Psychoneuroendocrinology 179 (2025) 107528

Contents lists available at ScienceDirect

Psychoneuroendocrinology

journal homepage: www.elsevier.com/locate/psyneuen

Body movements as biomarkers: Machine Learning-based prediction of HPA axis reactivity to stress

Luca Abel [a,*], Robert Richer [a], Felicitas Burkhardt [b], Miriam Kurz [b], Veronika Ringgold [b], Lena Schindler-Gmelch [c], Bjoern M. Eskofier [a,d], Nicolas Rohleder [b]

[a] *Machine Learning and Data Analytics Lab, Department Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Germany*
[b] *Chair of Health Psychology, Department of Psychology, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Germany*
[c] *Chair of Clinical Psychology and Psychotherapy, Department of Psychology, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Germany*
[d] *Institute of AI for Health, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany*

ARTICLE INFO

Keywords:
Cortisol
Movement
Posture
Stress
Machine Learning
TSST

ABSTRACT

Body movements and posture provide valuable insights into stress responses, yet their relationship with endocrine biomarkers of the stress response remains underexplored. This study investigates whether movement patterns during the Trier Social Stress Test (TSST) and the friendly-TSST (f-TSST) can predict cortisol reactivity. Using motion capturing, movement data from 41 participants were analyzed alongside salivary cortisol responses. Machine learning models achieved a classification accuracy of 65.2 % for distinguishing cortisol responders from non-responders and a regression mean absolute error of 2.94 nmol/l for predicting cortisol increase. Findings suggest that movement dynamics can serve as proxies of endocrine stress responses, contributing to objective, non-invasive stress assessment methods.

1. Introduction

Bodily movements carry valuable information about the inner state of an individual that goes beyond the observable physical actions, serving as a rich source of emotional and physiological cues. Subtle variations in gestures, posture, and gait can reflect underlying physiological and psychological processes that can be recognized by other humans. This ability is considered fundamental for survival, as movements are often detectable from greater distances than other signals such as vocalizations, allowing earlier anticipation and potential avoidance of threatening interactions (Dael et al., 2012; Ogren et al., 2019).

Previous research has shown that changes in inner states are measurable from the outside. For example, acute sickness can be detected from facial cues or body movements (Axelsson et al., 2018; Lasselin et al., 2020; Hansson et al., 2023). Furthermore, positive (e.g. happiness) and negative emotions (e.g. sadness, fear) can be detected through the observation of body posture and movements over time (Atkinson et al., 2004).

In the context of acute stress, body posture and movements have been described as promising indicators of pride and shame, which are associated with the origin of the stress response (Wallbott, 1998; Dickerson et al., 2004). Expansive postures, often associated with pride, and contracted postures, linked to shame, not only serve as social signals but also reflect underlying physiological states (Tracy and Matsumoto, 2008). For example, individuals who adopted a subordinate posture during a social-evaluative stress paradigm exhibited lower cortisol responses compared to those who maintained a dominant posture (Turan, 2015). Overall, changes in body posture and movements offer a promising window into the mechanisms of underlying stress responses. For that reason, researchers have proposed different approaches to leverage this information to gain insights into stress and stress-related constructs.

In the Trier Social Stress Test (TSST; Kirschbaum et al., 1993), thoracic movement is progressively reduced in healthy individuals, an effect that is more pronounced in individuals exhibiting a stronger cortisol response (Zito et al., 2019). This reduction is potentially associated with the so-called "freezing behavior", characterized by a reduction in heart rate and body sway, which can be experimentally triggered by presenting angry faces or threatening film scenes to individuals (Roelofs et al., 2010; Hagenaars et al., 2014). Similarly, exposing individuals to a social-evaluative stress test similar to the

* Corresponding author at: Machine Learning and Data Analytics Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Carl-Thiersch-Str. 2b, Erlangen 91052, Germany.
E-mail address: luca.abel@fau.de (L. Abel).

https://doi.org/10.1016/j.psyneuen.2025.107528
Received 8 April 2025; Received in revised form 23 June 2025; Accepted 25 June 2025
Available online 26 June 2025

Montreal Imaging Stress Task (Dedovic et al., 2005) led to reduced postural sway, indicating changes in postural control (Doumas et al., 2018). While the defensive freezing behavior in response to a threatening situation was reproduced in multiple studies (Buss et al., 2004; Mantella et al., 2008; Hagenaars et al., 2014; Noordewier et al., 2020), the connection to endocrine markers of stress remains unclear (Hashemi et al., 2021). Some studies found high basal cortisol levels in “strong freezers” (Buss et al., 2004; Mantella et al., 2008) while others showed that a delayed freezing recovery was linked to a low basal cortisol level (Niermann et al., 2017). Hashemi et al. (2021) found that threat-induced reductions in body sway were linked to a lower hair cortisol concentration. Hence, the link between hypothalamic-pituitary-adrenal (HPA) axis (re)activity and body posture and movements, especially in the context of social-evaluative acute stress, remains to be explored in detail.

Characterizing body posture and movement requires integrating information from multiple joints and body segments, resulting in high-dimensional data. Traditional statistical methods often struggle to model such complexity effectively. In contrast, machine learning (ML) techniques are well-suited for processing and analyzing high-dimensional data, being able to capture subtle differences in movement patterns with greater accuracy.

Previous research successfully employed ML techniques in the field of psychoneuroendocrinology by identifying complex relationships between physiological markers, behavioral indicators, and stress responses. For example, long-term ACTH (adrenocorticotropic hormone) increase or decrease was predicted with an accuracy of 81.2 % from self-reported quality of life and illness perception scores in a population of patients with breast cancer (Crumpei-Tanasă and Crumpei, 2021). Similarly, Dong et al. (2021) successfully used graph representation learning, a deep learning method, to predict salivary cortisol levels in pancreatic cancer patients based on wearable sensor data, including movement information. Baird et al. (2019) used speech features extracted during the TSST to predict sequentially measured cortisol levels, concluding that an acoustic-based approach is suitable for the prediction of cortisol as a stress marker. These studies underscore the potential of ML in capturing subtle, multivariate patterns in physiological and behavioral data.

In our previous research, we investigated movement and postural changes in response to the TSST and a modified control condition, the friendly-TSST (f-TSST; Wiemers et al., 2013). We demonstrated that exposure to acute psychosocial stress leads to “stress-induced movement inhibition”, a state of reduced bodily movements that is primarily characterized by changes in movements of the head, the upper extremities, and the trunk (Richer et al., 2024b). The implemented ML models allowed the classification of stress and the respective control condition with a mean accuracy of 73.4 %. However, in this prior work, stress was defined solely based on the experimental condition (TSST vs. f-TSST), rather than individual variations in HPA axis reactivity.

Although cortisol peaks delayed after stress onset, HPA axis activation begins during the stressor itself in response to social-evaluative threat (Dickerson and Kemeny, 2004). Movement changes observed during the TSST, such as reduced body sway or more contracted postures, may reflect this early activation phase. These behavioral adaptations can thus serve as external indicators of internal stress responses, providing insight into individual differences in cortisol reactivity.

