Classification of Acute Stress-Induced Response Patterns

Luca Abel
Machine Learning and Data Analytics Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
Erlangen, Germany
luca.abel@fau.de

Robert Richer
Machine Learning and Data Analytics Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
Erlangen, Germany
robert.richer@fau.de

Arne Küderle
Machine Learning and Data Analytics Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
Erlangen, Germany
arne.kuederle@fau.de

Stefan Gradl
Machine Learning and Data Analytics Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
Erlangen, Germany
stefan.gradl@fau.de

Bjoern M. Eskofier
Machine Learning and Data Analytics Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
Erlangen, Germany
bjoern.eskofier@fau.de

Nicolas Rohleder*
Chair of Health Psychology, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
Erlangen, Germany
nicolas.rohleder@fau.de

ABSTRACT

Modern machine learning techniques enable new possibilities for the analysis of psychological data. In the field of health psychology, it is of interest to explore the biological processes triggered by acute stress. This work introduces a method to automatically classify individuals into distinct stress responder groups based on these biological processes. Two important stress-sensitive markers were used: Salivary cortisol and Interleukin-6 (IL-6) in blood plasma. Controlled stress was induced using the Trier Social Stress Test on two consecutive days. Results show that Support Vector Machines performed best on the given dataset. We distinguished four different cortisol and three different IL-6 responder types with high mean accuracies (92.2 % ± 9.7 % and 91.2 % ± 6.3 %, respectively). Classification results were mainly limited by class imbalances and high intra-class standard deviations. Whereas promising as a first application of machine learning on such datasets, generalizability and real-world applicability of our results need to be proven by further research.

CCS CONCEPTS

• **Applied computing** → **Health informatics**; Consumer health;
• **Computing methodologies** → *Supervised learning*.

KEYWORDS

machine learning, health psychology, acute stress, psychosocial stress, TSST, cortisol, IL-6

*Also with Department of Psychology and Volen National Center for Complex Systems, Brandeis University.

PervasiveHealth '19, May 20–23, 2019, Trento, Italy

ACM ISBN 978-1-4503-9999-9/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

ACM Reference Format:
Luca Abel, Robert Richer, Arne Küderle, Stefan Gradl, Bjoern M. Eskofier, and Nicolas Rohleder. 2019. Classification of Acute Stress-Induced Response Patterns. In *PervasiveHealth '19: EAI International Conference on Pervasive Computing Technologies for Healthcare, May 20–23, 2019, Trento, Italy*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Stress is a hidden epidemic – the World Health Organization estimates that mental diseases, including stress-related disorders, will be the second leading cause of disabilities by the year 2020 [8]. The negative economic impact of stress is substantial, with an estimated loss of about $42 billion for absence and treatment of stress-related illness in the U.S. per year [8].

Stress can be differentiated into chronic and acute stress. As an example, social interactions with others can trigger an acute stress response. This response is characterized by strong biological reactions that affect the whole body through widely spread autonomic innervation and the secretion of stress hormones [14].

Past research has shown that these responses are highly individual, and can be categorized into healthy and multiple pathological groups [12, 14]. Whereas adequate stress responses are a crucial and healthy physiological reaction, defective stress responses have been linked to DNA damage, over-expression of inflammatory genes, and declines in cognitive functioning, which are well known markers of physiological and biological age [11].

Successful management of such stress-related diseases require early detection, as well as precision medicine approaches. Therefore, there is a need to classify these response patterns by observing the reaction to acute stress. To reliably induce psycho-social stress in adults, a well established procedure is the Trier Social Stress Test (TSST) [10].

During such an acute stress situation, an important pathway through which the organism is sending signals is the hypothalamic-pituitary-adrenal axis (HPA axis). Under stress, the hypothalamus releases CRH (Corticotrophin Releasing Hormone), which stimulates the production of ACTH (Adrenocorticotrophic Hormone) in the anterior pituitary. In turn, increased levels of ACTH stimulate secretion of cortisol in the adrenal cortex [13]. Therefore, cortisol has been established as a non-invasive biomarker for

stress [9, 15]. Additionally, due to the inflammatory effect of stress, the pro-inflammatory cytokine Interleukin-6 (IL-6) can also serve as a marker of acute stress reaction [1]. Therefore, measurements of cortisol and IL-6 during TSST allow a reliable and repeatable observation of the different response patterns.

