Contrastive Learning of Electrodermal Activity Representations for Stress Detection

Katie Matton^{*1,2}**, Robert Lewis**^{*1}**, John Guttag**^{2}**, Rosalind W. Picard**^{1}
[1]MIT Media Lab, Massachusetts Institute of Technology
[2]CSAIL, Massachusetts Institute of Technology

Abstract

Electrodermal activity (EDA), usually measured as skin conductance, is a biosignal that contains valuable information for health monitoring. However, building machine learning models utilizing EDA data is challenging because EDA measurements tend to be noisy and sparsely labelled. To address this problem, we investigate applying contrastive learning to EDA. The EDA signal presents different challenges than the domains to which contrastive learning is usually applied (e.g., text and images). In particular, EDA is non-stationary and subject to specific kinds of noise. In this study, we focus on designing contrastive learning methods that are tailored to EDA data. We propose novel transformations of EDA signals to produce sets of positive examples within a contrastive learning framework. We evaluate our proposed approach on the downstream task of stress detection. We find that the embeddings learned with our contrastive pre-training approach outperform baselines, including fully supervised methods.

Electrodermal activity (EDA) refers to electrical changes that arise in the skin, and it is most commonly recorded using a measure of electrical conductance on the surface of the skin (1). Changes in EDA are related to sympathetic nervous system activity, and are representative of changes in psychological or physiological arousal (1; 2). EDA can be measured through wrist-worn smartwatches, which makes it a valuable signal for remote health monitoring. A leading application of EDA is stress estimation, and numerous lab studies have validated its association with stress (3; 2; 4; 5). Stress is thought to be a significant factor in many health conditions (6; 7). As such, several recent longitudinal studies have incorporated wrist-worn EDA measurements to deepen our understanding of how stress mediates conditions such as suicidal thinking (8), substance-use disorder 9, and post-traumatic stress disorder (10). Beyond stress, EDA is an important signal for remote monitoring of other conditions and constructs such as seizures, sleep, and pain (11; 12; 13).

There are two significant limitations to the analysis of EDA data collected in ambulatory settings. First, modeling most EDA-related outcomes relies on self-reported patient labels (e.g., psychological stress level) or clinical labels (e.g., a diagnosis). As such, EDA datasets are sparsely labelled; there is a natural upper bound on the number of labels one can collect (at most a few per day from a patient or clinician) relative to the number of EDA measurements one can record (ambulatory EDA runs at sample rates $\geq$4Hz). Second, wrist-worn EDA measurements are noisy (14). This noise often results from physical disruptions to the positioning of the watch, e.g., when the electrodes do not make consistent contact with the skin, or environmental factors, e.g., temperature and humidity.

One approach that can help to overcome the challenges of working with noisy and sparsely labelled data is *unsupervised contrastive learning*. This is a self-supervised method for learning useful data representations from unlabelled data. It works by (1) generating transformed versions of the input examples, and (2) encouraging the representations of transformations of the same (i.e., positive) examples to be close together, and transformations of different (i.e., negative) examples to be farther

[*]Equal contribution. Correspondence to `kmatton@mit.edu, roblewis@media.mit.edu.`

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

apart. The efficacy of unsupervised contrastive learning methods is contingent on choosing good transformations: they should be both challenging (i.e., it is non-trivial to distinguish between positive and negative examples) and label-preserving.

In domains such as computer vision and speech processing, there is considerable literature surrounding the selection of data transformations and the use of unsupervised contrastive learning (15; 16; 17; 18). More recently, contrastive methods have also been applied to biosignal time-series in the health domain, for example to electrocardiogram (ECG) data for cardiac arrhythmia detection and stress detection (19; 20; 21), and to electroencephalogram (EEG) data for sleep stage scoring and eye state classification (22; 23; 20). In contrast, there is almost no existing work that examines unsupervised contrastive learning with electrodermal activity data. The EDA signal has unique properties that make it considerably different from many of these other modalities – it is a non-periodic / non-stationary time-series that is subject to specific sources of noise when measured at the wrist. Therefore, we do not expect the findings from other domains to generalize directly to EDA data. In particular, it remains unclear what types of transformations work best for contrastive learning applied to EDA.

In this study, we take steps towards closing this gap. We present ongoing work to develop an unsupervised contrastive learning framework that incorporates transformations that are specific to the nature of the EDA signal. We evaluate these transformations on their ability to generate embeddings that have predictive utility in the downstream task of stress estimation. We first outline the methodology of our approach. We then present interim results on an EDA dataset with stress labels (WESAD, 24).

