# ISPAAD: Integrated Stress, Physical Activity, and Amusement Dataset

*Paul* Schreiber[1,*] and *Maria* Maleshkova[1,**]

[1]Professorship of Data Engineering, Helmut Schmidt University, Hamburg, Germany

**Abstract.** Long-term stress exposure is a leading cause of sickness in adults and may even cause physiological issues such as heart diseases. Developing systems to detect and monitor stressful events based on physiological signals can help prevent stress-induced severe health implications. Research in this domain has shown that it is possible to reliably distinguish stress and its absence and differentiate between different types of stress. However, publicly available data in the stress detection domain is rare, which limits the research possibilities. Moreover, datasets are usually very heterogeneous owing to variations in experimental protocols and sensors used to record physiological signals, which limits the comparability. This work presents ISPAAD, an integrated stress, physical activity, and amusement dataset to overcome aforementioned limitations. The dataset is intended for stress detection, in-depth analysis of different stressors, and physical and affective states. The contributions of this paper are threefold: 1) the criteria to be considered for seamless data integration are outlined, 2) a vanilla data integration pipeline is introduced, and 3) a novel and multi-modal dataset is presented, bridging cognitive and socio-evaluative stress with physical activity as well as a positive affective state.

## 1 Introduction

Distress is linked to adverse effects on mental health, and may worsen symptoms of those suffering from migraines and stress disorders [1]. Especially the long-term exposure to stress may be detrimental to health and even cause physiological issues such as heart diseases [2]. Conventional stress assessments are conducted using psychological self-reports, such as the Perceived Stress Scale (PSS) [3], the State-Trait-Anxiety-Inventory (STAI) [4], or the Self-Assessment Manikins (SAM) [5], whereas the latter is also used to assess other types of affects. Albeit, such stress-assessments are limited in terms of practicability in everyday life, they are prone to erroneous subjectiveness, and they reflect on fixed moments and situations in time rather than being continuous evaluations. To address these issues, researchers have employed non-invasive wearable devices for continuous monitoring of physiological signals. In previous research, physiological signals including cardiac signals, skin conductance, respiration rate, or temperature have been studied as indicators for increased stress, to be used to develop automated stress recognition systems (e.g., [1]. These studies achieve promising results in predicting or classifying the exposure to stress stimuli. However, the data used to

[*]e-mail: schreibp@hsu-hh.de
[**]e-mail: maleshkm@hsu-hh.de

train such systems is usually obtained from small and homogeneous samples. Data recorded during studies is oftentimes not publicly available for other researchers, and due to limited standardization in study protocols, different datasets are highly heterogeneous. Hence, the domain lacks validation of study results and testing whether the systems employed generalize well to other samples.

The lack of validated and generalizable stress management systems can be attributed to number of reasons. First of all, publicly available datasets in the domain of stress detection are rare, as pointed out by [6]. Secondly, data collection and experimental setups in stress research and affective computing still lack standardized procedures that facilitate structural unity of datasets [7]. However, a standardized protocol to infer human emotions [8], and other affective states such as stress, would vastly benefit the development of reproducible and generalizable recognition systems. Furthermore, the datasets used to develop automated stress recognition systems often differ in sensor technologies, stress elicitation methods, and vital signs recorded, restricting the comparability of results (compare [9–11]). Lastly, and related to the previous issues, is the lack of proper annotation of the datasets. Datasets for stress recognition are often complex and characterized by their temporal, continuous, and multi-modal structure. Adequate labeling and annotation of such datasets is indispensable, specifically to facilitate the validation of research results by other researchers and to integrate data with other datasets.

This paper introduces a combined multi-modal dataset for stress research to aid the development of generalizable stress recognition systems. Three different datasets from the domain of stress recognition are integrated in an effort to increase the data volume necessary to effectively train machine- and deep-learning models. To this extent, publicly available stress datasets are evaluated based on specific criteria employed to determine the potential for feasible data integration. This research work makes the following contributions: 1) the criteria necessitated for data integration are presented from a data perspective, 2) a vanilla data integration pipeline used to obtain the dataset, and 3) a novel, combined, publicly available multi-modal dataset for automated stress detection and analysis, sampled from three existing datasets.