To address this gap, we aimed to extend previous findings by examining whether movement patterns during the (f-)TSST can predict cortisol reactivity. By moving beyond condition-based classification and instead focusing on individual neuroendocrine responses, we want to gain deeper insights into how body posture and movements relate to different facets of the stress response. Specifically, we employed two complementary approaches: (1) classifying cortisol responders versus non-responders and (2) predicting the cortisol response using ML-based regression, as outlined in the overview Fig. 1. The first approach treats cortisol reactivity as a *categorical* outcome, dividing participants into responders (a cortisol increase of at least 1.5 nmol/l increase, according to Miller et al., 2013) and non-responders, which enables the identification of general movement patterns linked to stress-related physiological activation. Regression, in contrast, predicts the *magnitude* of cortisol increase, offering a more fine-grained analysis of how movement dynamics relate to the intensity of the HPA axis response. Our approach leverages ML techniques to explore the predictive relationships between movement dynamics and cortisol reactivity, contributing to the growing body of research at the intersection of computational behavioral science and biopsychology, ultimately advancing the development of objective, non-invasive methods for stress assessment.

2. Methods

2.1. Data acquisition

We recruited 41 young healthy individuals (N = 18 women, Age 24.0 ± 3.5 years, BMI 22.1 ± 2.0 kg/m^2). Detailed descriptions of the recruitment process, inclusion, and exclusion criteria were described in a previous publication (Richer et al., 2024b; *Main Study*). Exclusion criteria followed established studies and HPA axis assessment guidelines, covering age (<18 or >40), BMI (<18 or >30 kg/m^2), physical or

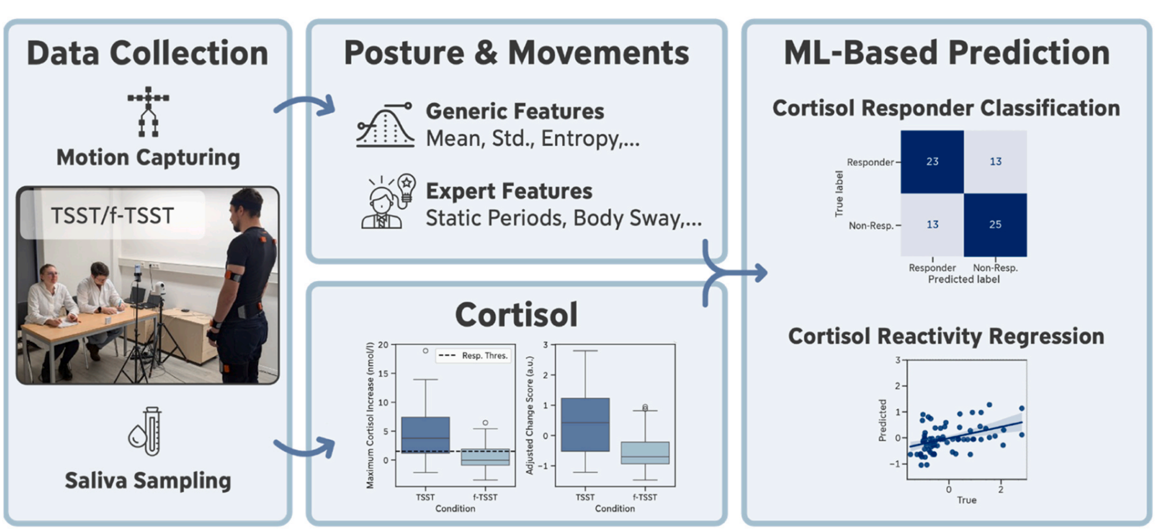


Fig. 1. Graphical Abstract: Body posture & movement data from an (f-)TSST study were used to predict the cortisol response and reactivity.

mental illness, medication use (e.g., beta-blockers, glucocorticoids, antidepressants), substance use (including cigarettes), elevated self-reported depression scores, and prior stress test experience.

All participants gave written informed consent, and the study was approved by the ethics committee of Friedrich-Alexander-Universität Erlangen-Nürnberg (protocol #493_20 B), following the principles of the Declaration of Helsinki.

Participants were exposed to the TSST and the f-TSST on two consecutive days in randomized order. Testing was conducted between 13:00 and 21:00 to minimize the influence of circadian rhythms (Smyth et al., 1997).

After arrival at the laboratory, the first saliva sample (S_0) was taken. All saliva samples were obtained using Salivettes (Sarstedt AG & Co. KG, Nümbrecht, Germany). Sample S_0 was only used to exclude participants with high cortisol levels on arrival, and not in any of the following evaluations. After assessing the required body measurements, for the Xsens MVN Awinda motion capture suit (Movella, Henderson, NV, USA), participants were equipped with the 17 sensors of the Xsens system. Following this, participants answered a set of pre-stress state questionnaires, whose results are not in the scope of this work. After the Xsens system was calibrated, the second saliva sample (S_1) was obtained right before the start of the (f-)TSST, which was conducted in a separate room.

The total duration of both tests was 15 min, split into three 5 min phases: Preparation, Speech and Mental Arithmetics. In contrast to the original f-TSST protocol, we modified the f-TSST to include a math task from the placebo-TSST (Het et al., 2009), where participants counted in steps of 15, starting from zero. During the (f-)TSST, the movement was recorded and subsequently exported as .mvnx files using the *Xsens MVN Analyze* software. The panel was equipped with a smartphone to log the timing of the (f-)TSST phases, allowing later separation of the Motion Capture recordings.

After finishing the (f-)TSST, participants were brought back to the first room, where six more saliva samples (S_2-S_7) were acquired at the respective time points (+15 min, +25 min, +35 min, +45 min, +60 min, +75 min relative to (f-)TSST start).

2.2. Cortisol assessment

The saliva samples were analyzed in the laboratory as described in previous publications (Janson and Rohleder, 2017; Ringgold et al., 2024). Samples with insufficient volume for cortisol concentration analysis were excluded ($n = 15$; 2.23 %). From the resulting cortisol values, we computed the maximum cortisol increase as a measure to quantify HPA axis reactivity (Miller et al., 2013). The maximum cortisol increase (Δc_{max}) was computed as:

$$\Delta c_{max} = \max(S_i) - S_1, \forall i \in [2, \ldots, 7],$$

where S_i is the cortisol value of the respective sample in nmol/l. For the classification task, cortisol responders were defined by an increase of Δc_{max} of at least 1.5 nmol/l, as proposed by Miller et al. (2013). In the initial regression experiment, we predicted Δc_{max} values without any additional transformation. In a subsequent regression experiment, the *law of initial value* (Wilder, 1962; Benjamin, 1963) was considered by regressing the standardized residuals, following the approach outlined in previous literature (Jin, 1992; Ramsay and Lewis, 2003; Miller et al., 2013). According to Dickerson and Kemeny (2004), who found cortisol peaks 21–40 min after stress onset, we set the cortisol values of S_1 and S_3 as baseline and peak levels and computed the change score for these values. A significant negative correlation (r = -.25, p = .033) between the change score and the baseline values (S_1) indicated that the change score is related to the baseline values, therefore the *law of initial value* should be considered (Miller et al., 2013). A regression model was then applied, with the change score as the dependent variable and the baseline value as the independent variable. The standardized residuals extracted from this model represent adjusted change scores, providing a refined measure of reactivity.