In reaction to acute stress an increase of these two biomarkers is expected. Cortisol levels peak about 15 min after stress, and then slowly decline to baseline after about 2 h. In contrast, IL-6 levels gradually increase with a peak after approx. 2 h.

Using the described study design, past works have already demonstrated that physiological and psychological measures are connected to certain stress response patterns. For instance, Breines et al. [1] showed that participants with higher self-compassion show lower IL-6 responses when performing the TSST whereas Chen et al. [3] discovered a relation between inflammatory reactivity and basal HPA axis activity as response to acute stress. Fiksdal et al. [4] suggest that symptoms of anxiety are linked to an attenuated cortisol stress response. In addition, they found exaggerated responses for subjects with depression symptoms. Furthermore, there is evidence that post-stress rumination predicts HPA axis responses to repeated acute stress [7].

Whereas these publications analyzed correlations between psychophysiological influences and stress responses, some publications suggest that the separation into discrete stress responder groups might bring further insight. For example, McEwen [11] formulated conditions for healthy and unhealthy responses. He manually separated the unhealthy patterns into four groups (missing response, exaggerated response, no return to baseline, and no habituation to repeated stress).

However, to the best of our knowledge, it was never attempted to classify these patterns of acute stress responses automatically. Therefore, the goal of this work is to use machine learning methods for classification. Compared to classification by a trained professional this approach has the potential to reduce the required time, as well as increase the objectivity of the grouping.

Our approach is similar to previous work performed in the context of Post-traumatic Stress Disorder (PTSD) by Galatzer-Levy et al. [5, 6]. They showed that machine learning techniques, applied to combined clinical, neuroendrocrine, psycho-physiological, and demographic data, were able to reliably predict a non-remitting PTSD diagnosis for the following 15 months based on data gathered within 10 days after the traumatic event.

These results have the potential to be the foundation of next generation precision therapies. For this reason, we aim to follow a similar approach for future individualized stress therapy by automatically classifying different responder groups.

2 METHODS

2.1 Study Data

Data was collected as part of a larger research project studying the effects of stress on aging over the course of two years and has been reported previously [1, 3, 4, 7].

71 healthy subjects (51 % female, age = 37.8 ± 18.4 years (M ± SD), BMI range = 18-35 kg/m^2, M = 25.2 kg/m^2) were included in our analysis. All participants were screened for medical and psychological conditions via telephone prior to being invited to participate.

All data was collected in the afternoon between 1:30pm and 6:30pm to minimize the impact of circadian variations in hormone concentrations. The specific procedures of the study were described in detail, and informed consent was obtained. The study procedure was approved by the Brandeis Institutional Review Board.

Subjects were exposed to the Trier Social Stress Test (TSST) on two consecutive days. The TSST consists of an anticipatory phase (about 10 min) and a test period (15 min), during which subjects have to perform free speech and mental arithmetic tasks in front of an audience of usually two adults [10]. The first phase is typically structured like a job interview. Additionally, video and audio recordings of the subject are taken in order to increase the induced stress and to provide records for further evaluation.

IL-6 was measured for all subjects, whereas cortisol measurements were only available for 63 subjects. The time points of sample collection (−45, −1, +16, +25, +45, +75, +135 min relative to TSST start) were chosen such that the complete cortisol response, including recovery, is expected to be covered. Due to the slow decline of IL-6 and the considerable effort of taking blood samples, only three data points were collected (−1, +45, +135 min relative to TSST start).

Table 1: Stress responder classes as derived from previous work [11].