The remainder of this research paper is structured as follows. Section 2 discusses the publicly available datasets in the domain of stress recognition. Section 3 presents the data integration pipeline, including the criteria used to determine potential datasets. Next, section 4 presents summary statistics of the combined dataset. The findings are discussed in section 5 and, finally, a conclusion is provided in section 6.

## 2 Related Work

Research in the domain of stress recognition is usually conducted based on experiments in controlled laboratory settings (e.g., [6, 12]) or in field studies (e.g., [13–15]). Table 1 provides an overview of the publicly available datasets in the domain of stress detection. The overview presented here includes both types of experimental settings mentioned above, yet the data integration pipeline presented in this paper essentially relates to data collected controlled laboratory settings. During the controlled experiments, stress levels and labels are inferred through momentary psychological self-reports as mentioned above (e.g., [5]), continuous self-reports/ annotations [16], labels based on (stress-) stimuli [10], and labels based on stimuli, situation, body position, location [7]. Subjects' physiological responses are collected throughout the entire experiment. These kinds of datasets are usually employed for supervised learning to classify and distinguish different affective states based on the physiological responses (e.g., [10]). Stress recognition datasets do not always include labels, and are subsequently used for unsupervised learning tasks (e.g. [17]). [17] employ unsupervised learning algorithms to differentiate between different kinds of stress, arguing that in real-life labels are not always available. [7] point out that laboratory experiments should exploit the controlled and sequential nature of laboratory studies to provide detailed labels and annotations of the experiments such that other researchers can reproduce study results and replicate the experiments.

| Authors | Year | Dataset | Context |
|---|---|---|---|
| Schmidt et al. [10] | 2018 | WESAD | Laboratory |
| Campanella et al. [18] | 2024 | EmpaticaE4Stress | Laboratory |
| Schreiber et al. [19] | 2025 | VitaStress | Laboratory |
| Birjandtalab et al. [17] | 2016 | Non-EEG | Laboratory |
| Iqbal et al. [20] | 2022 | Stress-Predict | Laboratory |
| Parent et al. [12] | 2020 | PASS | Laboratory |
| Markova et al. [21] | 2019 | CLAS | Laboratory |
| Hongn et al. [22] | 2025 | Wearable Dataset | Laboratory |
| Koldijk et al. [23] | 2014 | SWELL-KW | Knowledge Work |
| Schneegas et al. [24] | 2013 | Driver Workload | Driving |
| Taylor et al. [25] | 2013 | Driving Monitoring Dataset (DMD) | Driving |
| Mundnich et al. [14] | 2020 | TILES-18 | Hospital Work |
| Yau et al. [15] | 2022 | TILES-19 | Hospital Work |
| Hosseini [13] | 2022 | Physiological Stress Signals | Hospital Work |

Table 1: Overview of publicly available datasets in the context of stress recognition

The datasets provided by Iqbal et al. [20] and Schmidt et al. [10] provide detailed labels and time logs of the different phases of the experiments. This kind of complementary data is crucial in terms of understanding what happened during the experiments and further aids in analysis. Hongn et al. [22] also provide this kind of complementary data, however, respective files only show the timestamps without an associated event, making the data less intuitive. The labels are still viable in combination with an info-graphic provided by the authors. The datasets provided by Campanella et al. [18] and Birjindtalab et al. [17] do not provide labels or additional information to infer what happened during the experiments, other than a study protocol described in the corresponding research articles. However, study protocols described in research papers may not always be unambiguously mapped to the corresponding data records. Following the framework for data collection and experiments in the domain of affective computing [7], meta- or complementary data such labels and annotations are an integral part of enhancing the re-usability of data.

