have more idiosyncratic physiological correlates, or their combination may represent a more clearly definable psychophysiological state. Prior studies have also found that the combined use of EEG and HRV features is good for a more definable psychophysiological state, like stress[50], and that higher-order complexity features from EEG differ when predicting states like Focus and Relaxation separately[51].

Implications and Future Directions

Consumer-grade EEG/PPG devices like the Neurostellar's wearable device, combined with sophisticated multimodal data analysis, show significant potential for real-world mental state assessment, enabling differentiation of activity-related states and prediction of combined relaxation and focus. A key implication is the necessity for a paradigm shift in biosignal analysis, moving away from the pursuit of pristine, lab-standard signal quality in uncontrolled environments. Instead, the focus should increasingly be on robustly leveraging the rich information within multimodal features and understanding their complex interactions. Prior research already advocates such a paradigm shift in physical activity assessment from wearable accelerometer data[52][53] and mental state assessment from wearable HRV data[54]. Recent evidences also suggest that combining power spectral data and non-linear measures could reflect a deeper understanding of the multifaceted nature of brain function, which oscillations providing invaluable insights into the rhythmic and synchronized activity of local neural populations, while non-linear dynamics offering a powerful window into the global integration and computational complexity of the brain as a whole[55][56]. This approach, acknowledging the inherent noise in real-world data and incorporating the multifaceted nature of brain-body data, enhances resilience to artefacts and allows for the development of models adaptable to individual variability and the nuances of everyday conditions, better capturing subtle mental state shifts.

While the study demonstrates promise, variability in signal quality and feature correlations suggests ongoing opportunities for refinement in sensor technology, signal processing for noisy data, and feature selection. Future research should prioritise developing and validating advanced machine learning models that learn cross-modal relationships and temporal dependencies, alongside longitudinal studies to assess the evolution and stability of these multimodal signatures. Further investigation into specific feature interactions will provide deeper insights into the psychophysiological basis of mental states in naturalistic settings. Validation in larger, diverse populations across various everyday activities is crucial for translating these findings into impactful real-world applications.

While recent advancements have seen the successful application of sophisticated deep learning and foundation models for mental state prediction, the primary objective of this study was the foundational validation of the signal fidelity of Neurostellar's wearable device and the development of robust feature engineering techniques tailored for noisy, real-world multimodal data. Our approach, utilising well-established machine learning algorithms, aimed to provide interpretable benchmarks of the device's core capabilities in ecologically valid conditions. This characterisation of signal quality and the efficacy of specifically designed features in handling limited-channel, artefact-prone data are crucial prerequisites that can inform and enhance the subsequent application of more complex computational models to consumer-grade wearable biosignals.

Conclusion

In conclusion, this study provides compelling evidence that the EEG/PPG of Neurostellar's wearable device can capture physiologically relevant signals in real-world conditions, demonstrating substantial similarity to lab-standard equipment for many key HRV and several EEG features. When its multimodal data are analysed using robust feature engineering and appropriate machine learning frameworks, it can effectively differentiate between various mental states associated with common activities and significantly predict a combined subjective state of relaxation and focus. These results contribute to the growing body of evidence supporting the viability of consumer wearables for practical, everyday mental state monitoring, paving the way for more accessible and personalised psychophysiological insights.

Statements and Declarations

Conflicts of Interest

V.M., V.V., P.F.K., A.G., K.R., D.S., A.V. and K.V. are affiliated to Neurostellar. A.S. is neither affiliated to Neurostellar nor received direct financial compensation from Neurostellar for this study. All authors affirm that the study design, data collection, analysis, interpretation, and manuscript preparation were conducted with scientific objectivity and integrity.

Data Availability

The data required for evaluating the study would be made available by the corresponding author on reasonable request.

Acknowledgements

This study received no external funding. It was supported by Neurostellar, the developer of the wearable device, by providing infrastructure and the device. A.S. arranged the lab-standrard device.