2.3. Body movement feature extraction

We extracted the same set of body movement features as utilized in a previous study (Richer et al., 2024b). To characterize movement during the (f-)TSST, raw motion data were aggregated over the entire test duration, and a range of features was computed. These features were derived from multiple data channels, including acceleration, velocity, angular velocity, and rotation, and were extracted for all individual body parts as well as predefined body part groups (e.g., Total Body, Upper Extremities).

The extracted features can be categorized into two main types. *Generic features* do not require prior domain knowledge and include statistical measures (e.g., mean, standard deviation) as well as signal characteristics, such as entropy. In contrast, expert features are designed to capture specific movement patterns identified in previous research or observed during data collection. Examples of expert features include static periods, which quantify episodes of little to no movement, and the distance of the hands to the face, which serves as an indicator of face-touching behavior. Static periods were obtained by computing the variance of the signal in a 1 s window (50 % overlap), if the variance was below a pre-defined threshold, we classified the window as static. Subsequently, neighboring windows were taken together to allow the computation of various metrics (e.g. maximum duration, mean duration, ratio, etc.). In total, we extracted 587 features (509 generic and 78 expert features). A comprehensive overview of all extracted features is provided in Table A.1.

In prior statistical analyses comparing conditions, we observed movement inhibition in the TSST condition compared to the f-TSST condition (Richer et al., 2024b). This effect was reflected in both generic features (e.g., a reduction in the mean velocity of the Total Body) and expert features (e.g., prolonged static periods of the Trunk) (Richer et al., 2024b).

2.4. Classification & regression analyses

We employed machine learning techniques to investigate associations between neuroendocrine markers and movement data. Two separate analyses were conducted: (1) the classification of cortisol responders and non-responders in both conditions based on movement data, and (2) the regression of Δc_{max} and the adjusted change scores using movement data. ML offers the possibility of exploring the large feature space we are dealing with by inherently including only features in the model that are relevant for the prediction of the desired outcome. Furthermore, we can ensure the model's predictive performance on unseen data through appropriate validation strategies, as described in detail in the following paragraphs.

Our ML model is constructed as a pipeline consisting of several steps for feature scaling, transformation, and selection before the final prediction is performed:

First, we removed all features with zero variance. Next, we applied feature scaling using either Min-Max Scaling (rescaling features to a [0,1] range) or Standard Scaling (transforming features to have a mean of zero and unit variance). Next, we performed a feature selection step to remove irrelevant or redundant features, reduce computational costs, and improve generalization to new data. This is a crucial step for certain, but not all, ML algorithms due to the "curse of dimensionality," where having too many features can make models less effective and harder to train (Bellman, 1966). For feature selection, we utilized Select-k-Best (SkB) based on the ANOVA F-value and Recursive Feature Elimination (RFE) with a linear Support Vector Machine (SVM) as the base classifier.

In ML, according to the “no free lunch theorem” (Wolpert and Macready, 1997), no algorithm is inherently superior across all possible tasks. Therefore, systematically evaluating different algorithms and optimizing their internal hyperparameters are critical steps. For both analyses, we evaluated seven machine learning models in their respective classifier and regressor variants: K-Nearest-Neighbors (kNN), Support Vector Machine (SVM), Decision Tree (DT), AdaBoost (Ada), Multi-Layer Perceptron (MLP), and Random Forest (RF). Hyperparameters were optimized using a grid search approach in which each hyperparameter combination is systematically compared. Due to the exceptionally large number of hyperparameters in Random Forest models, a randomized search strategy was applied instead, where n = 40,000 hyperparameter combinations are randomly selected.

To prevent overfitting during hyperparameter optimization, we implemented a cross-validation (CV) approach. Overfitting occurs when a machine learning model learns the training data too well, including noise and random patterns, instead of capturing the underlying trend. This leads to high accuracy on training data but poor performance on new, unseen data, reducing the model's ability to generalize. In *k*-fold CV, the dataset is split into *k* equally sized subsets, where *k-1* subsets are used as the training set and the remaining subset as the test set. This procedure is repeated *k* times (folds), and the final ML performance is determined by averaging across folds. This ensures that the model evaluation is always performed on unseen data, mitigating the risk of overly optimistic performance estimates. To prevent data leakage, which means the models learn patterns that are not associated with the target variable, movement data from the same participant (i.e., data from both conditions) was never included in both the training and test sets. The ML pipeline with the best-performing hyperparameter set (with regard to the target metric) is then retrained on the entire dataset to give the model as many data points to learn from as possible. This CV-based hyperparameter optimization is the most common approach in ML (Yarkoni and Westfall, 2017). The detailed hyperparameter grid can be found in Table A.2.

To ensure a robust model evaluation, we repeated this process by embedding the hyperparameter optimization CV into another 5-fold CV, where the dataset is split into an 80 % training and a 20 % test set *before* performing the hyperparameter optimization. This means that the hyperparameter optimization in the "inner" CV was performed exclusively on the training set of the "outer" CV loop. The ML pipeline with the best-performing hyperparameter set, which was retrained on the entire dataset of the "inner" hyperparameter optimization CV, was then evaluated on the training set of the "outer" model evaluation CV for each fold, ensuring that final evaluations were conducted on completely unseen data.

Integrating these two levels of "nested" CV minimizes overfitting, prevents data leakage, and allows for an unbiased evaluation of ML pipelines. This makes it particularly valuable for models with complex hyperparameter spaces, providing a robust estimate of model generalization performance.

Model performance was assessed using accuracy for classification tasks (responder vs. non-responder) and mean absolute error (MAE) for regression tasks (Δc_{max}). Accuracy, defined as the proportion of correctly classified samples out of the total samples, was used because it is easy to interpret and well-suited for balanced datasets as in our case (36 responders vs. 37 non-responders after the exclusion of incomplete measurements). MAE is one of the standard measures for regression tasks and was chosen to enable comparison with previous work.

The metrics were averaged across the outer model evaluation CV folds to estimate the model's generalization ability reliably. To further assess model performance on the entire dataset and to compare to previously reported results, we computed Spearman's correlation coefficient between predicted and actual cortisol values for the regression tasks.

To interpret model predictions and identify key features influencing classification and regression outcomes, we analyzed SHAP (SHapley Additive exPlanation) values using the SHAP Python package (v0.44.0; Lundberg and Lee, 2017). SHAP values quantify the contribution of each feature to model predictions, offering insights into their relative importance. Features with higher absolute SHAP values are considered to have a stronger influence on model decisions. A positive SHAP value indicates an increased likelihood of a positive classification, whereas a negative value suggests a lower probability. Analogously, for regression models, SHAP values indicate if a feature pushes the predicted output above or below the average prediction.

2.5. Availability of data and code

Raw data is available on OSF (https://osf.io/va6t3/; Richer et al., 2024a). The source code for data processing, feature extraction, and reproducing all results, figures, and tables is available on GitHub (https://github.com/empkins/movement-cort-prediction). All analyses were performed in Python (v3.9.8), using the packages BioPsykit (v0.12.2; Richer et al., 2021), pingouin (v0.5.5; Vallat, 2018), and scikit-learn (v1.2.2; Pedregosa et al., 2011).