Laboratory based stress studies focus on controlled interventions to induce different types of stress in subjects. Usually, sequences of different stress stimuli are employed, including a neutral baseline recording at the beginning of each session as well as recovery breaks and self-reports. Figure 1 provides an exemplary sequence of a experimental protocol. Common types of stress used in experiments include: physical, socio-evaluative, cognitive (or mental), and emotional stress. Physical stress has been induced either by immersing the hand of a subject in ice-cold water [9, 26, 27] or during physical exercise [17, 22]. Cognitive stress is usually induced by employing a mental arithmetic task, as seen in the Trier Social Stress Test (TSST) [28], the Stroop Color and Word Test [6], or even video games [12]. Socio-evaluative stress is elicited through a semi-public speaking task [28], during which subjects present a task before a jury. The public speaking task is conducted under the pretense of evaluation by an expert and is also part of the TSST. Considering the many studies and applications in automated stress detection systems, it becomes apparent that standardized protocols such as the TSST are reliable standards to replicate and reproduce experiments and data [7]. Furthermore, the data collected following such protocols also enhances the comparability of the data, as participants of independent cohorts have been subject to the same stress-eliciting procedure.

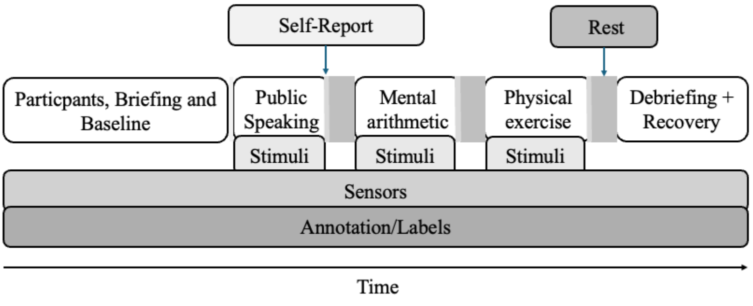



Figure 1: Experimental Setup following [7]

## 3 Data and Methods

This section describes the datasets used for the data integration and provides an overview of the integration process. The combined dataset can be downloaded from Zenodo https://doi.org/10.5281/zenodo.16842624.

### 3.1 Dataset Description

#### *3.1.1 VitaStress*

The VitaStress dataset was introduced by Schreiber et al. [19] to study physiological responses to different types of stress. The dataset comprises data of 21 subjects. Physiological signals were collected employing the Corsano Cardiowatch 287-2B, including the following data records: green, red, and infrared photoplethysmography (PPG, skin conductance (SC), accelerometer (ACC), temperature (TEMP). The dataset was recorded in a controlled lab setting and bridges cognitive and socio-evaluative stress with physical stress. The participants performed various tasks to induce the different stress types (e.g. arithmetic, public speaking, physical exercise). Furthermore, the dataset includes informative labels and annotations, and self-reports of participants. The different conditions served as the gold standard for stress assessment rather than the self-reports, although these are also included.

#### *3.1.2 WESAD*

The WESAD dataset was introduced by Schmidt et al. [10] to study different affective states. WESAD comprises data of 15 subjects. Physiological signals were collected employing the Empatica E4, whereas the data records available include the following physiological signals: Blood-Volume Pulse (BVP), Accelerometer (ACC), Electrodermal Activity (EDA), and Temperature (TEMP). The dataset was recorded in a controlled lab setting and bridges cognitive and socio-evaluative stress with an amusement state. The participants performed various tasks to induce stress or amusement (e.g., arithmetic, public speaking, watching video clips).

Furthermore, the dataset includes informative metadata, self-reports of participants, and labels (i.e., events during the experiment). The different conditions served as the gold standard for stress assessment rather than the self-reports, although these are also included.