References

1. ^*de Barros EC (2024). "Understanding The Influence Of Digital Technology On Human Cognitive Functions: A Narrative Review." IBRO Neurosci Rep.* ***17****:415–422. doi:10.1016/j.ibneur.2024.11.006.*
2. ^*Firth J, Torous J, Stubbs B, Firth JA, Steiner GZ, Smith L, Alvarez-Jimenez M, Gleeson J, Vancampfort D, Armitage CJ, Sarris J (2019). "The "Online Brain": How The Internet May Be Changing Our Cognition." World Psychiatry.* ***18*** *(2):119–129. doi:10.1002/wps.20617.*

3. ^Shanmugasundaram M, Tamilarasu A (2023). "The Impact Of Digital Technology, Social Media, And Artificial Intelligence On Cognitive Functions: A Review." Front Cognit. **2**. doi:10.3389/fcogn.2023.1203077.
4. ^Montero-Marin J, Prado-Abril J, Piva Demarzo MM, Gascon S, García-Campayo J (2014). "Coping With Stress And Types Of Burnout: Explanatory Power Of Different Coping Strategies." PLoS ONE. **9**(2):e89090. doi:10.1371/journal.pone.0089090.
5. ^Oishi S (2019). "Reevaluating The Strengths And Weaknesses Of Self-Report Measures Of Subjective Well-Being." https://www.semanticscholar.org/paper/Reevaluating-the-Strengths-and-Weaknesses-of-Self-Oishi/0d063b91daf0bc252b318372200a3aba72fc298a.
6. ^Morgado FFR, Meireles JFF, Neves CM, Amaral ACS, Ferreira MEC (2017). "Scale Development: Ten Main Limitations And Recommendations To Improve Future Research Practices." Psicol Reflex Crit [Psychology, Reflection and Criticism]. **30**:3. doi:10.1186/s41155-016-0057-1.
7. ^Sparby T, Edelhäuser F, Weger UW (2020). "The Sense Of Certainty And The Reliability Of First Person Reports: An Investigation Using Micro-Phenomenological Self-Inquiry." Scand J Psychol. **61**(6):775–783. doi:10.1111/sjop.12671.
8. ^Sparby T, Leass M, Weger UW, Edelhäuser F (2023). "Training Naive Subjects In Using Micro-Phenomenological Self-Inquiry To Investigate Pain And Suffering During Headaches." Scand J Psychol. **64**(1):60–70. doi:10.1111/sjop.12858.
9. ^Ziegler M, Poropat A, Mell J (2014). "Does The Length Of A Questionnaire Matter?" J Individ Differ. https://econtent.hogrefe.com/doi/10.1027/1614-0001/a000147.
10. ^Chakrabartty SN (2023). Psychology And Mental Health Care. 8.
11. ^Scognamiglio P, Tully PJ, Khosravi M, Monaco F (2023). "Editorial: Psychometrics In Psychiatry 2022: Psychological Therapy And Psychosomatics." Front Psychiatry. **14**:1295716. doi:10.3389/fpsyt.2023.1295716.
12. a, bBorghare PT, Methwani DA, Pathade AG (2024). "A Comprehensive Review On Harnessing Wearable Technology For Enhanced Depression Treatment." Cureus. **16**(8):e66173. doi:10.7759/cureus.66173.
13. ^Tsirmpas C, Andrikopoulos D, Fatouros P, Eleftheriou G, Anguera JA, Kontoangelos K, Papageorgiou C (2022). "Feasibility, Engagement, And Preliminary Clinical Outcomes Of A Digital Biodata-Driven Intervention For Anxiety And Depression." Front Digit Health. **4**. doi:10.3389/fdgth.2022.868970.
14. ^Alhejaili R, Alomainy A (2023). "The Use Of Wearable Technology In Providing Assistive Solutions For Mental Well-Being." Sensors. **23**(17):Article 17. doi:10.3390/s23177378.
15. ^Khakurel J, Pöysä S, Porras J (2017). "The Use Of Wearable Devices In The Workplace—A Systematic Literature Review." In: Gaggi O, Manzoni P, Palazzi C, Bujari A, Marquez-Barja JM, editors. Smart Objects And Technologies For Social Good. Springer International Publishing. p. 284–294. doi:10.1007/978-3-319-61949-1_30.

16. ^Maltseva K (2020). "Wearables In The Workplace: The Brave New World Of Employee Engagement." Bus Horiz. **63**(4):493–505. doi:10.1016/j.bushor.2020.03.007.

