

Original Paper

# A Comparison of Personalized and Generalized Approaches to Emotion Recognition Using Consumer Wearable Devices: Machine Learning Study

---

Joe Li; Peter Washington, PhD

Information and Computer Sciences, University of Hawai'i at Mānoa, Honolulu, HI, United States

**Corresponding Author:**

Peter Washington, PhD

Information and Computer Sciences

University of Hawai'i at Mānoa

1680 East-West Road, Room 312

Honolulu, HI, 96822

United States

Phone: 1 000000000

Email: [pyw@hawaii.edu](mailto:pyw@hawaii.edu)

## Abstract

**Background:** There are a wide range of potential adverse health effects, ranging from headaches to cardiovascular disease, associated with long-term negative emotions and chronic stress. Because many indicators of stress are imperceptible to observers, the early detection of stress remains a pressing medical need, as it can enable early intervention. Physiological signals offer a noninvasive method for monitoring affective states and are recorded by a growing number of commercially available wearables.

**Objective:** We aim to study the differences between personalized and generalized machine learning models for 3-class emotion classification (neutral, stress, and amusement) using wearable biosignal data.

**Methods:** We developed a neural network for the 3-class emotion classification problem using data from the Wearable Stress and Affect Detection (WESAD) data set, a multimodal data set with physiological signals from 15 participants. We compared the results between a participant-exclusive generalized, a participant-inclusive generalized, and a personalized deep learning model.

**Results:** For the 3-class classification problem, our personalized model achieved an average accuracy of 95.06% and an  $F_1$ -score of 91.71%; our participant-inclusive generalized model achieved an average accuracy of 66.95% and an  $F_1$ -score of 42.50%; and our participant-exclusive generalized model achieved an average accuracy of 67.65% and an  $F_1$ -score of 43.05%.

**Conclusions:** Our results emphasize the need for increased research in personalized emotion recognition models given that they outperform generalized models in certain contexts. We also demonstrate that personalized machine learning models for emotion classification are viable and can achieve high performance.

(JMIR AI 2024;3:e52171) doi: [10.2196/52171](https://doi.org/10.2196/52171)

## KEYWORDS

affect detection; affective computing; deep learning; digital health; emotion recognition; machine learning; mental health; personalization; stress detection; wearable technology

## Introduction

Stress and negative affect can have long-term consequences for physical and mental health, such as chronic illness, higher mortality rates, and major depression [1-3]. Therefore, the early detection and corresponding intervention of stress and negative emotions greatly reduces the risk of detrimental health conditions appearing later in life [4]. Since negative stress and

affect can be difficult for humans to observe [5-7], automated emotion recognition models can play an important role in health care. Affective computing can also facilitate digital therapy and advance the development of assistive technologies for autism [8-13].

Physiological signals, including electrocardiography (ECG), electrodermal activity (EDA), and photoplethysmography (PPG), have been shown to be robust indicators of emotions [14-16].

The noninvasive nature of physiological signal measurement makes it a practical and convenient method for emotion recognition. Wearable devices such as smartwatches have become increasingly popular, and products such as Fitbit have already integrated the sensing of heart rate, ECG, and EDA data into their smartwatches. The accessibility of wearable devices indicates that an emotion recognition model using biosignals can have practical applications in health care.

The vast majority of research in recognizing emotions from biosignals involves machine learning models that are generalizable, which means that the models were trained on one group of subjects and tested on a separate group of subjects [17-28]. Prior studies emphasize the need for personalized or subject-dependent models [18,29,30], and some investigations, albeit few, analyze personalized models [31,32]. Both generalized and personalized models have potential benefits; for example, generalized models can train on more data than personalized models, and personalized models do not need to address the problem of inter-subject data variance [33]. However, it is still unclear how personalized models compare against generalized models in many contexts.