#### *3.1.3 Wearable Dataset*

Another publicly available dataset is provided by Hongn et al. [22]. Similar to [17], this dataset aims to incorporate physiological responses during cognitive stress and physical activity to facilitate structural comparisons between cognitive and physical stress. The experimental protocol is carried out to evoke cognitive stress during several well-established stimuli (e.g., Stroop Color Word Test, mental arithmetic). Furthermore, the experimental protocol included anaerobic and aerobic exercises for a detailed analysis of physical activity.

### 3.2 Data Integration Pipeline

The data integration criteria on which the datasets are sampled is based on the overall goal of providing a dataset that is as coherent and uniform as possible considering the heterogeneity of the source datasets.

- **Physiological Signals.** To provide a complete dataset, data from two different sources may be integrated if they contain the same or a relevant subset of data.
- **Experimental Protocol.** Data from two different sources may be integrated if the underlying experimental protocols of the sources are semantically similar, and their combination thus enriches the information provided in each dataset.
- **Labels.** Data from two different sources may be integrated if the sources each contain informative labels and annotations. Labels and annotations refers to information about the sequence of the experiment, its different phases and stimuli.

The aim of this vanilla data integration process, is to combine several datasets with the VitaStress dataset [19]. To this extent, the datasets presented in the previous subsections were used for the integration process. The combined dataset may be used for classification of different types of stress as well as forecasting of various physiological signals. The final dataset contains physiological signals of 71 subjects.

The first step in the data integration process consists of dataset collection and selection. Based on Table 1 we initially select the WESAD dataset [10], and the Wearable dataset [22]. These datasets were selected as they meet the criteria defined above. Physiological signals of these datasets were collected using the Empatica E4, ensuring that data records are in a similar format. The experimental protocols are similar and each dataset includes informative labels. With regards to physiological signals, all the dataset selected contain a subset of the data records contained in the VitaStress dataset. While the datasets selected include BVP sampled at 64Hz VitaStress contains PPG signals (green, red, infrared, ambient) sampled at 64Hz, which can be used to calculate BVP.

Following the selection of relevant datasets, global and local schemata were defined to ensure seamless data integration. The global scheme are depicted in Figure 2. The data integration pipeline employed in this work is depicted in Figure 3.

As mentioned above, the VitaStress dataset contained PPG signals instead of BVP. To calculate BVP, the Empatica E4 employs two green and two red PPG LEDs. To approximate this calculation, all but the ambient PPG signal were used from the VitaStress dataset to calculate BVP per subject. This was achieved by fusing the different signals (taking an average), and bandpass filtering the fused signal to obtain a clean BVP signal [29, 30].

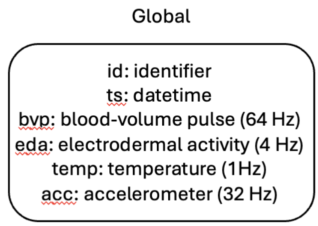



Figure 2: Global Scheme for Seamless Data Integration

To overcome misalignment in sampling frequency, skin conductance and temperature data were resampled. In particular, the skin conductance values sampled from the Corsano Cardiowatch 287-2B are sampled at 25 Hz and measured in pico-siemens [1], while the conductance values from the Empatica E4 are sampled at 4 Hz and measured in micro-siemens. To resolve this misalignment, the conductance values from the Corsano system were down-sampled to 4 Hz and converted into micro-siemens. Furthermore, temperature data from the Empatica E4 was down-sampled to 1 Hz to align with the sampling frequency of the temperature data from the Corsano system.

Considering the labeling, the combined dataset maps stressors to categories, rather than task specific labels. For example, stressors such as the Stroop Word Color Test, the TMCT, or TSST-mental arithmetic are labeled as cognitive stress. Other tasks such as presenting one's real or opposite opinion as well as the TSST-public speaking task are labeled as socio-evaluative stress [31]. Unlike the Wearable dataset, which provides task-level timestamps, WESAD labels only the TSST. Following the protocol [10]—first 5 min public speaking, next 5 min mental arithmetic—we split each TSST segment at its midpoint to separate socio-evaluative from cognitive stress.