17. a, bMikhaylov D, Saeed M, Husain Alhosani M, F. Al Wahedi Y (2024). "Comparison Of EEG Signal Spectral Characteristics Obtained With Consumer- And Research-Grade Devices." Sensors. **24**(24):Article 24. doi:10.3390/s24248108.

18. ^Di Flumeri G, Giorgi A, Germano D, Ronca V, Vozzi A, Borghini G, Tamborra L, Simonetti I, Capotorto R, Ferrara S, Sciaraffa N, Babiloni F, Aricò P (2023). "A Neuroergonomic Approach Fostered By Wearable EEG For The Multimodal Assessment Of Drivers Trainees." Sensors. **23**(20):Article 20. doi:10.3390/s23208389.

19. a, bMark JA, Curtin A, Kraft AE, Ziegler MD, Ayaz H (2024). "Mental Workload Assessment By Monitoring Brain, Heart, And Eye With Six Biomedical Modalities During Six Cognitive Tasks." Front Neuroergon. **5**:1345507. doi:10.3389/fnrgo.2024.1345507.

20. ^Mullen TR, Kothe CAE, Chi M, Ojeda A, Kerth T, Makeig S, Jung T-P, Cauwenberghs G (2015). "Real-Time Neuroimaging And Cognitive Monitoring Using Wearable Dry EEG." IEEE Trans Biomed Eng. **62**(11):2553–2567. doi:10.1109/TBME.2015.2481482.

21. ^Anders C, Arnrich B (2022). "Wearable Electroencephalography And Multi-Modal Mental State Classification: A Systematic Literature Review." Comput Biomed. **150**:106088. doi:10.1016/j.compbiomed.2022.106088.

22. ^Fekri Azgomi H, F. Branco LR, Amin MR, Khazaei S, Faghih RT (2023). "Regulation Of Brain Cognitive States Through Auditory, Gustatory, And Olfactory Stimulation With Wearable Monitoring." Sci Rep. **13**(1):12399. doi:10.1038/s41598-023-37829-z.

23. ^Friedman N, Fekete T, Gal K, Shriki O (2019). "EEG-Based Prediction Of Cognitive Load In Intelligence Tests." Front Hum Neurosci. **13**. doi:10.3389/fnhum.2019.00191.

24. ^Moontaha S, Kappattanavar A, Hecker P, Arnrich B (2023). "Wearable EEG-Based Cognitive Load Classification By Personalized And Generalized Model Using Brain Asymmetry." Proc 16th Int Jt Conf Biomed Eng Syst Technol. 41–51. doi:10.5220/0011628300003414.

25. ^Basile LFH, Anghinah R, Ribeiro P, Ramos RT, Piedade R, Ballester G, Brunetti EP (2007). "Interindividual Variability In EEG Correlates Of Attention And Limits Of Functional Mapping." Int J Psychophysiol. **65**(3):238–251. doi:10.1016/j.ijpsycho.2007.05.001.

26. ^Huang G, Zhao Z, Zhang S, Hu Z, Fan J, Fu M, Chen J, Xiao Y, Wang J, Dan G (2023). "Discrepancy Between Inter- And Intra-Subject Variability In EEG-Based Motor Imagery Brain-Computer Interface: Evidence From Multiple Perspectives." Front Neurosci. **17**. doi:10.3389/fnins.2023.1122661.

27. ^Agarwal N, Chinta U, Bhimanapati VBR, Jain S, Jain S (2021). "EEG Based Focus Estimation Model For Wearable Devices." SSRN Scholarly Paper No. 4982710. Social Science Research Network. https://papers.ssrn.com/abstract=

4982710.

28. ^Bréchet L, Michel CM (2022). "EEG Microstates In Altered States Of Consciousness." Front Psychol. **13**. doi:10.3389/fpsyg.2022.856697.

29. ^Zhao M, Jia W, Jennings S, Law A, Bourgon A, Su C, Larose M-H, Grenier H, Bowness D, Zeng Y (2024). "Monitoring Pilot Trainees' Cognitive Control Under A Simulator-Based Training Process With EEG Microstate Analysis." Sci Rep. **14**(1):24632. doi:10.1038/s41598-024-76046-0.

30. ^Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Goj R, Jas M, Brooks T, Parkkonen L, Hämäläinen M (2013). "MEG And EEG Data Analysis With MNE-Python." Front Neurosci. **7**. doi:10.3389/fnins.2013.00267.