3. Results & discussion

Two participants were excluded from the study: one due to incomplete participation in the TSST, and the other due to corrupted motion capture data. The final dataset comprised 73 data points from 37 participants (N = 16 women, Age 23.8 ± 3.6 years, BMI 22.0 ± 1.9 kg/m^2), with one f-TSST condition missing. A data point was defined as one condition (TSST or f-TSST) with complete cortisol and motion capture data from a single participant.

The maximum cortisol increase ranged from −3.5 to 18.9 nmol/l, and the range of the adjusted change scores was between −1.5 and 2.8. The distribution per condition for both measures can be seen in Fig. 2a and 2b. As shown in Fig. 2c, 29 % of the participants were classified as a cortisol non-responder in the TSST condition, similarly, 27 % of the participants in the friendly control condition were classified as cortisol responders. This highlights the need for a predictive model that does not rely on the condition to assess whether a stress reaction is present, as in our previous piloting work (Richer et al., 2024b), but one that can predict the actual physiological response, as performed in our following experimental analyses.

3.1. Classification cortisol responders vs. non-responders

Classification results showed a maximum accuracy of 65.2 ± 7.3 % using *MinMaxScaler* for scaling, *SelectKBest* for feature selection, and *RandomForestClassifier* as the classification model. The corresponding confusion matrix (Fig. 3), indicates that the classification of the two classes was equally possible. Detailed results for each model combination are listed in Table A.3. The accuracy is slightly lower than the previously reported 73.4 % for the condition classification (Richer et al., 2024b), which is expected given that there are various factors influencing the cortisol response, which will be discussed in detail in the limitation section.

To assess whether classification accuracy is influenced by the sex of the participants, we analyzed the accuracy for men and women separately. There was only a small difference in classification accuracy between male and female participants (61 % vs. 68 %), indicating that there is no sex-specific bias in the classification of cortisol (non-) responders.

Analyses of SHAP values revealed that out of the 40 top features, 31 were expert features (9 generic features). Previously reported results for the condition classification found notably fewer expert features in the top 40 (11 expert / 29 generic features; Richer et al., 2024b). This might be an indicator that the expert features are better suited at capturing the physiological response instead of the condition than the generic features. The body parts contributing most to the model output were upper extremities (12 features), head (10 features), as well as trunk and lower extremities (9 features each). When comparing this to the features contributing to the condition classification, notably more features from the Trunk (9 vs. 1) and Lower Extremities (9 vs. 2) were selected for the

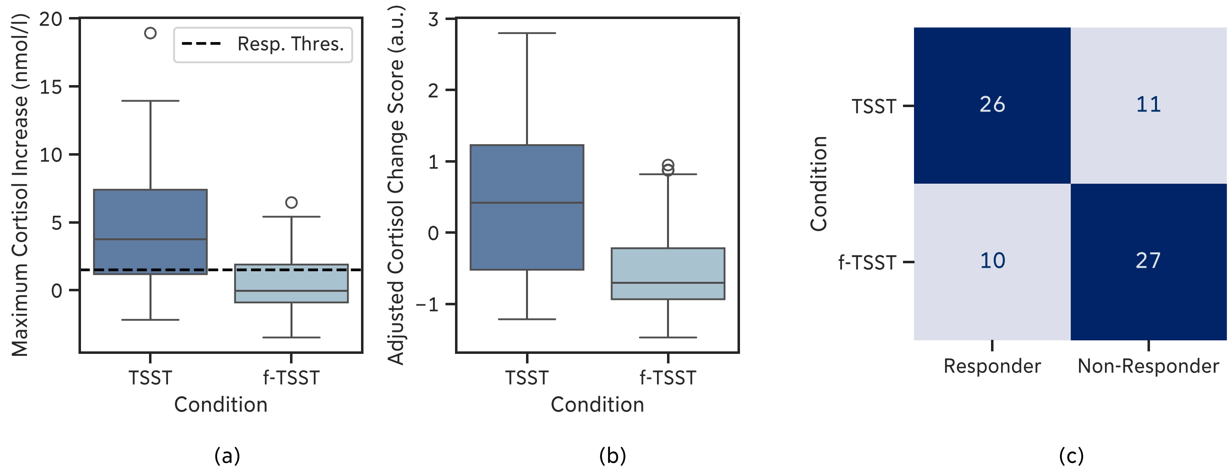


Fig. 2. (a) Maximum cortisol increase (Δc_{max}) per condition (Resp. Thres.: Responder Threshold). (b) Adjusted change scores per condition. (c) Cortisol responders/ non-responders per condition.

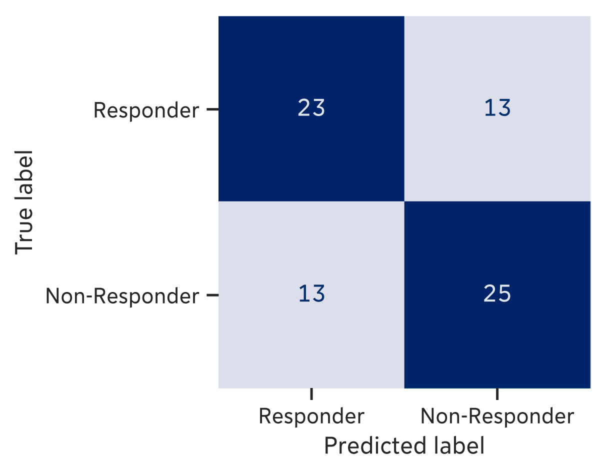


Fig. 3. Confusion matrix of responder vs. non-responder classification.

cortisol responder/non-responder classification, which is a starting point for further investigations. An overview of the SHAP values of the top 20 features is shown in Fig. 4, notable observations are: A reduced head movement could be associated with the responder class, showing for example in the feature ratio of static periods. A high distance of the hands seems to be associated to the non-responder class (negative SHAP values).

3.2. Prediction of the cortisol response

The best results for predicting the maximum cortisol increase (Δc_{max}) with ML-based regression were obtained using a pipeline consisting of *StandardScaler* for feature scaling, *SelectKBest* for feature selection, and *RandomForestRegressor* as the regression model. This model achieved an MAE of 2.95 $\pm$ 0.47 nmol/l averaged over all model selection folds.

A moderate correlation was achieved when evaluating the model performance for the prediction on Δc_{max} on the dataset level ($r = .36$, $p = .001$). This indicates that the cortisol response can be explained to some degree by body posture and movements during a stress event.

For the second regression analysis, predicting the adjusted change scores (standardized residuals) a mean absolute error of 0.70 $\pm$ 0.09 was achieved over all folds using a pipeline with *MinMaxScaler*, *SelectKBest* ($k = 18$), and *MLPRegressor*. 17.0 % of the variability of the adjusted change scores was explained by this model. The predicted vs. true adjusted change scores for the best-performing model can be found in Fig. 5a, and an overview of the performance of all pipeline combinations can be found in Table A.4.