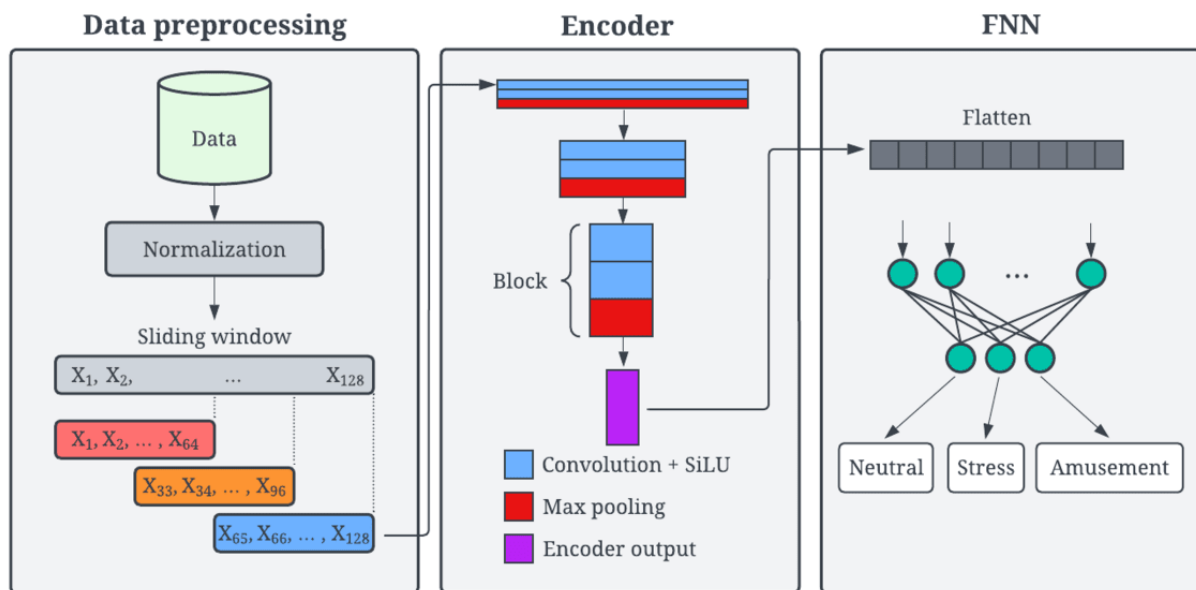
We present 1 personalized and 2 generalized machine learning approaches for the 3-class emotion classification problem (neutral, stress, and amusement) on the Wearable Stress and Affect Detection (WESAD) data set, a publicly available data set that includes both stress and emotion data [18]. The two generalized models are trained using participant-inclusive and participant-exclusive procedures. We compare the performance of these 3 models, finding that the personalized machine learning approach consistently outperforms the generalized approach on the WESAD data set.

## Methods

### Overview

To classify physiological data into the neutral, stress, and amusement classes, we developed a machine learning framework and evaluated the framework using data from the WESAD data set. Our machine learning framework consists of data preprocessing, a convolutional encoder for feature extraction, and a feedforward neural network for supervised prediction (Figure 1). Using this model architecture, we compared generalized and personalized approaches to the 3-class emotion classification task (neutral, stress, and amusement).

**Figure 1.** Overview of our model architecture for the 3-class emotion classification task. FNN: feedforward neural network; SiLU: sigmoid linear unit.



### Data Set

We selected WESAD, a publicly available data set that combines both stress and emotion annotations. WESAD consists of multimodal physiological data in the form of continuous time-series data for 15 participants and corresponding annotations of 4 affective states: neutral, stress, amusement, and meditation. However, we only considered the neutral, stress, and amusement classes since the objective of WESAD is to provide data for the 3-class classification problem, and the benchmark model in WESAD ignores the meditation state as well. Our model incorporated data from 8 modalities recorded

in WESAD: ECG, EDA, electromyogram (EMG), respiration, temperature, and acceleration (x, y, and z axes). In the data set, measurements for each of the 8 modalities were sampled by a RespiBAN sensor at 700 Hz to enforce uniformity, and data were collected for approximately 36 minutes per participant.