The data files from the Empatica system were slightly adjusted to match the local schemata as shown above. The first two rows of the Empatica files contain a unix timestamp indicating the start of the recording, as well as the sampling frequency. This information is used to compute Python's datetime timestamps.

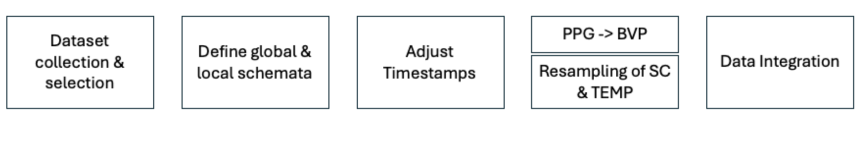


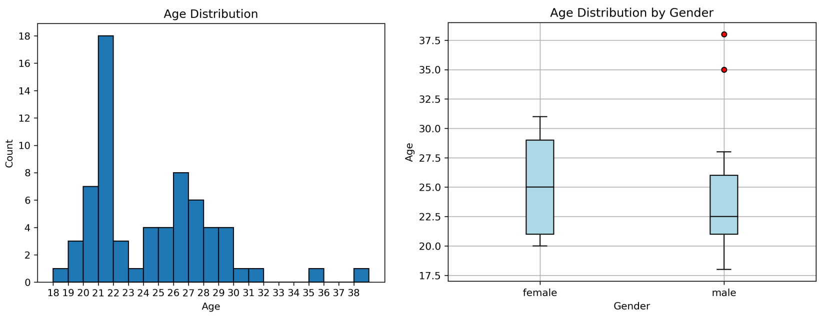

Figure 3: Data Integration Pipeline

## 4 Summary Statistics

Figure 4a show the age distribution of subjects in the combined dataset, which ranges from 18 to 38 years with an average age of 24,03 and a standard deviation of 4. Most participants

[1] https://corsano.com/knowledge-base/content-of-exported-data-file/

fall within the 21 to 28 age group, with a noticeable peak at 21 years. The distribution of age in the combined dataset is skewed to the right, indicating a higher concentration of younger participants and fewer individuals as age increases. In total, *age* is missing for five subjects.




(a) Age Distribution of Subjects (b) Gender Distribution of Subjects

Figure 4: Basic Demographic Information about the Subjects in the Combined Dataset

In terms of gender, the dataset consists of 49 male and 23 female participants, totaling 72 subjects. The boxplot 4b shows the age distribution by gender for a subset of the dataset with a value for *age*. For female participants, the mean age is 25 years. The inter-quartile range (IQR) spans from 21 to 29 years, with the overall age range extending from 20 to 31 years. For male subjects, the mean age is lower, with 22 years. The IQR ranges from 20 to 26 years, and the full range goes from 18 to 28 years. As pointed out above, two outliers are present at 35 and 38 years. Furthermore, the boxplot shows a male-dominant sample (in total 49 male, 23 female).

Figure 5 shows the distribution of the different classes of interest across all subjects and datasets. Baseline recordings are common at the beginning of stress experiments to record physiological signals during a neutral cognitive state; usually the baseline recordings take longer than other stimuli (e.g. [10]). Cognitive and social stressors each account for about 200 minutes, indicating a balanced inclusion of cognitive and socio-evaluative stimuli. Physical stress exposure is slightly less represented than the previous two stressors, owing to the fact that the physical stress condition from just the VitaStress dataset was considered. The amusement stimuli which is the least represented in the combined dataset was only used during the recording of the WESAD dataset, not however during the other two datasets considered here. The demographic factors described above as well as the class imbalance should be taken into account when analyzing this dataset and interpreting the results.