1 Methods

1.1 Contrastive Learning Formulation and Model Architecture

In line with Chen et al.(16), we formulate a self-supervised contrastive learning set up where we train a model to distinguish *positive examples* of the input signal from *negative examples*. We segment the EDA signal into windows of a fixed length ($|x_i| = M$ samples). We generate transformed versions of each segment $\tilde{x}_i = t(x_i)$ by applying a transform t that is randomly sampled from a set of transforms T (cf. Section 1.2). We define positive examples as those that have the same base segment (e.g., $\tilde{x}_i = t(x_i)$ and $\tilde{x}'_i = t'(x_i)$ where $t, t' \sim T$.), and negative examples as those that have different base segments ($x_j, \forall j : j \neq i$).

Our model architecture consists of an encoder, $f(\cdot)$, followed by a linear projection head, $g(\cdot)$. The encoder produces lower-dimensional embeddings of the input signal $h_i = f(\tilde{x}_i)$, which can be transferred to downstream prediction tasks. We implement it using a 1-D convolutional neural network (CNN). The projection head g further reduces the dimensionality of the input, producing $z_i = g(h_i)$. We implement this as a single linear layer. We use the *InfoNCE* loss (15) to optimize model parameters during pre-training:

$$L = -log \frac{exp(sim(z_i, z'_i)/\tau)}{\sum_j exp(sim(z_i, z'_j)/\tau)} \tag{1}$$

where τ is a temperature parameter, z_i, z'_i are the embeddings of a positive pair, z_i, z'_j are the embeddings of a negative pair, and $sim(z_i, z'_j)$ is the *cosine similarity* between embeddings. Further details on the model architecture and hyperparameters can be found in Appendix A.3.

1.2 Domain-Specific Transformations of the EDA Signal

We collaborated with several EDA experts to develop EDA-specific transformations. Details and visualizations are in Figure 1 and Appendix A.2.

Transformations that alter the high-frequency components of the signal. Existing work has found that nearly all of the information related to sympathetic nervous system activity is restricted to the low frequency components of EDA ($<= 0.24$ Hz) (25). Motivated by this finding, we design several transformations that primarily alter the parts of the signal where relevant information *is not* expected:

- *Low-Pass Filter:* We apply a filter that passes only signals with a lower frequency than a chosen cutoff frequency. We examine cutoff frequencies: $f = \{0.1, 0.25, 0.5, 0.75, 1.0\}$.
- *Band-Stop Filter:* We apply a filter that rejects / attenuates a specific band of frequencies, while letting all others pass. We examine reject frequencies: $f = \{0.1, 0.25, 0.5, 0.75, 1.0\}$.