We additionally recomputed the adjusted change scores while controlling for age, gender, and BMI. Hormonal contraceptive use and menstrual cycle phase were not included in this re-analysis, as participants using hormonal contraceptives were excluded and all female participants were tested during the luteal phase, thereby keeping these factors constant. Importantly, the newly computed scores were highly correlated with the original adjusted scores that corrected only for baseline levels ($r = .96$, $p < .001$), indicating minimal influence on the outcome. The results of our machine learning-based regression analysis remained stable when predicting these newly adjusted cortisol scores ($r = .45$).

The correlation coefficient ($r = .45$, $p < .001$) is notably higher ($r = .36$ vs $r = .45$; Z = -2.61, $p < .001$) than in the first analysis. Given that the adjusted change scores account for baseline differences (Ramsay and Lewis, 2003), this result indicates that the movement data during the (f-)TSST carries information about the true cortisol reactivity. The following in-depth analyses refer to the second regression experiment predicting the adjusted cortisol change scores.

Compared to previous research by Baird et al. (2019) who sequentially predicted cortisol samples using speech-based features, the correlation coefficient is slightly higher (.45 vs .42), which was obtained for the cortisol sample + 20 min after TSST start. These results indicate that movement and posture may encode information about HPA axis reactivity similar to that conveyed by speech features.

Analyses of the SHAP values of the 18 selected features (Fig. 5b) revealed that the body parts contributing the most to the prediction were head (9 features), upper extremities (7 features), and lower extremities (2 features). Selected features are similar to the classification task, with the 4 most contributing features in both models being different measures of static periods of the Head (maximum duration, mean duration, ratio, and standard deviation of the duration).

Interestingly, the distance between the hands, as well as the distance between the left hand and the head, were in the 18 selected features. In our previous work, these features did not show a significant difference for the conditions and were not in the top 18 features for the ML model predicting the condition (Richer et al., 2024b). A greater distance between the left hand and the head was associated with a positive SHAP value, indicating higher cortisol levels. This finding aligns with previous research by Turan (2015), which demonstrated that dominant postures – described as being "expansive, taking up more space with open limbs" – are linked to an increased cortisol response.

The paired boxplot (Fig. 6) suggests a systematic bias in the model's predictions of cortisol reactivity to acute psychosocial stress, where lower cortisol reactivity values tend to be overestimated, while higher

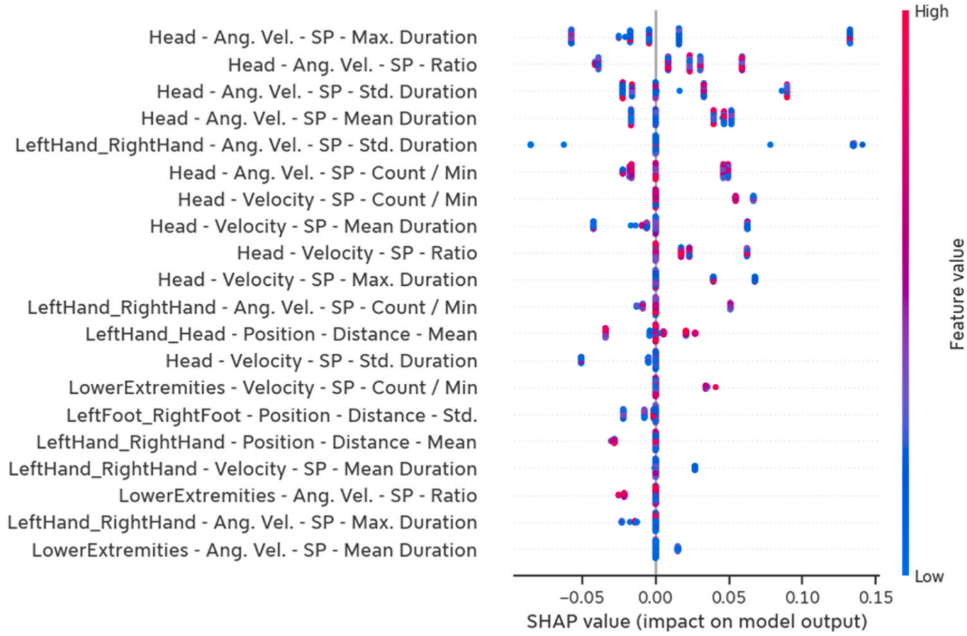
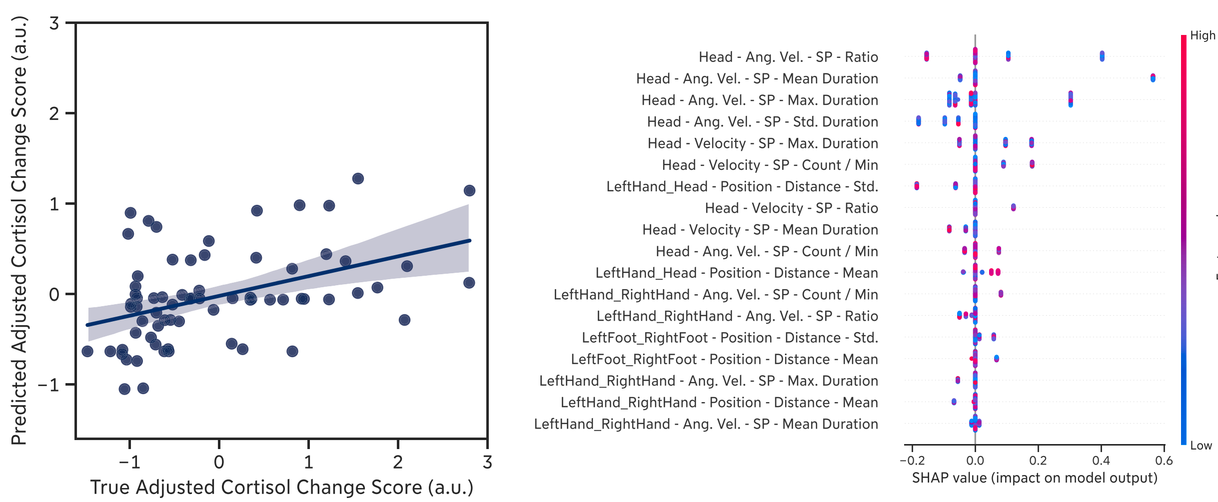


Fig. 4. SHAP values showing the 20 most important features for classifying cortisol responders versus non-responders. Features names are defined as in the following example: "Head - Ang. Vel. - SP - Max. Duration" represents the maximum duration of a static period (SP) in the angular velocity (Ang. Vel.) of the head. I.e., the longest continuous time interval during which the head's angular velocity remained very low, indicating a prolonged moment of little to no rotational movement.



(a) (b)

Fig. 5. (a) Predicted and true standardized change scores. (b) SHAP values for the selected features in the regression task. Feature names can be read as follows: "Head - Ang. Vel. - SP - Ratio" represents the ratio of total static periods (SP), during which the head's angular velocity (Ang. Vel.) is very low, to the overall measurement duration, reflecting the proportion of time spent with minimal rotational head movement.

values are underestimated. This pattern may reflect a regression-to-the-mean effect, potentially arising from an imbalanced distribution of cortisol responses in the training data. Such bias might indicate that the model struggles to capture extreme stress responses accurately, possibly due to individual variability in HPA axis activation.