## 5 Discussion

This paper addresses the task of data integration in the domain of stress detection. The ISPAAD dataset bridges different stressors with physical activity as well as a positive affect, maintaining important semantic information about the different phases and stressors. Other publicly available datasets (see section 2), are smaller in terms of sample size, lacking rich diversity of subjects required to conduct in-depth analysis. The amount of samples provided in the ISPAAD datasets, can aid to train machine- and especially deep learning models more effectively, as the latter usually require large amounts of training data. Furthermore, instead of focusing on task-specific labeling, the combined dataset maps certain tasks to a broader category, such that the stressors *Stroop Word Color Test*, *TMCT* [22] or *TSST-mental arithmetic*

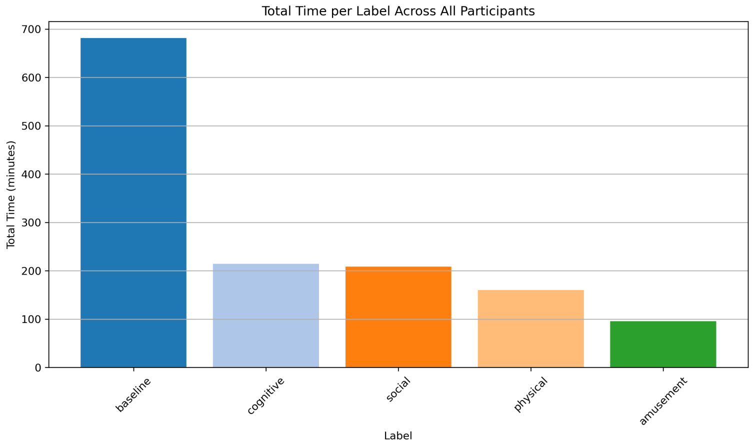



Figure 5: Distribution of (stress-) stimuli across all datasets

[10] are classified as cognitive stress. By including these annotations for different phases, the ISPAAD dataset enhances usability, and is suitable for both, supervised and unsupervised learning. To the best of our knowledge, this is the first work in the domain of stress detection to integrate multiple datasets.

A key limitation of this dataset is the slight under-representation of physical activity and amusement compared to cognitive and socio-evaluative stress. Moreover, the over-representation of the baseline period may bias models towards detecting rest over stress, however this can be addressed by down-sampling the baseline class. Another limitation owes to the synthetic label generation carried out for the WESAD dataset. As the authors pointed out in [10], they conducted both socio-evaluative and cognitve stressors consecutively, labeled as tsst instead of differentiating between these two types of stressors. This introduced uncertainty when splitting the tsst label into two new different synthetic labels. In addition, the demographics of the proposed dataset clearly show a gender imbalance and a bias towards young adults. While these biases should be taken into account when interpreting results, they underline that gender distributions are generally not balanced in current stress detection datasets, which needs to be addressed.

## 6 Conclusion

This research paper details the construction of ISPAAD, a novel, multi-modal, and integrated dataset intended to be utilized in stress detection research. Three different stress detection datasets are integrated, each including physiological signals such as blood volume pulse, skin conductance, skin temperature and 3-axis accelerometer data. Currently, the ISPAAD dataset comprises data of 72 subjects, and can be used to train machine- or deep learning models to classify stressful events, different types of stress, or forecast physiological signals during certain events. The contributions of this research work are threefold: 1) criteria to be considered for seamless data integration are outlined, 2) a vanilla data integration pipeline is presented, and 3) a novel and multi-modal dataset is presented that bridges cognitive and socio-evaluative stress with physical activity as well as a positive affective state. In particular, the diversity of cognitive and socio-evaluative stressors enables in-depth analysis and comparison between each stressor. Due to the scope of this work, benchmarking with predictive models is not included. As part of ongoing work, we aim to establish baseline performances trained on the combined dataset and augment the dataset.