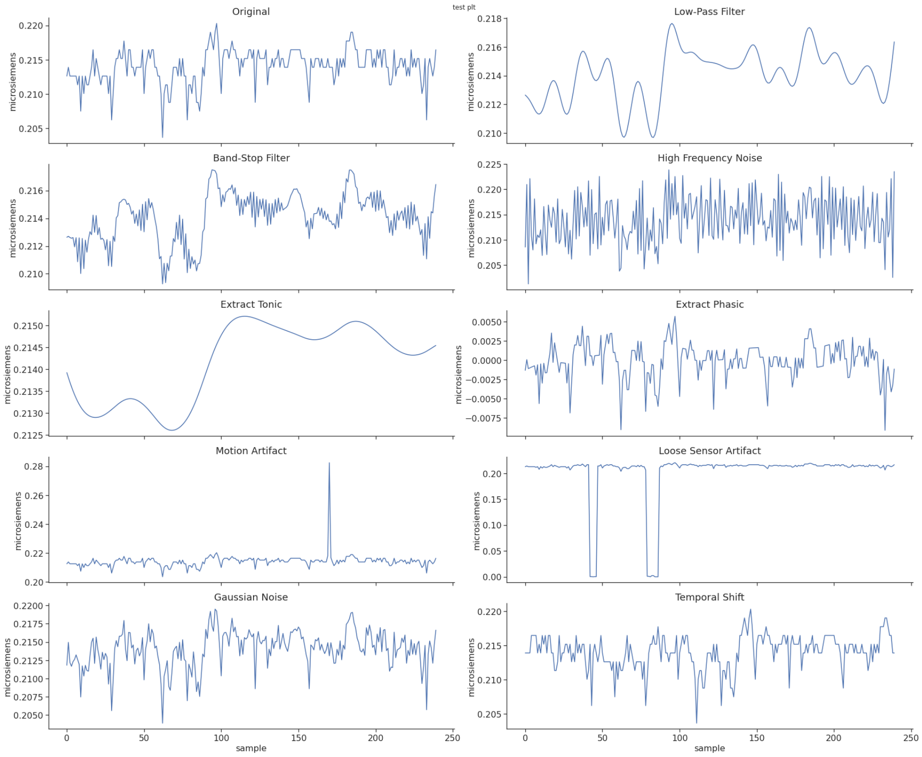

Figure 1: Example EDA segment with each data transformation applied. Parameters used for each transformation are the best parameters selected based on the validation performance (cf. Table 1). NB: the range of the y-axis varies between subplots.

- *High Frequency Noise:* We add noise to the high frequency bands of the signal. To do this, we map the signal to the frequency domain using an FFT, and add Gaussian noise to the frequency components starting at a chosen cutoff. We examine cutoff frequencies: $\{0.5, 1.0\}$ and $\sigma = \{0.1, 0.2, ..., 0.9, 1.0\}$ for the noise distribution.

Transformations based on signal decomposition. EDA signals are characterized by two components: (1) a slow-varying *tonic* component, which represents the skin conductance level (SCL) and (2) a faster-varying *phasic* component that represents the skin conductance response (SCR) (1; 14). Because SCRs can reflect responses to stimuli (such as external stressors), the phasic component of the signal is of particular interest for stress detection (26):

- *Extract Tonic:* Isolates the tonic component of the EDA signal.

- *Extract Phasic:* Isolates the phasic component of the EDA signal.

Transformations that simulate artifacts. EDA data, even when collected in lab settings, are prone to artifacts, i.e., changes in the signal that are *not* a result of the electrodermal system (27; 28). Because artifacts do not reflect sympathetic nervous system activity, we would like our model predictions to be invariant to them. To help achieve this, we simulate adding two common types of artifacts that stem from the data recording process:

- *Motion Artifact:* We add simulated motion artifacts to the signal, i.e., artifacts that arise when the placement of wearable sensors is altered due to movement. Motion artifacts often appear as abrupt peaks in the signal.

- *Loose Sensor Artifact:* We simulate the addition of artifacts that appear when the sensors lose contact with the skin. This typically results in the signal dropping to near-zero values.

Standard data transformations. We include transformations commonly used on other signals:

- *Gaussian Noise:* We add noise by sampling from a Gaussian distribution with $\sigma = \{0.1, 0.2, ..., 0.9, 1.0\}$. We scale σ for each signal to control the signal-to-noise ratio (SNR).

Transform	Best Parameters	Accuracy	F1
Low-Pass Filter	$f = 0.25$ Hz	0.78 (0.10)	0.54 (0.30)
Band-Stop Filter	$f = 1.0$ Hz	0.75 (0.11)	0.52 (0.30)
High Frequency Noise	$f = 0.5$ Hz, $\sigma = 0.4$	0.77 (0.13)	0.59 (0.26)
Extract Tonic	-	0.62 (0.21)	0.44 (0.27)
Extract Phasic	-	0.74 (0.11)	0.51 (0.34)
Motion Artifact	-	0.73 (0.18)	0.54 (0.32)
Loose Sensor Artifact	-	0.66 (0.19)	0.49 (0.33)
Gaussian Noise	$\sigma = 0.3$	0.79 (0.09)	0.58 (0.25)
Temporal Shift	$l = 190$	0.86 (0.10)	0.77 (0.18)

Table 1: Test performance of contrastive pre-training with a single transform on 10% of the labelled data. Best parameters are selected on validation performance. Mean (SD) is reported across 5 seeds.

- *Temporal Shift:* We shift the signal forward or backward in time. We examine shifts of length $l = \{10, 20, 30, ..., 240\}$ samples.

2 Experiments

2.1 Experimental Settings

Dataset. We use the WESAD dataset (24), a multimodal wearable sensor dataset for stress and affect detection. WESAD includes data collected from 15 subjects in a laboratory setting. Subjects were exposed to experimental conditions that were designed to elicit different affective states. We focus on the binary classification task of distinguishing between the stress and baseline (i.e., neutral) conditions. The WESAD dataset contains multiple physiological measures, but since our study focuses on EDA data, we use only the EDA wrist data in our experiments. As in prior work (24), we segment the EDA data into 60-second wide windows, overlapping with a window shift of 0.25 seconds. This produces 103,172 segments with a stressed or non-stressed label (based on the experimental condition). There are more non-stressed segments (65%) than stressed segments (35%).