3.3. Limitations

Several limitations should be considered when interpreting the results of the current study. First, the relatively small sample size restricts the generalizability of the findings. However, a key strength of the study design lies in the inclusion of a control condition and the use of a within-subject design, which effectively doubled the number of data points. However, it also increases the complexity of the prediction task compared to previous studies that focused solely on predicting cortisol responses to a stress task (e.g. Baird et al., 2019). This enhances the generalizability of the results, especially because the f-TSST protocol is less strict in its standardization compared to the TSST. Despite these strengths, future research involving larger and more diverse populations is needed to assess the robustness of the observed relationships between movement and cortisol.

Importantly, the generalizability of the findings is further limited by the homogeneity of the sample, which consisted exclusively of young, healthy, German-speaking university students. While this approach reduces confounding variables and increases internal validity, it restricts applicability to broader populations with more diverse sociodemographic and cultural backgrounds. Future studies should include participants across different age groups, health statuses, and cultural

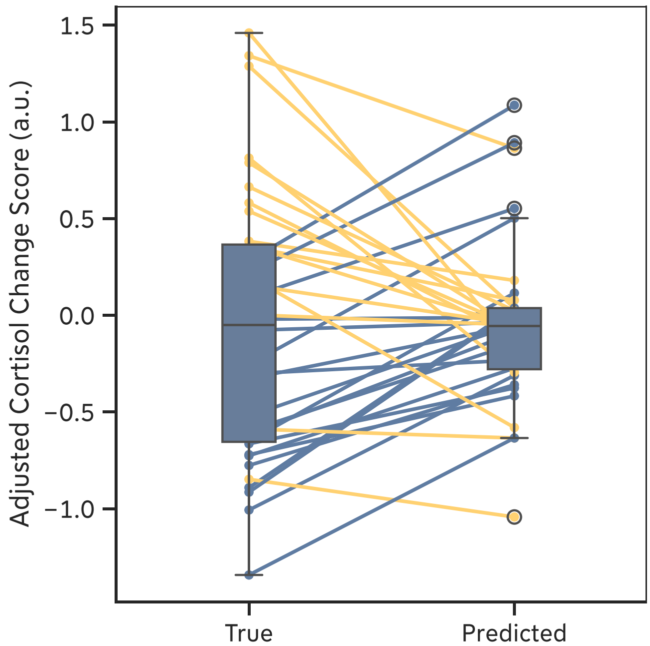


Fig. 6. Paired boxplot for true and predicted adjusted cortisol change scores (yellow lines indicate a smaller predicted value, and blue lines indicate a larger predicted value compared to the true values).

contexts to better understand how individual and contextual factors shape the relationship between movement and cortisol responses.

The complexity of the physiological stress response presents a notable limitation, as cortisol secretion is modulated by various factors, including genetic predisposition, hormonal fluctuations, environmental stressors, and age (Russell and Lightman, 2019). Additionally, the cortisol response to repeated acute stressors is known to be subject to habituation effects (Petrowski et al., 2012). Although our protocol did not include repeated stress exposure per se, we showed that some participants had a substantial cortisol response to the friendly control condition, which potentially led to the aforementioned habituation effects. If the body posture and movement changes show similar habituation effects should be a topic for future investigations. The models employed in this study did not account for these factors, instead focusing exclusively on posture and movement data. This limitation may have contributed to individual differences in observed cortisol levels and movement patterns, which we were unable to fully predict. Future research should consider incorporating these factors into predictive models to enhance their accuracy and reliability.

We used inertial sensor-based motion capture technology to measure movement patterns, which brings several advantages. These systems are portable and cost-effective, operating without external cameras or line-of-sight constraints, though they come with certain challenges. Technical issues, such as data corruption or sensor malfunctions, led to the exclusion of some participants, thus reducing the amount of usable data. Additionally, the attachment of the sensors might interfere with the natural movement of the participants (Jeon et al., 2023). To address this limitation, future research should explore contactless motion capture using camera-based or radar technology. Camera-based tracking methods, in particular, show significant potential, as recent advancements have introduced numerous pose estimation frameworks with promising accuracy and robustness. Furthermore, the TSST protocol already incorporates camera recordings, providing an opportunity to develop a large and diverse dataset for motion analysis.

Furthermore, our evaluation was based exclusively on features aggregated over the entire (f-)TSST duration. While this approach enhances simplicity and interpretability, it inevitably omits part of the information present in the raw movement data. Future research might leverage temporal data, examining how the body posture and movements change throughout the stress task, potentially improving the prediction of the neuroendocrine response.

4. Conclusion & outlook

We investigated in this study the relationship between body posture, movement patterns, and cortisol reactivity following the (f-)TSST. Using machine learning techniques, we classified cortisol responders and non-responders with a moderate accuracy of 65.2 % and predicted the maximum cortisol increase with a mean absolute error of 2.95 nmol/l. Additionally, we achieved a correlation of.45 for the prediction of adjusted change scores. These findings reinforce the idea that movement patterns encode valuable information about physiological stress responses, highlighting the potential of body posture and motion as non-invasive markers of stress reactivity.

Despite the mentioned limitations, body posture and movements provide valuable insights into the stress response. To the best of our knowledge, we are the first to present full body motion as a potential digital biomarker, complementing traditional measures. By integrating multimodal approaches, including heart rate variability, speech analysis, and facial expressions, the predictive power of stress assessment models could be enhanced. Especially speech analysis is a promising candidate, as we were able to show in previous work that the classification of (f-)TSST conditions from speech features was possible with 80 % accuracy (Oesten et al., 2023).

Given the temporal alignment of movement changes with the stressor, future work should also investigate associations with autonomic indices such as heart rate, skin conductance, or vagally mediated heart rate variability. These markers, reflecting sympathetic and parasympathetic activity, may be more closely related to arousal and motor inhibition than cortisol and could yield improved prediction accuracy. Such integration could offer a more comprehensive view of the physiological correlates of stress-related movement.

Additionally, future work should also examine the extent to which the most informative movement features identified by our models can be perceived and interpreted by human observers. Preliminary findings from an ongoing study, in which human raters view motion capture recordings to classify stress conditions, will help clarify this. Notably, some features, such as reduced hand movement or a downward-tilted head during the TSST, were based on prior observational insights, suggesting that at least a subset of these patterns may be accessible to trained observers. However, more subtle or high-dimensional features likely require computational approaches to be reliably detected.

Beyond stress detection, body movement patterns may also reflect broader physiological processes. For example, acute sickness is associated with a low-grade inflammatory response, which can be detected through alterations in movement patterns (Lasselin et al., 2020). Similarly, acute psychosocial stress elicits an inflammatory response, representing a potential pathway linking stress to disease progression (Rohleder, 2019). Given this association, further research is needed to systematically investigate the relationship between body movements and inflammation to enhance our understanding of stress-related health outcomes.

Future research should also explore real-world applications of movement-based stress detection. Wearable technologies, such as accelerometers and smart clothing, offer the potential to continuously monitor body movements in naturalistic settings. Implementing such technologies might pave the way for real-time stress assessment in workplace environments, healthcare settings, and everyday life, facilitating early stress detection and personalized intervention strategies.