Class	*Abbr.*	Description
Habituation	*Hab*	High day 1 response Low day 2 response
Non-Habituation	*NonHab*	High responses on both days
Sensitization	*Sens*	Low or none day 1 response Higher day 2 response
Non-Responder	*NonResp*	Low or none response on both days
Anticipatory	*Anticip*	Highest response *before* TSST on one of the days

2.2 Data Labeling

Based on the previous work of McEwen five class labels were derived (Table 1) [11]. Mean cortisol and IL-6 values of all subjects in the respective classes are shown in Figure 1. Class labels were assigned by a trained professional based on the measured hormone levels. The percentual distributions of the classes are shown in Figure 2. Due to the complex interplay between HPA axis and immune system, no direct dependency can be assumed. Therefore, cortisol and IL-6 were labeled and analyzed separately. These labels served as ground truth for our supervised learning approach on the same input data. In the IL-6 data no subjects with *NonResp* or *Anticip* patterns could be identified. Hence, the respective classes were omitted for this part of the analysis.

2.3 Feature Extraction & Selection

We computed several features based on the measured hormonal trends, which aim to mimic the labeling approach by a trained

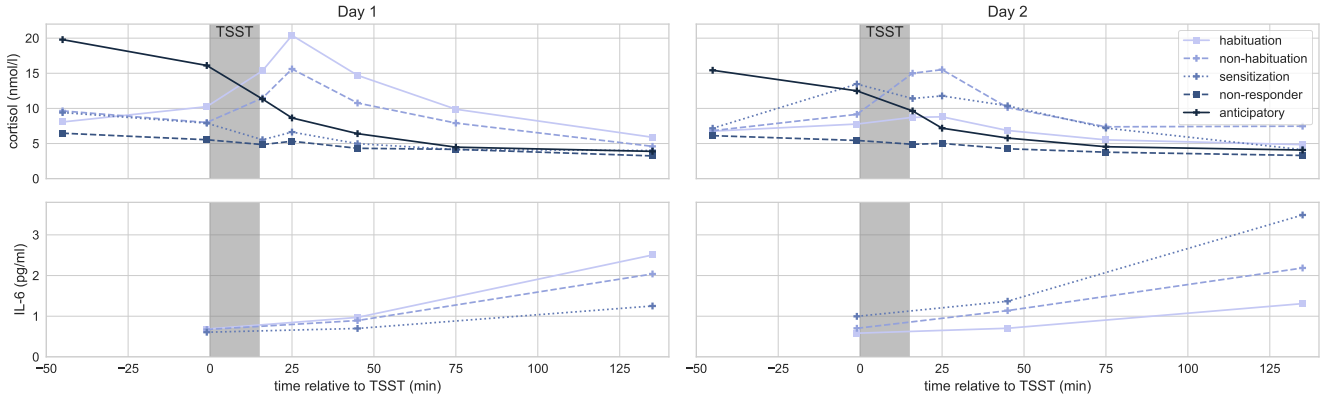

Figure 1: Mean cortisol and IL-6 levels for each class as labeled by an expert. Class definitions according to Table 1. Standard deviation not shown for better visualization.

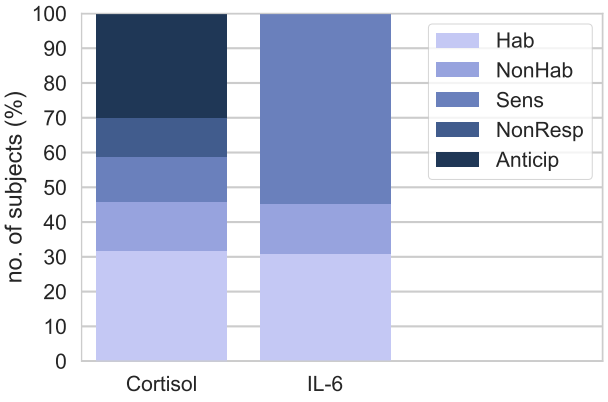

Figure 2: Class distributions for cortisol and IL-6 in %.

professional (e.g. Minimum-Maximum Difference, Time of Maximum). However, results of the feature selection (Select-K-Best with ANOVA F-value; $k \in [5, 20]$) revealed that the calculated features did not improve classification performance compared to the raw data values. Therefore, the normalized measurement values (14 for cortisol and 6 for IL-6) were directly used as features.