Evaluation. To assess the utility of our proposed contrastive pre-training approach, we freeze the pre-trained encoder and examine the performance of a linear classifier trained on top of it. As baselines, we consider two other models with the same architecture (i.e., the same encoder and linear classifier): a model trained end-to-end with supervised learning (SL), and a linear classifier trained on top of a frozen, randomly initialized encoder (RE). During the contrastive learning phase we train with all examples (and assume no access to labels). During the supervised learning phase, we train on $x\%$ of the labelled data (we vary x from 10% to 100%). This setup allows us to examine how each method performs when we make different assumptions about the sparsity of labels. We perform leave-one-subject-out (LOSO) cross-validation to assess how well each method performs on *held-out* subjects. The details of our model architecture and hyper-parameter selection are provided in Appendix A.3.

2.2 Results and Discussion

Impact of individual transformations. To understand the utility of each transform we examine versions of the contrastive pre-training encoder trained with only one transform applied (i.e., for positive pairs of examples, one segment is transformed $\tilde{x}_i = t(x_i)$, while the other is not $\tilde{x}_i = x_i$). Table 1 displays the test accuracy and F1 score for each single-transform experiment, where 10% of the labelled data is used. The best parameters for each transform are selected based on the validation accuracy (see Appendix A.4). We see that the *Temporal Shift* transform achieves the best accuracy (Acc=0.86), followed by *Gaussian Noise* (0.79) and *Low-Pass Filter* (0.78). By contrast, we find that the *Extract Tonic* and *Loose Sensor Artifact* transforms perform worst with accuracies of 0.62 and 0.66, respectively.

We hypothesize that the strong performance of the *Temporal Shift* transform is partially due to the labelling procedure of the WESAD dataset. The labels are derived from experimental conditions, which are contiguous in time, so the *Temporal Shift* transform is necessarily label-preserving. In future work, we will investigate if this result holds in real-world datasets, where stress can fluctuate quickly, producing more frequent changes in the labels. The poor performance of the *Loose Sensor*

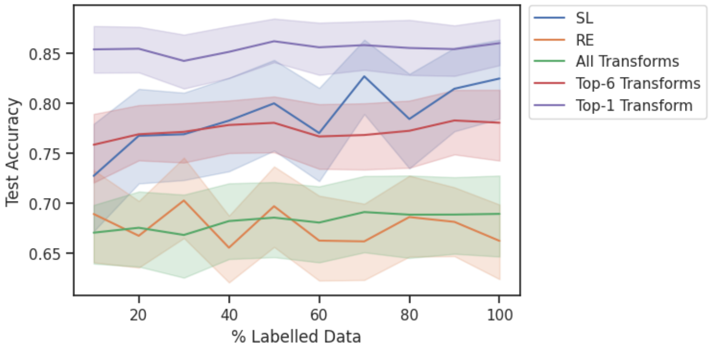

Figure 2: Test performance of models with contrastive pre-training compared to the *supervised learning* (SL) and *random encoder* (RE) baselines on different fractions of labelled data. Three versions of the contrastive pre-training encoder are shown: one with all of the transforms, one with the top 6 transforms (selected based on validation performance), and one with only the best performing single transform (*Temporal Shift*). Results are reported across 5 seeds.

Artifact transform may be related to the controlled data collection setting, where we expect to see few instances of improper sensor placement. The limited utility of the *Extract Tonic* transform aligns with previous findings that the phasic component of EDA is the most useful for stress detection (26).

Impact of amount of labelled data. We consider the performance of contrastive learning models relative to the baselines of supervised learning (SL) and random encoder (RE). Figure 2 compares these models over different fractions of the labelled data. We examine three different versions of the contrastive learning approach: one that uses all transforms, one that uses only the best transform (*Temporal Shift*, Top-1), and one that uses the Top-6 transforms (*Temporal Shift*, *Gaussian Noise*, *Low-Pass Filter*, *High Frequency Noise*, *Band-Stop Filter*, and *Motion Artifact*), where the transforms are ranked on validation performance from the single transform experiments (cf. Appendix A.4).