### Preprocessing and Partitioning

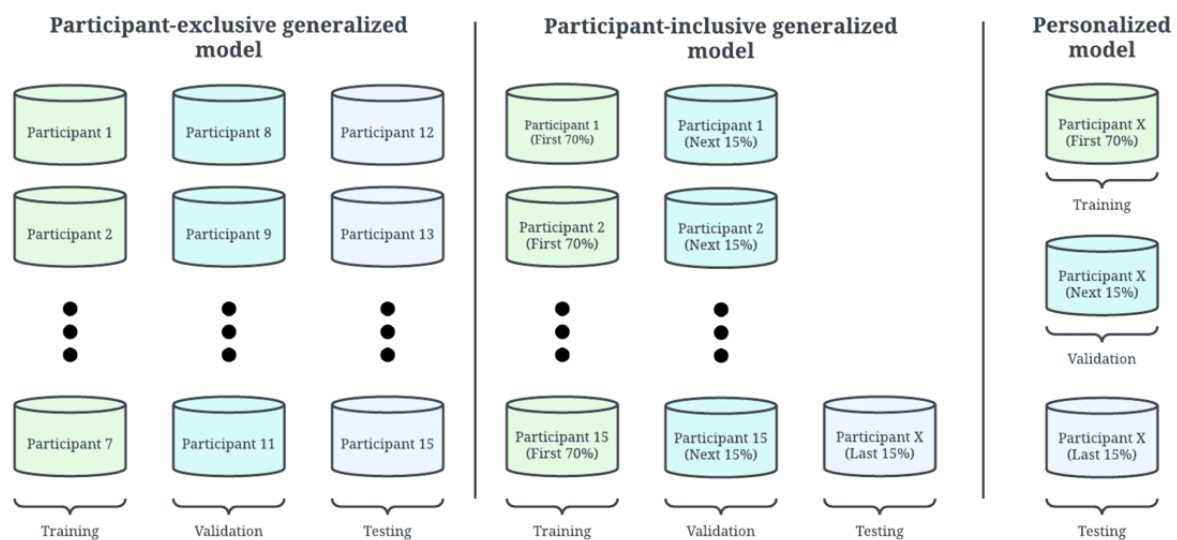
Each data modality was normalized with a mean of 0 and an SD of 1. We used a sliding window algorithm to partition each modality into intervals consisting of 64 data points, with a 50% overlap between consecutive intervals. We ensured that all 64 data points within an interval shared a common annotation,

which allowed us to assign a single affective state to each interval. The process of normalization, followed by a sliding window partition, is illustrated in Figure 1. These intervals were partitioned into training, validation, and testing sets.

For the personalized model, we partitioned the training, validation, and testing sets as follows: each participant in the data set had their own model that was trained, validated, and tested independently of other participants. For each affective state (neutral, stress, and amusement), we allocated the initial 70% of intervals with that affective state for training, the next

15% for validation, and the final 15% for testing. This guaranteed that the relative frequencies of each affective state were consistent across all 3 sets. Simply using the first 70% of all intervals for the training data would skew the distribution of affective states, given the nature of the WESAD data set. Furthermore, our partitioning of intervals according to sequential time order rather than random selection helped prevent overfitting by guaranteeing that 2 adjacent intervals with similar features would be in the same set. The partitioning of training, validation, and testing sets for the personalized model is shown in Figure 2.

**Figure 2.** A comparison of different generalized and personalized approaches to the 3-class emotion classification task. The participant-exclusive generalized model mimics generalized approaches used in other papers. The participant-exclusive generalized model shown in the figure differs from what we use in this paper.



Standard generalized models partition the training, validation, and testing sets by participant [18]. We denote these standard models as participant-exclusive generalized models, as shown in Figure 2. Through this partitioning method, it is impossible to compare the performances of generalized and personalized models since they are solving two separate tasks. Therefore, we present a modified participant-exclusive generalized model that solves the same task as the personalized model. The testing set for our participant-exclusive generalized model consisted of the last 15% of intervals for each affective state for 1 participant. The training set consisted of the first 70% of intervals for each affective state for all participants except the 1 participant in the testing set, and the validation set consisted of the next 15% of intervals for all participants except the 1 participant in the testing set. The training and testing sets for this approach contained data from mutually exclusive sets of participants; this is where the name of the model, participant-exclusive, is derived from. Since the testing sets for the participant-exclusive generalized and personalized models are equivalent, it is possible to compare generalized and personalized approaches. This participant-exclusive generalized model served as our first generalized model baseline.