In conclusion, our findings contribute to the growing field of computational behavioral science by demonstrating the feasibility of using body movements as a non-invasive stress marker. By refining these approaches and integrating additional physiological and behavioral indicators, the development of objective and scalable stress assessment tools might be possible, which ultimately improve stress management and mental health interventions.

CRediT authorship contribution statement

Robert Richer: Writing – review & editing, Methodology, Investigation, Conceptualization. **Luca Abel:** Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Miriam Kurz:** Writing – review & editing, Conceptualization. **Felicitas Burkhardt:** Writing – review & editing, Conceptualization. **Veronika Ringgold:** Writing – review & editing, Conceptualization. **Eskofier Bjoern M:** Writing – review & editing, Supervision, Funding acquisition. **Lena Schindler-Gmelch:** Writing – review & editing, Conceptualization. **Nicolas Rohleder:** Writing – review & editing, Supervision, Funding acquisition.

Funding

This work was partly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1483 – Project-ID 442419336, EmpkinS.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT (GPT-4o) (OpenAI, San Francisco, CA, USA) in order to enhance readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of Competing Interest

The authors declare no conflicts of interest.

Acknowledgments

We would like to thank R. Buberl, J. Gutheil, S. Heumann, F. Hirschbeck, D. Iajgunovici, J. Kaup, and N. Merz for their efforts and support in data collection as well as all study participants for their participation.

Preprint

This work has been published as preprint on OSF (https://doi.org/10.31219/osf.io/kwdmf_v1; 07.04.2025).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.psyneuen.2025.107528.

References

Atkinson, A.P., Dittrich, W.H., Gemmell, A.J., Young, A.W., 2004. Emotion perception from dynamic and static body expressions in point-light and full-light displays. Perception 33, 717–746. https://doi.org/10.1068/p5096.

Axelsson, J., Sundelin, T., Olsson, M.J., Sorjonen, K., Axelsson, C., Lasselin, J., Lekander, M., 2018. Identification of acutely sick people and facial cues of sickness. Proc. R. Soc. B Biol. Sci. 285, 20172430. https://doi.org/10.1098/rspb.2017.2430.

Baird, A., Amiriparian, S., Cummins, N., Sturmbauer, S., Janson, J., Messner, E.-M., Baumeister, H., Rohleder, N., Schuller, B.W., 2019. Using Speech to Predict Sequentially Measured Cortisol Levels During a Trier Social Stress Test. Interspeech 2019. Presented at the Interspeech 2019. ISCA, pp. 534–538. https://doi.org/10.21437/Interspeech.2019-1352.

Bellman, R., 1966. Dynamic programming. Science 153, 34–37. https://doi.org/10.1126/science.153.3731.34.

Benjamin, L.S., 1963. Statistical Treatment of the Law of Initial Values (LIV) in Autonomic Research: A Review and Recommendation. Biopsychosoc. Sci. Med. 25, 556.

Buss, K.A., Davidson, R.J., Kalin, N.H., Goldsmith, H.H., 2004. Context-specific freezing and associated physiological reactivity as a dysregulated fear response. Dev. Psychol. 40, 583–594. https://doi.org/10.1037/0012-1649.40.4.583.

Crumpei-Tanasă, I., Crumpei, I., 2021. A machine learning approach to predict stress hormones and inflammatory markers using illness perception and quality of life in breast cancer patients. Curr. Oncol. 28, 3150–3171. https://doi.org/10.3390/curroncol28040275.

Dael, N., Mortillaro, M., Scherer, K.R., 2012. Emotion expression in body action and posture. Emotion 12, 1085–1101. https://doi.org/10.1037/a0025737.

Dedovic, K., Renwick, R., Mahani, N.K., Engert, V., Lupien, S.J., Pruessner, J.C., 2005. The Montreal Imaging Stress Task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain. J. Psychiatry Neurosci. 30, 319–325.

Dickerson, S.S., Kemeny, M.E., 2004. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. Psychol. Bull. 130, 355–391. https://doi.org/10.1037/0033-2909.130.3.355.

Dickerson, S.S., Gruenewald, T.L., Kemeny, M.E., 2004. When the social self is threatened: shame, physiology, and health. J. Pers. 72, 1191–1216. https://doi.org/10.1111/j.1467-6494.2004.00295.x.

Dong, G., Boukhechba, M., Shaffer, K.M., Ritterband, L.M., Gioeli, D.G., Reilley, M.J., Le, T.M., Kunk, P.R., Bauer, T.W., Chow, P.I., 2021. Using graph representation learning to predict salivary cortisol levels in pancreatic cancer patients. J. Healthc. Inform. Res. 5, 401–419. https://doi.org/10.1007/s41666-021-00098-4.

Doumas, M., Morsanyi, K., Young, W.R., 2018. Cognitively and socially induced stress affects postural control. Exp. Brain Res. 236, 305–314. https://doi.org/10.1007/s00221-017-5128-8.

Hagenaars, M.A., Roelofs, K., Stins, J.F., 2014. Human freezing in response to affective films. Anxiety Stress Coping 27, 27–37. https://doi.org/10.1080/10615806.2013.809420.

Hansson, L.S., Lasselin, J., Tognetti, A., Axelsson, J., Olsson, M.J., Sundelin, T., Lekander, M., 2023. The walking sick: perception of experimental sickness from biological motion. Brain Behav. Immun. 113, 319–327. https://doi.org/10.1016/j.bbi.2023.07.020.

Hashemi, M.M., Zhang, W., Kaldewaij, R., Koch, S.B.J., Smit, A., Figner, B., Jonker, R., Klumpers, F., Roelofs, K., 2021. Human defensive freezing: associations with hair cortisol and trait anxiety. Psychoneuroendocrinology 133, 105417. https://doi.org/10.1016/j.psyneuen.2021.105417.

Het, S., Rohleder, N., Schoofs, D., Kirschbaum, C., Wolf, O.T., 2009. Neuroendocrine and psychometric evaluation of a placebo version of the 'Trier Social Stress Test'. Psychoneuroendocrinology 34, 1075–1086. https://doi.org/10.1016/j.psyneuen.2009.02.008.

Janson, J., Rohleder, N., 2017. Distraction coping predicts better cortisol recovery after acute psychosocial stress. Biol. Psychol. 128, 117–124. https://doi.org/10.1016/j.biopsycho.2017.07.014.

Jeon, J., Kwon, S., Lee, Y., Hong, J., Yu, J., Kim, J., Kim, S.-G., Lee, D., 2023. Influence of the Hawthorne effect on spatiotemporal parameters, kinematics, ground reaction force, and the symmetry of the dominant and nondominant lower limbs during gait. J. Biomech. 152, 111555. https://doi.org/10.1016/j.jbiomech.2023.111555.

Jin, P., 1992. Toward a reconceptualization of the law of initial value. Psychol. Bull. 111, 176–184. https://doi.org/10.1037/0033-2909.111.1.176.