## References

[1] K. Hovsepian, M. Al'Absi, E. Ertin, T. Kamarck, M. Nakajima, S. Kumar, cStress: towards a gold standard for continuous stress assessment in the mobile environment, in *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing* (2015), pp. 493–504

[2] G.E. Miller, E. Blackwell, Turning up the heat: Inflammation as a mechanism linking chronic stress, depression, and heart disease, Current Directions in Psychological Science **15**, 269 (2006).

[3] S. Cohen, T. Kamarck, R. Mermelstein, A global measure of perceived stress, Journal of health and social behavior pp. 385–396 (1983).

[4] C.D. Spielberger, F. Gonzalez-Reigosa, A. Martinez-Urrutia, L.F. Natalicio, D.S. Natalicio, The state-trait anxiety inventory, Revista Interamericana de Psicologia/Interamerican journal of psychology **5** (1971).

[5] M.M. Bradley, P.J. Lang, Measuring emotion: the self-assessment manikin and the semantic differential, Journal of behavior therapy and experimental psychiatry **25**, 49 (1994).

[6] M. Benchekroun, P.E. Velmovitsky, D. Istrate, V. Zalc, P.P. Morita, D. Lenne, Cross dataset analysis for generalizability of hrv-based stress detection models, Sensors **23**, 1807 (2023).

[7] P. Schreiber, F. Grensing, M. Maleshkova, TRRRACED–Towards Reproducible, Replicable and Reusable Affective Computing Experiments and Data, in *2024 IEEE 15th International Conference on Cognitive Infocommunications (CogInfoCom)* (IEEE, 2024), pp. 000147–000154

[8] J. Miranda, M. Canabal, J.M. Lanza-Gutiérrez, M. Portela-García, C. López-Ongil, T. Riesgo Alcaide, Meaningful data treatment from multiple physiological sensors in a cyber-physical system (2017).

[9] V. Mishra, G. Pope, S. Lord, S. Lewia, B. Lowens, K. Caine, S. Sen, R. Halter, D. Kotz, Continuous detection of physiological stress with commodity hardware, ACM transactions on computing for healthcare **1**, 1 (2020).

[10] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, K. Van Laerhoven, Introducing wesad, a multimodal dataset for wearable stress and affect detection, in *Proceedings of the 20th ACM international conference on multimodal interaction* (2018), pp. 400–408

[11] M. Gjoreski, H. Gjoreski, M. Luštrek, M. Gams, Continuous stress detection using a wrist device: in laboratory and real life, in *proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: Adjunct* (2016), pp. 1185–1193

[12] M. Parent, I. Albuquerque, A. Tiwari, R. Cassani, J.F. Gagnon, D. Lafond, S. Tremblay, T.H. Falk, Pass: a multimodal database of physical activity and stress for mobile passive body/brain-computer interface research, Frontiers in Neuroscience **14**, 542934 (2020).

[13] S. Hosseini, R. Gottumukkala, S. Katragadda, R.T. Bhupatiraju, Z. Ashkar, C.W. Borst, K. Cochran, A multimodal sensor dataset for continuous stress detection of nurses in a hospital, Scientific Data **9**, 255 (2022).

[14] K. Mundnich, B.M. Booth, M. l'Hommedieu, T. Feng, B. Girault, J. L'hommedieu, M. Wildman, S. Skaaden, A. Nadarajan, J.L. Villatte et al., Tiles-2018, a longitudinal physiologic and behavioral data set of hospital workers, Scientific Data **7**, 354 (2020).

[15] J.C. Yau, B. Girault, T. Feng, K. Mundnich, A. Nadarajan, B.M. Booth, E. Ferrara, K. Lerman, E. Hsieh, S. Narayanan, Tiles-2019: A longitudinal physiologic and behavioral data set of medical residents in an intensive care unit. scientific data 9, 1 (sept. 2022), 536 (2022)

[16] K. Sharma, C. Castellini, E.L. Van Den Broek, A. Albu-Schaeffer, F. Schwenker, A dataset of continuous affect annotations and physiological signals for emotion analysis, Scientific data **6**, 196 (2019).