A second generalized model baseline was created, called the participant-inclusive generalized model. Like the testing sets

for the participant-exclusive generalized and personalized models, the testing set for this model contained the last 15% of intervals for each affective state for a single participant. The training set consisted of the first 70% of intervals for each affective state for all participants, and the validation set consisted of the next 15%. The set of participants in the training and testing sets overlapped by 1 participant—the subject in the testing set—which is why this model is called the participant-inclusive generalized model. This is illustrated in Figure 2.

### Model Architecture

The model architecture consisted of an encoder network followed by a feedforward head, which is shown in Figure 1. A total of 8 channels, representing the 8 modalities we used from WESAD, served as input into an encoder network, which was modeled after the encoder section of U-Net [34]. The encoder network had 3 blocks, with each block consisting of two 1D convolutional layers (kernel size of 3) followed by 1D max pooling (kernel size of 2). The output of each convolution operation was passed through a sigmoid linear unit (SiLU) activation function. Between each block, we doubled the number of channels and added a dropout layer (15%) to reduce overfitting. The output of the encoder was flattened and passed

through 2 fully connected layers with SiLU activation to produce a 3-class probability distribution. Table 1 shows the hyperparameters that determine the model structure. These were

consistent between the participant-exclusive generalized, participant-inclusive generalized, and personalized models.

**Table 1.** Hyperparameters relating to model structure.

Hyperparameter	Value
Encoder depth (number of blocks), n	3
Dropout rate, %	15
Number of fully connected layers, n	2
Convolutional kernel size, n	3
Max pooling kernel size, n	2
Activation function	SiLU <sup>a</sup>

<sup>a</sup>SiLU: sigmoid linear unit.

### Model Training

We trained the 2 generalized baseline models and the personalized model under the same hyperparameters to guarantee a fair comparison. Both models were trained with cross-entropy loss using AdamW optimization. All models were written using PyTorch [35]. Within 1000 epochs, models with the lowest validation loss were saved for testing. A Nvidia GeForce RTX 4090 GPU was used for training. A separate personalized model was trained for each of the 15 participants. The participant-exclusive generalized model was trained 15 times, and the participant-inclusive generalized model was trained once. For model comparison, all models were tested on each of the 15 participants.

### Ethical Considerations

This study did not require institutional review board (IRB) review because we exclusively used a commonly analyzed publicly available data set. We did not work with any human subjects.

## Results

For the 3-class emotion classification task (neutral, stress, and amusement), Tables 2 and 3 illustrate the accuracy and  $F_1$ -score of the personalized and generalized models when tested on each of the 15 participants. We include  $F_1$ -score, a balanced evaluation metric consisting of the harmonic mean of precision

and recall, to accommodate for the imbalanced class distribution in WESAD [18]. In order to guarantee a fair comparison between the models, they had the same random seeds for model initialization, and their architecture and hyperparameters were the same. The accuracy and  $F_1$ -score for the personalized model exceeded those of the participant-inclusive generalized model for all participants except participant 1, and the personalized model outperformed the participant-exclusive generalized model in terms of accuracy and  $F_1$ -score for all participants. The personalized models for participants 1 and 2 also indicate subpar performance compared to other participants, which we address in the Discussion section.

Table 4 shows the average and SD of the accuracies and  $F_1$ -scores across all participants for the 3 models. We achieved an average accuracy of 95.06%, 66.95%, and 67.65% for the personalized, participant-inclusive generalized, and participant-exclusive generalized models, respectively. We also achieved an average  $F_1$ -score of 91.72%, 42.50%, and 43.05% for the personalized, participant-inclusive generalized, and participant-exclusive generalized models, respectively. Observing the error margins in Table 4, the differences in accuracy and  $F_1$ -score between the personalized model and both generalized models are statistically significant. As shown in Table 5, we evaluated the  $P$  values between each model type for accuracy and  $F_1$ -score through pairwise 2-tailed  $t$  tests to determine statistical significance.