2.4 Classification

We tested three well established classifiers on the feature set: k-Nearest-Neighbors (kNN), Support Vector Machines (SVMs), and Random Forest (RF).

The selected classifiers were validated using a 5-fold stratified cross-validation. To counter class imbalances, we oversampled the dataset using the Synthetic Minority Oversampling Technique (*SMOTE*) with $k = 3$ [2].

As the performance of the used classifier heavily depends on the selected training parameters, a grid search was used in combination with a 5-fold inner cross-validation for kNN and SVM, whereas a randomized search was used for RF. The parameter space for kNN was $k \in [1, 20]$. For the SVM we tested Linear- or Radial-Basis Function kernels and values for C in range of 10^0 and 10^4 (logarithmic steps), and γ in range of 10^{-2} and 10^{-3}. RF parameters were optimized within a standard parameter space. The parameter optimization was performed for each fold of the external cross validation.

We selected the best classifier based on the mean accuracy over all cross-validation folds. To obtain a final optimal set of classifier parameters, the selected classifier was re-trained on the entire dataset.

3 RESULTS

Linear SVMs performed best for both biological variables with an accuracy of 78.1 % ± 11.2 % for five cortisol classes and 91.2 % ± 6.3 % for three IL-6 classes. Table 2 shows the results for all classifiers. Individual cortisol and IL-6 classes are shown in the confusion matrices (Figure 3). The optimal values for C, based on the parameter optimization, were $C = 10^3$ for cortisol and $C = 10^4$ for IL-6.

The cortisol results revealed that the majority of misclassification occurred between the *Sens* and *NonResp* group. Merging these two groups improved accuracy to 92.2 % ± 9.7 %. This will be further discussed in later sections of this paper.

Table 2: Accuracies of different classifiers in % (M ± SD).

	kNN	SVM	RF
Cortisol	64.6 ± 12.2	78.1 ± 11.2	66.2 ± 11.0
IL-6	84.3 ± 9.2	91.2 ± 6.3	77.4 ± 8.9

4 DISCUSSION

Our results indicate that an automatic classification of stress response groups based on cortisol and IL-6 levels is possible and allows a more objective approach than manual labeling.

As noted in the results, *Sens* samples are often misclassified as *NonResp* (and vice versa). It became apparent that the biological difference between both classes is not clear enough for reliable

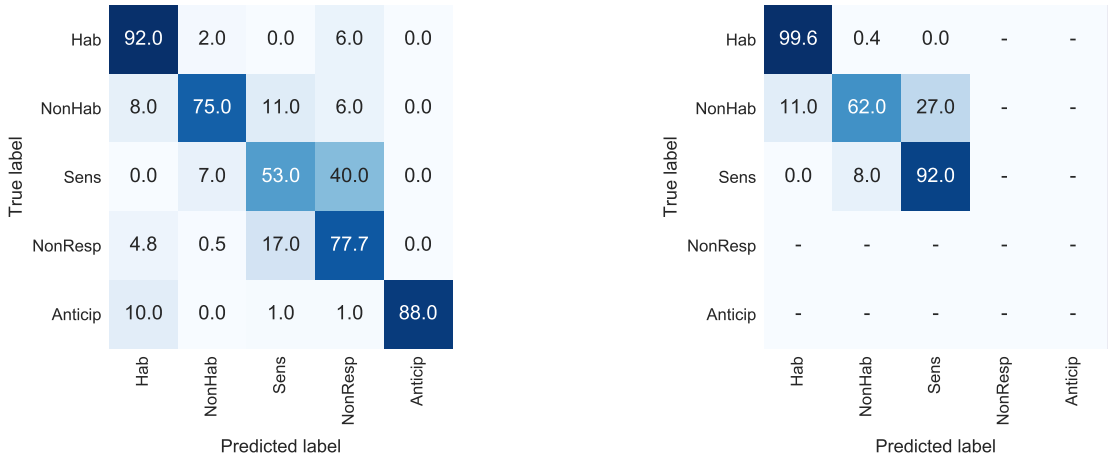

Figure 3: Confusion matrices for cortisol (left) and IL-6 (right) response classification (Linear SVM). Normalized on true labels (%).