The baseline model performances are as expected: the accuracy of RE is low and does not increase notably with more labelled data, reflecting its limited capacity to learn, while the accuracy of SL gradually increases with more labelled data. Regarding the contrastive models, we see that Top-1 considerably outperforms all other models at all label fractions. Adding more transforms to the pre-training procedure *worsens* performance, with the *Top-6* model only exceeding SL performance at label fractions $\leq 20\%$, and the *All Transforms* model performing on par with RE. The performance of the *Temporal Shift* transform is remarkably strong; it beats the fully supervised model for all fractions of data (even 100%). The finding that the other transforms do not add utility beyond *Temporal Shift* is unexpected. However, most of the other transforms involve adding noise to the signal, and since the WESAD data was collected in a lab, it is relatively clean. In future work, it will be interesting to see if the utility of our proposed transforms is more apparent on noisier, naturalistic data.

3 Conclusion

This work shows the promise of using contrastive learning to address label sparsity and noise in EDA analysis. In future work, there are several limitations that we will address. First, the WESAD dataset was collected in an artificial laboratory setting, so we will need to investigate additional datasets to understand how our results translate to more naturalistic data. Second, stress is likely best categorized by multiple physiological modalities and therefore we will extend this work to consider multi-modal contrastive learning methods.

References

[1] W. Boucsein, *Electrodermal activity*. Springer Science & Business Media, 2012.

[2] M. E. Dawson, A. M. Schell, and D. L. Filion, “The electrodermal system.,” 2017.

[3] N. Sharma and T. Gedeon, “Objective measures, sensors and computational techniques for stress recognition and classification: A survey,” *Computer Methods and Programs in Biomedicine*, vol. 108, no. 3, pp. 1287–1301, 2012.

[4] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on psychological stress detection using biosignals," *IEEE Transactions on Affective Computing*, vol. 13, no. 1, pp. 440–460, 2019.

[5] Y. S. Can, B. Arnrich, and C. Ersoy, "Stress detection in daily life scenarios using smart phones and wearable sensors: A survey," *Journal of biomedical informatics*, vol. 92, p. 103139, 2019.

[6] H. Yaribeygi, Y. Panahi, H. Sahraei, T. P. Johnston, and A. Sahebkar, "The impact of stress on body function: A review," *EXCLI journal*, vol. 16, p. 1057, 2017.

[7] "Stress effects on the body."

[8] E. M. Kleiman, K. H. Bentley, J. S. Maimone, H.-I. S. Lee, E. N. Kilbury, R. G. Fortgang, K. L. Zuromski, J. C. Huffman, and M. K. Nock, "Can passive measurement of physiological distress help better predict suicidal thinking?," *Translational psychiatry*, vol. 11, no. 1, pp. 1–6, 2021.

[9] S. Carreiro, K. K. Chintha, S. Shrestha, B. Chapman, D. Smelson, and P. Indic, "Wearable sensor-based detection of stress and craving in patients during treatment for substance use disorder: A mixed methods pilot study," *Drug and Alcohol Dependence*, vol. 209, p. 107929, 2020.

[10] S. A. McLean, K. Ressler, K. C. Koenen, T. Neylan, L. Germine, T. Jovanovic, G. D. Clifford, D. Zeng, X. An, S. Linnstaedt, *et al.*, "The aurora study: a longitudinal, multimodal library of brain biology and function after traumatic stress exposure," *Molecular psychiatry*, vol. 25, no. 2, pp. 283–296, 2020.

[11] K. T. Johnson and R. W. Picard, "Advancing neuroscience through wearable devices," *Neuron*, vol. 108, no. 1, pp. 8–12, 2020.

[12] V. Bhatkar, R. Picard, and C. Staahl, "Combining electrodermal activity with the peak-pain time to quantify three temporal regions of pain experience," *Frontiers in Pain Research*, vol. 3, 2022.

[13] M. C. Ortega, E. Bruno, and M. P. Richardson, "Electrodermal activity response during seizures: A systematic review and meta-analysis," *Epilepsy & Behavior*, vol. 134, p. 108864, 2022.

[14] H. F. Posada-Quintero and K. H. Chon, "Innovations in electrodermal activity data collection and signal processing: A systematic review," *Sensors*, vol. 20, no. 2, p. 479, 2020.

[15] A. v. d. Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," *arXiv preprint arXiv:1807.03748*, 2018.

[16] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*, pp. 1597–1607, PMLR, 2020.

[17] A. Baevski, Y. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A framework for self-supervised learning of speech representations," *Advances in Neural Information Processing Systems*, vol. 33, pp. 12449–12460, 2020.

[18] A. Jaiswal, A. R. Babu, M. Z. Zadeh, D. Banerjee, and F. Makedon, "A survey on contrastive self-supervised learning," *Technologies*, vol. 9, no. 1, p. 2, 2020.