**Table 2.** A comparison of model accuracy between the personalized and generalized models.

Participant	Model accuracy, %		
	Personalized model	Participant-inclusive generalized model	Participant-exclusive generalized model
1	68.36	82.69	53.94
2	82.32	67.12	81.91
3	99.99	82.81	82.81
4	99.90	82.86	82.31
5	98.02	82.94	74.67
6	99.57	54.57	54.03
7	100.00	82.05	83.23
8	100.00	53.72	53.70
9	100.00	51.86	51.83
10	93.69	82.05	79.85
11	100.00	60.86	62.11
12	98.34	53.53	53.60
13	99.81	53.26	65.35
14	100.00	53.47	53.54
15	85.83	60.43	81.91

**Table 3.** A comparison of  $F_1$ -score between the personalized and generalized models.

Participant	$F_1$ -score, %		
	Personalized model	Participant-inclusive generalized model	Participant-exclusive generalized model
1	58.14	61.91	23.36
2	58.88	44.55	58.53
3	99.98	62.05	62.05
4	99.87	61.95	61.50
5	96.87	61.99	54.74
6	99.35	24.94	23.59
7	100.00	61.16	62.09
8	100.00	23.38	23.29
9	100.00	22.85	22.89
10	94.29	61.04	59.23
11	100.00	38.27	40.15
12	97.40	26.79	26.90
13	99.75	24.47	44.63
14	100.00	23.93	24.09
15	71.28	38.26	58.71

**Table 4.** Average accuracy and  $F_1$ -score of models across all participants.

Model type	Accuracy, mean (SD [%])	$F_1$ -score, mean (SD [%])
Personalized	95.06 (9.24)	91.72 (15.33)
Participant-inclusive generalized	66.95 (13.76)	42.50 (17.37)
Participant-exclusive generalized	67.65 (13.48)	43.05 (17.20)

**Table 5.** *P* values of accuracy and  $F_1$ -score comparisons between model types.

Model comparison	<i>P</i> value for accuracy	<i>P</i> value for $F_1$ -score
Personalized versus participant-inclusive generalized	$P < .001$	$P < .001$
Personalized versus participant-exclusive generalized	$P < .001$	$P < .001$
Participant-inclusive generalized versus participant-exclusive generalized	.81	.88

## Discussion

### Principal Findings

We demonstrated that a personalized deep learning model outperforms a generalized model in both the accuracy and  $F_1$ -score metrics for the 3-class emotion classification task. By establishing two generalized model baselines through the participant-inclusive and participant-exclusive models, we created an alternative approach to the standard generalization technique of separating the training and testing sets by participant, and as a result, we were able to compare personalized and generalized approaches. Our personalized model achieved an accuracy of 95.06% and an  $F_1$ -score of 91.72%, while our participant-inclusive generalized model achieved an accuracy of 66.95% and an  $F_1$ -score of 42.50% and our participant-exclusive generalized model achieved an accuracy of 67.65% and an  $F_1$ -score of 43.05%.

Our work indicates that personalized models for emotion recognition should be further explored in the realm of health care. Machine learning methods for emotion classification are clearly viable and can achieve high accuracy, as shown by our personalized model. Furthermore, given that numerous wearable technologies collect physiological signals, data acquisition is both straightforward and noninvasive. Combined with the popularity of consumer wearable technology, it is feasible to scale emotion recognition systems. This can ultimately play a major role in the early detection of stress and negative emotions, thus serving as a preventative measure for serious health problems.