Kirschbaum, C., Pirke, K.-M., Hellhammer, D.H., 1993. The "Trier Social Stress Test" – a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology 76–81. https://doi.org/10.1159/000119004.

Lasselin, J., Sundelin, T., Wayne, P.M., Olsson, M.J., Paues Göranson, S., Axelsson, J., Lekander, M., 2020. Biological motion during inflammation in humans. Brain Behav. Immun. 84, 147–153. https://doi.org/10.1016/j.bbi.2019.11.019.

Lundberg, S., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions. https://doi.org/10.48550/arXiv.1705.07874.

Mantella, R.C., Butters, M.A., Amico, J.A., Mazumdar, S., Rollman, B.L., Begley, A.E., Reynolds, C.F., Lenze, E.J., 2008. Salivary cortisol is associated with diagnosis and severity of late-life generalized anxiety disorder. Psychoneuroendocrinology 33, 773–781. https://doi.org/10.1016/j.psyneuen.2008.03.002.

Miller, R., Plessow, F., Kirschbaum, C., Stalder, T., 2013. Classification criteria for distinguishing cortisol responders from nonresponders to psychosocial stress: evaluation of salivary cortisol pulse detection in panel designs. Psychosom. Med. 75, 832–840. https://doi.org/10.1097/PSY.0000000000000002.

Niermann, H.C.M., Figner, B., Tyborowska, A., van Peer, J.M., Cillessen, A.H.N., Roelofs, K., 2017. Defensive freezing links Hypothalamic-Pituitary-Adrenal-axis activity and internalizing symptoms in humans. Psychoneuroendocrinology 82, 83–90. https://doi.org/10.1016/j.psyneuen.2017.05.001.

Noordewier, M.K., Scheepers, D.T., Hilbert, L.P., 2020. Freezing in response to social threat: a replication. Psychol. Res. 84, 1890–1896. https://doi.org/10.1007/s00426-019-01203-4.

Oesten, M., Richer, R., Abel, L., Rohleder, N., Eskofier, B.M., 2023. VoStress–Voice-based Detection of Acute Psychosocial Stress. 2023 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI). IEEE, pp. 1–4.

Ogren, M., Kaplan, B., Peng, Y., Johnson, K.L., Johnson, S.P., 2019. Motion or emotion: Infants discriminate emotional biological motion based on low-level visual information. Infant Behav. Dev. 57, 101324. https://doi.org/10.1016/j.infbeh.2019.04.006.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Petrowski, K., Wintermann, G.-B., Siepmann, M., 2012. Cortisol response to repeated psychosocial stress. Appl. Psychophysiol. Biofeedback 37, 103–107. https://doi.org/10.1007/s10484-012-9183-4.

Ramsay, D., Lewis, M., 2003. Reactivity and regulation in cortisol and behavioral responses to stress. Child Dev. 74, 456–464. https://doi.org/10.1111/1467-8624.7402009.

Richer, R., Küderle, A., Ullrich, M., Rohleder, N., Eskofier, B.M., 2021. BioPsyKit: a Python package for the analysis of biopsychological data. J. Open Source Softw. 6, 3702. https://doi.org/10.21105/joss.03702.

Richer, R., Koch, V., Abel, L., Hauck, F., Kurz, M., Ringgold, V., Müller, V., Küderle, A., Schindler-Gmelch, L., Eskofier, B.M., Rohleder, N., 2024b. Machine learning-based detection of acute psychosocial stress from body posture and movements. Sci. Rep. 14, 8251. https://doi.org/10.1038/s41598-024-59043-1.

[dataset] Richer, R., Abel, L., Hauck, F., Kurz, M., Ringgold, V., Küderle, A., Schindler-Gmelch, L.,Eskofier, B., Rohleder, N., 2024a. Effect of acute psychosocial stress on body posture and movements – Main Study Dataset. https://doi.org/10.17605/OSF.IO/VA6T3.

Ringgold, V., Abel, L., Eskofier, B.M., Rohleder, N., 2024. Validation of the Virtual Reality Stroop Room: effects of inhibiting interfering information under time-pressure and task-switching demands. Comput. Hum. Behav. Rep. 16, 100497. https://doi.org/10.1016/j.chbr.2024.100497.

Roelofs, K., Hagenaars, M.A., Stins, J., 2010. Facing freeze: social threat induces bodily freeze in humans. Psychol. Sci. 21, 1575–1581. https://doi.org/10.1177/0956797610384746.

Rohleder, N., 2019. Stress and inflammation - the need to address the gap in the transition between acute and chronic stress effects. Psychoneuroendocrinology 105, 164–171. https://doi.org/10.1016/j.psyneuen.2019.02.021.

Russell, G., Lightman, S., 2019. The human stress response. Nat. Rev. Endocrinol. 15, 525–534. https://doi.org/10.1038/s41574-019-0228-0.

Smyth, J.M., Ockenfels, M.C., Gorin, A.A., Catley, D., Porter, L.S., Kirschbaum, C., Hellhammer, D.H., Stone, A.A., 1997. Individual differences in the diurnal cycle of cortisol. Psychoneuroendocrinology 22, 89–105. https://doi.org/10.1016/S0306-4530(96)00039-X.

Tracy, J.L., Matsumoto, D., 2008. The spontaneous expression of pride and shame: evidence for biologically innate nonverbal displays. Proc. Natl. Acad. Sci. U. S. A. 105, 11655–11660. https://doi.org/10.1073/pnas.0802686105.

Turan, B., 2015. Is a submissive posture adaptive when being evaluated negatively? Effects on cortisol reactivity. Neuroendocrinol. Lett. 36, 394–398.

Vallat, R., 2018. Pingouin: statistics in Python. J. Open Source Softw. 3, 1026. https://doi.org/10.21105/joss.01026.

Wallbott, H.G., 1998. Bodily expression of emotion. Eur. J. Soc. Psychol. 28, 879–896. https://doi.org/10.1002/(SICI)1099-0992(1998110)28:6<879::AID-EJSP901>3.0.CO;2-W.

Wiemers, U.S., Schoofs, D., Wolf, O.T., 2013. A friendly version of the Trier Social Stress Test does not activate the HPA axis in healthy men and women. Stress 16, 254–260. https://doi.org/10.3109/10253890.2012.714427.

Wilder, J., 1962. Basimetric Approach (law of Initial Value) to Biological Rhythms. Ann. N. Y. Acad. Sci. 98, 1211–1220. https://doi.org/10.1111/j.1749-6632.1962.tb30629.x.

Wolpert, D.H., Macready, W.G., 1997. No free lunch theorems for optimization. IEEE Trans. Evol. Comput. 1, 67–82. https://doi.org/10.1109/4235.585893.

Yarkoni, T., Westfall, J., 2017. Choosing prediction over explanation in psychology: lessons from machine learning. Perspect. Psychol. Sci. 12, 1100–1122. https://doi.org/10.1177/1745691617693393.

Zito, G.A., Apazoglou, K., Paraschiv-Ionescu, A., Aminian, K., Aybek, S., 2019. Abnormal postural behavior in patients with functional movement disorders during exposure to stress. Psychoneuroendocrinology 101, 232–239. https://doi.org/10.1016/j.psyneuen.2018.11.020.