[19] D. Kiyasseh, T. Zhu, and D. A. Clifton, "Clocs: Contrastive learning of cardiac signals across space, time, and patients," in *International Conference on Machine Learning*, pp. 5606–5615, PMLR, 2021.

[20] J. Y. Cheng, H. Goh, K. Dogrusoz, O. Tuzel, and E. Azemi, "Subject-aware contrastive learning for biosignals," *arXiv preprint arXiv:2007.04871*, 2020.

[21] S. Rabbani and N. Khan, "Contrastive self-supervised learning for stress detection from ECG data," *Bioengineering*, vol. 9, no. 8, p. 374, 2022.

[22] M. N. Mohsenvand, M. R. Izadi, and P. Maes, "Contrastive representation learning for electroencephalogram classification," in *Machine Learning for Health*, pp. 238–253, PMLR, 2020.

[23] N. Wagh, J. Wei, S. Rawal, B. Berry, L. Barnard, B. Brinkmann, G. Worrell, D. Jones, and Y. Varatharajah, "Domain-guided self-supervision of eeg data improves downstream classification performance and generalizability," in *Machine Learning for Health*, pp. 130–142, PMLR, 2021.

[24] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and 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*, ICMI '18, (New York, NY, USA), p. 400–408, Association for Computing Machinery, 2018.

[25] H. F. Posada-Quintero, J. P. Florian, Á. D. Orjuela-Cañón, and K. H. Chon, "Highly sensitive index of sympathetic activity based on time-frequency spectral analysis of electrodermal activity," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 311, no. 3, pp. R582–R591, 2016.

[26] O. N. Rahma, A. P. Putra, A. Rahmatillah, Y. S. K. A. Putri, N. D. Fajriaty, K. Ain, and R. Chai, "Electrodermal activity for measuring cognitive and emotional stress level," *Journal of Medical Signals and Sensors*, vol. 12, no. 2, p. 155, 2022.

[27] S. Taylor, N. Jaques, W. Chen, S. Fedor, A. Sano, and R. Picard, "Automatic identification of artifacts in electrodermal activity data," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1934–1937, IEEE, 2015.

[28] S. Gashi, E. Di Lascio, B. Stancu, V. D. Swain, V. Mishra, M. Gjoreski, and S. Santini, "Detection of artifacts in ambulatory electrodermal activity data," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 2, pp. 1–31, 2020.

[29] D. Spathis, I. Perez-Pozuelo, S. Brage, N. J. Wareham, and C. Mascolo, "Self-supervised transfer learning of physiological representations from free-living wearable data," in *Proceedings of the Conference on Health, Inference, and Learning*, pp. 69–78, 2021.

[30] V. Dissanayake, S. Seneviratne, R. Rana, E. Wen, T. Kaluarachchi, and S. Nanayakkara, "Sigrep: Toward robust wearable emotion recognition with contrastive representation learning," *IEEE Access*, vol. 10, pp. 18105–18120, 2022.

[31] A. Saeed, F. D. Salim, T. Ozcelebi, and J. Lukkien, "Federated self-supervised learning of multisensor representations for embedded intelligence," *IEEE Internet of Things Journal*, vol. 8, no. 2, pp. 1030–1040, 2020.

[32] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and S. H. A. Chen, "NeuroKit2: A python toolbox for neurophysiological signal processing," *Behavior Research Methods*, vol. 53, pp. 1689–1696, feb 2021.

A Appendix

A.1 Related Work

Self-supervised learning (SSL) using *contrastive learning* (CL) techniques such as Contrastive Predictive Coding (CPC), SimCLR, and Wav2Vec has enabled significant progress on learning from noisy and sparsely labelled datasets in domains such as computer vision and speech processing (15; 16; 17; 18). More recently, contrastive methods have also been applied in health domains, for example to electrocardiogram (ECG) data for cardiac arrhythmia detection (19; 20) and stress detection (21), and to electroencephalogram (EEG) data for e.g., sleep stage scoring and eye state classification (open vs. closed) (22; 23; 20). Multimodal contrastive learning methods that use the signals from more than one sensor in a wearable have been considered in studies to predict generic health and demographic characteristics such as VO_2 max and age (29). Regarding CL methods for electrodermal activity (EDA), EDA has been included as a modality in multimodal CL systems that apply standard transforms across all signals generically (i.e., using transforms that are agnostic to the nature of the sensor signals) (30; 31). However, to the best of our knowledge, no previous work has focused exclusively on EDA in a domain-specific way that carefully considers the nature of the EDA signal – which is non-stationary and subject to specific sources of noise when measured at the wrist.