labeling. Specifically, most of *Sens* subjects only show a marginally higher reaction than *NonResp* subjects on the second day. After consultation with a health psychology expert, we merged these two classes, which considerably improved classification accuracy from 78.1 % to 92.2 %.

This observation indicates that a clearer consent on the class definitions might be required. Furthermore, this issue highlights the well-known paradigm that the classifier can only perform as good as the provided training labels. In further research this could be counteracted by multiple experts independently labeling the same dataset.

IL-6 pattern classification showed a high accuracy for all classes. This is also supported by inspection of the raw data, which revealed smaller intra-class variability compared to cortisol. Furthermore, results suggest that three samples per day are enough to reliably distinguish the selected three groups. Further samples taken before the TSST might be required to investigate the existence of additional IL-6 patterns. However, such changes to the study design would require more blood samples, which leads to higher costs and induces additional, unintended stress to the subjects.

From the three different patterns the classification of *NonHab* performed worst. It is the minority class with only 10 of 71 subjects (approx. 14 %). This could indicate that the applied oversampling could not fully compensate the effects of class imbalances.

Whereas the presented results look promising, there are biological and technical aspects which might limit the reliability of a TSST-based stress responder classification. For instance, IL-6 is primarily an inflammation marker and only indirectly linked to stress. Therefore, the IL-6 response to acute stress could be superimposed by other inflammation-related processes. Furthermore, the measured cortisol values highly depend on the used analysis device and method. Changing the analysis pipeline could lead to baseline shifts or changes in amplitude scaling. Therefore, the classifier trained in this work would potentially need to be re-trained, using additional data, in order to increase generalizability of the underlying classification model.

5 CONCLUSION & OUTLOOK

We introduced a method capable of classifying four different cortisol and three different IL-6 responder types with high accuracy (92.2 %± 9.7 % and 91.2 % ± 6.3 %, respectively). Additionally, we hope to motivate further research in this interdisciplinary challenge of applying machine learning techniques to health psychology data. Our results might lay the foundation for the automatic detection of unhealthy stress patterns. This could be implemented in the context of a personalized medicine approach, where patients undergo the two day TSST procedure to predict stress-related diseases based on automatic stress response classification.

Currently, we are collecting additional data with a similar test protocol. This data will be used for the demonstrate of generalizability of our model and to further improve the accuracy of our classification.

On a larger scale, we are looking for alternatives to the presented TSST based study design. Due to the high effort required to perform the test and collect the biological samples, it is infeasible for large scale screening of stress responder types. For example, as an alternative method for stress induction Virtual Reality based approaches could be considered, whereas biological samples could be replaced by questionnaires or video based observations. Furthermore, classification of stress responder types based on easier to measure parameters should be considered. Besides demographic data, especially psychological measures are prime candidates, as they have shown to correlate with stress reaction patterns [1].

ACKNOWLEDGMENTS

Bjoern M. Eskofier gratefully acknowledges the support of the German Research Foundation (DFG) within the framework of the Heisenberg professorship programme (grant number ES 434/8-1).

We would like to thank the research assistants in the Laboratory for Biological Health Psychology at Brandeis University who assisted with data collection, as well as all subjects who participated in the study.

REFERENCES

[1] Juliana G. Breines, Myriam V. Thoma, Danielle Gianferante, Luke Hanlin, Xuejie Chen, and Nicolas Rohleder. 2014. Self-compassion as a predictor of interleukin-6 response to acute psychosocial stress. *Brain, Behavior, and Immunity* 37 (2014), 109–114.