31. ^Vallat R, Walker MP (2021). "An Open-Source, High-Performance Tool For Automated Sleep Staging." eLife. **10**:e70092. doi:10.7554/eLife.70092.

32. ^Tang S, Li Z (2024). "EEG Complexity Measures For Detecting Mind Wandering During Video-Based Learning." Sci Rep. **14**(1):8209. doi:10.1038/s41598-024-58889-9.

33. ^Burns T, Rajan R (2015). "Combining Complexity Measures Of EEG Data: Multiplying Measures Reveal Previously Hidden Information." F1000Res. **4**:137. doi:10.12688/f1000research.6590.1.

34. ^Elgendi M, Norton I, Brearley M, Abbott D, Schuurmans D (2013). "Systolic Peak Detection In Acceleration Photoplethysmograms Measured From Emergency Responders In Tropical Conditions." PLoS ONE. **8**(10):e76585. doi:10.1371/journal.pone.0076585.

35. ^Lipponen JA, Tarvainen MP (2019). "A Robust Algorithm For Heart Rate Variability Time Series Artefact Correction Using Novel Beat Classification." J Med Eng Technol. **43**(3):173–181. doi:10.1080/03091902.2019.1640306.

36. ^Orphanidou C, Bonnici T, Charlton P, Clifton D, Vallance D, Tarassenko L (2015). "Signal-Quality Indices For The Electrocardiogram And Photoplethysmogram: Derivation And Applications To Wireless Monitoring." IEEE J Biomed Health Inform. **19**(3):832–838. doi:10.1109/JBHI.2014.2338351.

37. ^Charlton PH, Bonnici T, Tarassenko L, Clifton DA, Beale R, Watkinson PJ (2016). "An Assessment Of Algorithms To Estimate Respiratory Rate From The Electrocardiogram And Photoplethysmogram." Physiol Meas. **37**(4):610–626. doi:10.1088/0967-3334/37/4/610.

38. ^van Gent P, Farah H, van Nes N, van Arem B (2019). "HeartPy: A Novel Heart Rate Algorithm For The Analysis Of Noisy Signals." Transp Res Part F Traffic Psychol Behav. **66**:368–378. doi:10.1016/j.trf.2019.09.015.

39. ^Makowski D, Pham T, Lau ZJ, Brammer JC, Lespinasse F, Pham H, Schölzel C, Chen SHA (2021). "NeuroKit2: A Python Toolbox For Neurophysiological Signal Processing." Behav Res Methods. **53**(4):1689–1696. doi:10.3758/s13428-020-01516-y.