A.2 Implementation Details of Data Transformations

- *Low-pass Filter:* We apply a low-pass Butterworth filter of order 4 using the *signal filter* function provided in NeuroKit 2 (32).
- *Band-stop Filter:* We apply an IRR notch filter with quality factor $Q = 0.707$.
- *High Frequency Noise:* We apply a fast Fourier transform (FFT) to map the signal to the frequency space. We sample noise from a Gaussian distribution with a standard deviation proportional to σ. We scale σ based on the mean power of the signal, to control the resulting signal-to-noise ratio (SNR) across signals. We add noise to the FFT of the signal and then map it back to the time domain.
- *Extract Tonic/Phasic:* For these two transforms, we separate the signal into its two components using the high-pass filtering extraction method provided in NeuroKit2 (32)
- *Motion Artifact:* We sample a number of sensor spikes s (i.e., sharp peaks in the signal due to movement) from a uniform distribution with support $\{1, 2, 3\}$. For each spike s, we sample a spike location l from all possible indices, and a height factor h from a uniform distribution over $0.1 - 0.5$. To simulate a spike artifact, we compute the PDF of a Gaussian distribution centered at the spike location, scale this by the spike height factor, and add it to the original signal.
- *Loose Sensor Artifact:* We sample a number of sensor drops d (i.e., occurrences of the sensor losing contact with the skin) from a uniform distribution with support $\{1, 2, 3\}$. For each drop d, we randomly sample a start location s from all possible start indices, and a drop length l from a uniform distribution spanning $1 - 3$ seconds. We simulate a sensor drop by subtracting the mean power of the signal and zeroing out all resulting negative values.
- *Gaussian Noise:* We sample noise from a Gaussian distribution with zero-mean and standard deviation proportional to σ. We scale σ by the mean power of the signal to control the signal-to-noise ratio (SNR).
- *Temporal Shift:* We randomly sample whether to shift forward or backward in time with equal probability. We shift the signal by l samples, where l is the the length of the shift.

A.3 Model Architecture and Hyper-parameters

Model architecture. The following modules are used to create the different models in our experiments:

(a) **Encoder,** $f(\cdot)$: 1-D convolutional layers followed by 1-D max-pooling are used to transform the M-dimensional input signal, x_i, to a k-dimensional embedding, h_i. *Batch normalisation* and *dropout* are used in each convolutional *block* and *ReLU* is used as the activation function. For the experiments of this paper, we use 3 convolutional blocks in the encoder, transforming the input from a 240 dimensional input segment (60 seconds at 4Hz) to a 64 dimensional embedding. The input and output channels of these blocks are Conv_1(in=1, out=4), Conv_2(in=4, out=16) and Conv_3(in=16, out=32). A kernel size of 7 and a stride of 1 are used in the convolutional layers and a kernel size of 2 is used in the max-pooling operations. A single linear layer with *ReLU* activation then maps the output to 64 dimensions.

(b) **Projection head,** $g(\cdot)$: is implemented as a single linear layer that maps the 64-D embedding, h_i, to a 32-D embedding, z_i. The loss function 1 is computed on these 32-D embeddings.

(c) **Linear classifier for task prediction**: a linear classifier is used for supervised learning on the downstream prediction task (binary classification in the case of WESAD). This is implemented as a single linear layer mapping from the 64-D embeddings, h_i, to 1-D for the binary prediction. This model is trained for this task with binary cross entropy loss.

For the **contrastive learning models** in the *pre-training* phase, we compose a trainable version of the encoder module (a) and the projection head module (b). Then, for the downstream prediction phase, the pre-trained encoder is frozen and transferred to the downstream task, where the linear classifier (c) replaces the projection head and parameters in the linear classifier are optimized using the binary cross entropy loss.

For the **supervised learning (SL)** and **random encoder (RE)** models, there is only the downstream prediction phase, and these models comprise modules (a) and (c). In the case of SL, the model is trainable *end-to-end* (i.e., all parameters in (a) and (c) can be optimized), whereas for RE only the linear classifier module (c) can be trained.

Hyperparameters. Different hyperparameters are used for the contrastive pre-training and the supervised learning phases of the modeling.