[2] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philipp Kegelmeyer. 2002. SMOTE: Synthetic Minority Over-sampling Technique. *J Artif Intell Res* 16 (2002), 321–357.

[3] Xuejie Chen, Danielle Gianferante, Luke Hanlin, Alexander Fiksdal, Juliana G. Breines, Myriam V. Thoma, and Nicolas Rohleder. 2017. HPA-axis and inflammatory reactivity to acute stress is related with basal HPA-axis activity. *Psychoneuroendocrinology* 78 (2017), 168–176. https://doi.org/10.1016/j.psyneuen.2017.01.035

[4] Alexander Fiksdal, Luke Hanlin, Yuliya Kuras, Danielle Gianferante, Xuejie Chen, Myriam V. Thoma, and Nicolas Rohleder. 2019. Associations between symptoms of depression and anxiety and cortisol responses to and recovery from acute stress. *Psychoneuroendocrinology* 102 (apr 2019), 44–52.

[5] Isaac R. Galatzer-Levy, Karen-Inge Karstoft, Alexander Statnikov, and Arieh Y. Shalev. 2014. Quantitative forecasting of PTSD from early trauma responses: A Machine Learning application. *J Psychiatr Res* 59 (2014), 68–76.

[6] Isaac R. Galatzer-Levy, Sisi Ma, Alexander Statnikov, Rachel Yehuda, and Arieh Y. Shalev. 2017. Utilization of machine learning for prediction of post-traumatic stress: A re-examination of cortisol in the prediction and pathways to non-remitting PTSD. *Transl Psychiatry* 7, 3 (2017), 1–8.

[7] Danielle Gianferante, Myriam V. Thoma, Luke Hanlin, Xuejie Chen, Juliana G. Breines, Peggy M. Zoccola, and Nicolas Rohleder. 2014. Post-stress rumination predicts HPA axis responses to repeated acute stress. *Psychoneuroendocrinology* 49, 1 (2014), 244–252.

[8] Madhu Kalia. 2002. Assessing the economic impact of stress-the modern day hidden epidemic. *Metabolism: clinical and experimental* 51, 6 (2002), 49–53.

[9] Clemens Kirschbaum and Dirk H. Hellhammer. 1989. Salivary cortisol in psychobiological research: an overview. *Neuropsychobiology* 22, 3 (1989), 150–69.

[10] Clemens Kirschbaum, Karl-Martin Pirke, and Dirk H. Hellhammer. 1993. The 'Trier social stress test' - A tool for investigating psychobiological stress responses in a laboratory setting. In *Neuropsychobiology*.

[11] Bruce S. McEwen. 1998. Protective and Damaging Effects of Stress Mediators. *N Engl J Med* 338, 3 (1998), 171–179.

[12] Christine M. McInnis, Myriam V. Thoma, Danielle Gianferante, Luke Hanlin, Xuejie Chen, Juliana G. Breines, Suzi Hong, and Nicolas Rohleder. 2014. Measures of adiposity predict interleukin-6 responses to repeated psychosocial stress. *Brain, Behavior, and Immunity* 42 (2014), 33–40.

[13] Sheina Orbell, Havah Schneider, and Sabrina Esbitt. 2013. Hypothalamic-Pituitary-Adrenal Axis. In *Encyclopedia of Behavioral Medicine*. Springer New York, New York, NY, 1017–1018.

[14] Robert M. Sapolsky, L. Michael Romero, and Allan U. Munck. 2000. How Do Glucocorticoids Influence Stress Responses? Integrating Permissive, Suppressive, Stimulatory, and Preparative Actions. *Endocr. Rev.* 21, 1 (2000), 55–89.

[15] Nina Smyth, Frank Hucklebridge, Lisa Thorn, Phil Evans, and Angela Clow. 2013. Salivary cortisol as a biomarker in social science research. *Soc Personal Psychol Compass* 7, 9 (2013), 605–625.