[17] J. Birjandtalab, D. Cogan, M.B. Pouyan, M. Nourani, A non-EEG biosignals dataset for assessment and visualization of neurological status, in *2016 IEEE International Workshop on Signal Processing Systems (SiPS)* (IEEE, 2016), pp. 110–114

[18] S. Campanella, A. Altaleb, A. Belli, P. Pierleoni, L. Palma, Ppg and eda dataset collected with empatica e4 for stress assessment, Data in Brief **53**, 110102 (2024).

[19] P. Schreiber, S. Burbach, B. Cinar, L. Mackert, M. Maleshkova, VitaStress: A multimodal dataset for stress detection, in *Proceedings of the 27th International Conference on Multimodal Interaction* (2025), pp. 673–681

[20] T. Iqbal, A.J. Simpkin, D. Roshan, N. Glynn, J. Killilea, J. Walsh, G. Molloy, S. Ganly, H. Ryman, E. Coen et al., Stress monitoring using wearable sensors: A pilot study and stress-predict dataset, Sensors **22**, 8135 (2022).

[21] V. Markova, T. Ganchev, K. Kalinkov, Clas: A database for cognitive load, affect and stress recognition, in *2019 International Conference on Biomedical Innovations and Applications (BIA)* (IEEE, 2019), pp. 1–4

[22] A. Hongn, F. Bosch, L.E. Prado, J.M. Ferrández, M.P. Bonomini, Wearable physiological signals under acute stress and exercise conditions, Scientific Data **12**, 520 (2025).

[23] S. Koldijk, M. Sappelli, S. Verberne, M.A. Neerincx, W. Kraaij, The swell knowledge work dataset for stress and user modeling research, in *Proceedings of the 16th international conference on multimodal interaction* (2014), pp. 291–298

[24] S. Schneegass, B. Pfleging, N. Broy, F. Heinrich, A. Schmidt, A data set of real world driving to assess driver workload, in *Proceedings of the 5th international conference on automotive user interfaces and interactive vehicular applications* (2013), pp. 150–157

[25] P. Taylor, N. Griffiths, A. Bhalerao, D. Watson, X. Zhou, T. Popham, Warwick-JLR driver monitoring dataset (DMD): A public dataset for driver monitoring research, in *Proc. Cogn. Load-Vehicle Hum.-Mach. Interact.(CLW)* (2013), p. 27

[26] K. Plarre, A. Raij, S.M. Hossain, A.A. Ali, M. Nakajima, M. Al'Absi, E. Ertin, T. Kamarck, S. Kumar, M. Scott et al., Continuous inference of psychological stress from sensory measurements collected in the natural environment, in *Proceedings of the 10th ACM/IEEE international conference on information processing in sensor networks* (IEEE, 2011), pp. 97–108

[27] V. Mishra, S. Sen, G. Chen, T. Hao, J. Rogers, C.H. Chen, D. Kotz, Evaluating the reproducibility of physiological stress detection models, Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies **4**, 1 (2020).

[28] C. Kirschbaum, K.M. Pirke, D.H. Hellhammer, The 'trier social stress test'–a tool for investigating psychobiological stress responses in a laboratory setting, Neuropsychobiology **28**, 76 (1993).

[29] M. Elgendi, I. Norton, M. Brearley, D. Abbott, D. Schuurmans, Systolic peak detection in acceleration photoplethysmograms measured from emergency responders in tropical conditions, PloS one **8**, e76585 (2013).

[30] S.H. Chen, Y.C. Chuang, C.C. Chang, Development of a portable all-wavelength ppg sensing device for robust adaptive-depth measurement: A spectrometer approach with

a hydrostatic measurement example, Sensors **20**, 6556 (2020).
[31] P. Schmidt, A. Reiss, R. Dürichen, K. Van Laerhoven, Wearable-based affect recognition—a review, Sensors **19**, 4079 (2019).