- **Pre-training hyperparameters**: Adam is used as the optimizer with batch size of 256, learning rate of 0.001, and 400 epochs. The dropout probability is 0.1. Early stopping is implemented using the training InfoNCE loss.
- **Supervised learning hyperparameters**: Adam is used as the optimizer with batch size of 32, learning rate of 0.001, and 200 epochs. The dropout probability is 0.1. Early stopping is implemented using the validation error

A.4 Extended Single Transform Results

The validation accuracy and F1 score for the single-transform experiments with our proposed transformations and with transformations that are commonly applied to other signals are shown in tables 2 and 3 respectively.

Transform		Accuracy		F1	
Low-Pass Filter					
cutoff hz		mean	std	mean	std
0.10		0.732	0.068	0.451	0.209
0.25		0.786	0.068	0.590	0.176
0.50		0.782	0.063	0.582	0.171
0.75		0.769	0.070	0.541	0.198
1.0		0.769	0.065	0.554	0.170
Band-Stop Filter					
reject hz		mean	std	mean	std
0.10		0.740	0.061	0.462	0.194
0.25		0.716	0.052	0.391	0.180
0.50		0.749	0.066	0.506	0.183
0.75		0.728	0.066	0.466	0.199
1.0		0.772	0.065	0.564	0.171
High Frequency Noise					
cutoff hz	sigma	mean	std	mean	std
0.5	0.1	0.787	0.065	0.588	0.182
	0.2	0.790	0.073	0.595	0.194
	0.3	0.773	0.073	0.575	0.194
	0.4	0.792	0.075	0.608	0.194
	0.5	0.717	0.058	0.442	0.181
	0.6	0.728	0.067	0.471	0.202
	0.7	0.700	0.056	0.424	0.211
	0.8	0.710	0.049	0.394	0.204
	0.9	0.699	0.048	0.376	0.187
	1.0	0.738	0.073	0.512	0.206
1.0	0.1	0.757	0.059	0.513	0.178
	0.2	0.756	0.057	0.512	0.171
	0.3	0.758	0.059	0.510	0.186
	0.4	0.746	0.059	0.481	0.192
	0.5	0.762	0.060	0.521	0.186
	0.6	0.757	0.061	0.517	0.184
	0.7	0.765	0.057	0.523	0.181
	0.8	0.771	0.062	0.553	0.176
	0.9	0.766	0.064	0.531	0.191
	1.0	0.785	0.065	0.581	0.188
Extract Tonic		mean	std	mean	std
		0.718	0.061	0.448	0.194
Extract Phasic		mean	std	mean	std
		0.747	0.067	0.489	0.201
Motion Artifact		mean	std	mean	std
		0.773	0.082	0.569	0.216
Loose Sensor Artifact		mean	std	mean	std
		0.749	0.075	0.514	0.212

Table 2: Validation performance for single-transform experiments with our proposed transformations.

Transform	Accuracy		F1	
Gaussian Noise				
sigma	mean	std	mean	std
0.1	0.754	0.063	0.505	0.193
0.2	0.781	0.061	0.581	0.167
0.3	0.801	0.059	0.628	0.150
0.4	0.767	0.069	0.541	0.206
0.5	0.799	0.055	0.623	0.148
0.6	0.784	0.062	0.591	0.168
0.7	0.776	0.076	0.613	0.180
0.8	0.732	0.073	0.461	0.221
0.9	0.733	0.064	0.457	0.205
1.0	0.704	0.053	0.382	0.217
Temporal Shift				
shift length	mean	std	mean	std
10	0.798	0.055	0.626	0.152
20	0.803	0.060	0.644	0.153
30	0.820	0.069	0.677	0.167
40	0.832	0.064	0.710	0.139
50	0.829	0.068	0.706	0.149
60	0.849	0.067	0.744	0.144
70	0.844	0.062	0.733	0.142
80	0.836	0.067	0.716	0.156
90	0.851	0.068	0.740	0.155
100	0.845	0.068	0.736	0.141
110	0.862	0.069	0.772	0.137
120	0.838	0.065	0.718	0.148
130	0.852	0.073	0.748	0.155
140	0.862	0.069	0.769	0.139
150	0.858	0.067	0.762	0.138
160	0.860	0.070	0.757	0.155
170	0.865	0.063	0.773	0.130
180	0.867	0.068	0.773	0.146
190	0.877	0.059	0.797	0.115
200	0.857	0.064	0.752	0.141
210	0.865	0.056	0.770	0.120
220	0.857	0.053	0.757	0.113
230	0.850	0.057	0.740	0.127
240	0.843	0.067	0.717	0.154

Table 3: Validation performance for single-transform experiments with transformations that are commonly used for other signals.