40. ^Jeppesen J, Beniczky S, Johansen P, Sidenius P, Fuglsang-Frederiksen A (2014). "Using Lorenz Plot And Cardiac Sympathetic Index Of Heart Rate Variability For Detecting Seizures For Patients With Epilepsy." Annu Int Conf IEEE Eng Med Biol Soc. 4563–4566. doi:10.1109/EMBC.2014.6944639.
41. ^Toichi M, Sugiura T, Murai T, Sengoku A (1997). "A New Method Of Assessing Cardiac Autonomic Function And Its Comparison With Spectral Analysis And Coefficient Of Variation Of R–R Interval." J Auton Nerv Syst. **62**(1):79–84. doi:10.1016/S0165-1838(96)00112-9.
42. ^Nagaraj N, Balasubramanian K, Dey S (2013). "A New Complexity Measure For Time Series Analysis And Classification." Eur Phys J Spec Top. **222**(3):847–860. doi:10.1140/epjst/e2013-01888-9.
43. ^Kathpalia A, Nagaraj N (2021). "Time-Reversibility, Causality And Compression-Complexity." Entropy. **23**(3):327. doi:10.3390/e23030327.
44. ^Nagaraj N, Balasubramanian K (2017). "Dynamical Complexity Of Short And Noisy Time Series." Eur Phys J Spec Top. **226**(10):2191–2204. doi:10.1140/epjst/e2016-60397-x.
45. ^Ahsan Awais M, Ward T, Redmond P, Healy G (2024). "From Lab To Life: Assessing The Impact Of Real-World Interactions On The Operation Of Rapid Serial Visual Presentation-Based Brain-Computer Interfaces." J Neural Eng. **21**(4):046011. doi:10.1088/1741-2552/ad5d17.
46. ^Jackson G, Radhu N, Yinming Sun, Tallevi K, Ritvo P, Daskalakis ZJ, Grundlehner B, Penders J, Cafazzo JA (2013). "Comparative Evaluation Of An Ambulatory EEG Platform Vs. Clinical Gold Standard." 2013 35th Annual International Conference Of The IEEE Engineering In Medicine And Biology Society (EMBC). 1222–1225. doi:10.1109/EMBC.2013.6609727.
47. ^Pazuelo J, Juez JY, Moumane H, Pyrzowski J, Mayor L, Segura-Quijano FE, Valderrama M, Le Van Quyen M (2024). "Evaluating The Electroencephalographic Signal Quality Of An In-Ear Wearable Device." Sensors. **24**(12):3973. doi:10.3390/s24123973.
48. ^Vasilyev V, Borisov V, Syskov A, Kublanov V (2019). "Case Study Of Interrelation Between Brain-Computer Interface Based Multimodal Metric And Heart Rate Variability." Proc 12th Int Jt Conf Biomed Eng Syst Technol. 532–538. doi:10.5220/0007694505320538.
49. ^Subramanian S, Brown E, Barbieri R (2020). "Multimodal Vs Unimodal Estimation Of Sympathetic-Driven Arousal States." 2020 Comput Cardiol Conf. doi:10.22489/CinC.2020.290.
50. ^Attar ET, Balasubramanian V, Subasi E, Kaya M (2021). "Stress Analysis Based On Simultaneous Heart Rate Variability And EEG Monitoring." IEEE J Transl Eng Health Med. **9**:1–7. doi:10.1109/JTEHM.2021.3106803.
51. ^Richer R, Zhao N, Amores J, Eskofier BM, Paradiso JA (2018). "Real-Time Mental State Recognition Using A Wearable EEG." 2018 40th Annual International Conference Of The IEEE Engineering In Medicine And Biology Society (EMBC). 5495–5498. doi:10.1109/EMBC.2018.8513653.

52. ^*Tiwari A, Cassani R, Kshirsagar S, Tobon DP, Zhu Y, Falk TH (2022). "Modulation Spectral Signal Representation For Quality Measurement And Enhancement Of Wearable Device Data: A Technical Note." Sensors. **22**(12):4579. doi:10.3390/s22124579.*

53. ^*Troiano RP, McClain JJ, Brychta RJ, Chen KY (2014). "Evolution Of Accelerometer Methods For Physical Activity Research." Br J Sports Med. **48**(13):1019–1023. doi:10.1136/bjsports-2014-093546.*

54. ^*Tiwari A, Falk TH (2021). "New Measures Of Heart Rate Variability Based On Subband Tachogram Complexity And Spectral Characteristics For Improved Stress And Anxiety Monitoring In Highly Ecological Settings." Front Signal Process. **1**:737881. doi:10.3389/frsip.2021.737881.*

55. ^*Balkenhol J, Händel B, Biswas S, Grohmann J, Kistowski JV, Prada J, Bosman CA, Ehrenreich H, Wojcik SM, Kounev S, Blum R, Dandekar T (2024). "Beyond-Local Neural Information Processing In Neuronal Networks." Comput Struct Biotechnol J. **23**:4288–4305. doi:10.1016/j.csbj.2024.10.040.*

56. ^*Guevara Erra R, Perez Velazquez JL, Rosenblum M (2017). "Neural Synchronization From The Perspective Of Non-Linear Dynamics." Front Comput Neurosci. **11**:98. doi:10.3389/fncom.2017.00098.*

Declarations

Funding: This study received no external funding. It was supported by Neurostellar, the developer of the Orbit device, by providing infrastructure and the Orbit device. A.S. provided the lab-standard device.

Potential competing interests: V.M., V.V., P.F.K., A.G. and K.R. are affiliated to Neurostellar. A.S. is neither affiliated to Neurostellar nor received direct financial compensation from Neurostellar for this study. All authors affirm that the study design, data collection, analysis, interpretation, and manuscript preparation were conducted with scientific objectivity and integrity.