### Comparison With Previous Work

#### Generalized Models

The vast majority of prior studies using WESAD developed generalized approaches to the emotion classification task. Schmidt et al [18], the pioneers of WESAD, created several feature extraction models and achieved accuracies up to 80% for the 3-class classification task. Huynh et al [22] developed a deep neural network, trained on WESAD wrist signals, to outperform past approaches by 8.22%. Albaladejo-González et al [36] achieved an  $F_1$ -score of 88.89% using an unsupervised local outlier factor model and 99.03% using a supervised multilayer perceptron. Additionally, they analyzed the transfer learning capabilities of different models between the WESAD and SWELL-KW (SWELL knowledge work) [37] data sets. Ghosh et al [38] achieved 94.8% accuracy using WESAD chest data by encoding time-series data into Gramian Angular Field images and employing deep learning techniques. Bajpai et al [39] investigated the k-nearest neighbor algorithm to explore the tradeoff between performance and the total number of

nearest neighbors using WESAD. Through federated learning, Almadhor et al [40] achieved 86.82% accuracy on data in WESAD using a deep neural network. Behinaein et al [41] developed a novel transformer approach and achieved state-of-the-art performance using only one modality from WESAD.

#### Personalized Models

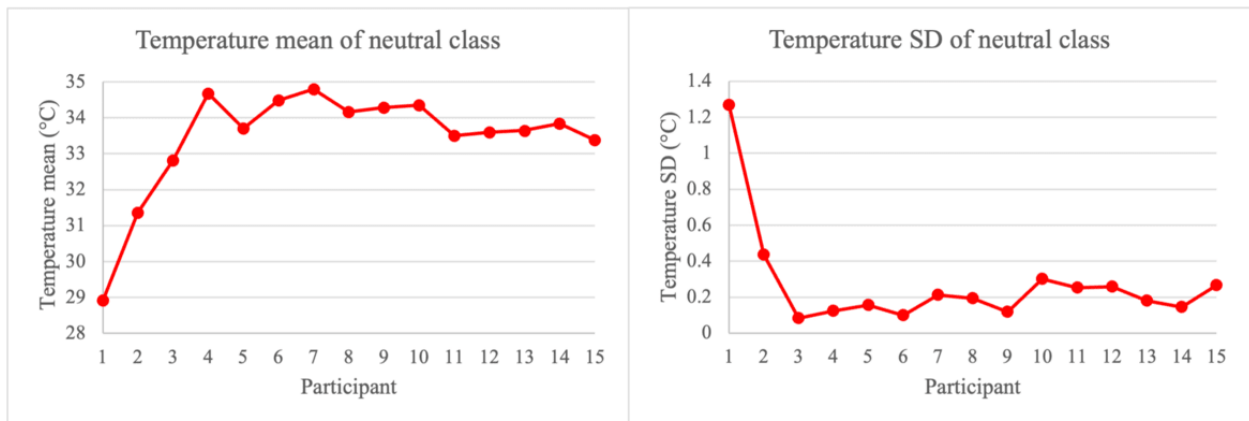
Sah and Ghasemzadeh [30] developed a generalized approach using a convolutional neural network using 1 modality from WESAD. For the 3-class classification problem, they achieved an average accuracy of 92.85%. They used the leave-one-subject-out (LOSO) analysis to highlight the need for personalization. Indikawati and Winiarti [31] directly developed a personalized approach for the 4-class classification problem in WESAD (neutral, stress, amusement, and meditation). Using different feature extraction machine learning models, they achieved accuracies ranging from 88%-99% for the 15 participants. Liu et al [32] developed a federated learning approach using data from WESAD with the goal of preserving user privacy. In doing so, they developed a personalized model as a baseline, which achieved an average accuracy of 90.2%. Nkurikiyeyezu et al [42] determined that personalized models (95.2% accuracy) outperform generalized models (42.5% accuracy) for the stress versus no-stress task. By running additional experiments to further understand how personalized models compare to generalized models for the 3-class emotion classification task and by developing participant-inclusive and participant-exclusive versions of the generalized models, our work concretely demonstrates how personalization outperforms generalization and thus supports the conclusions of Nkurikiyeyezu et al [42].

#### Limitations and Future Work

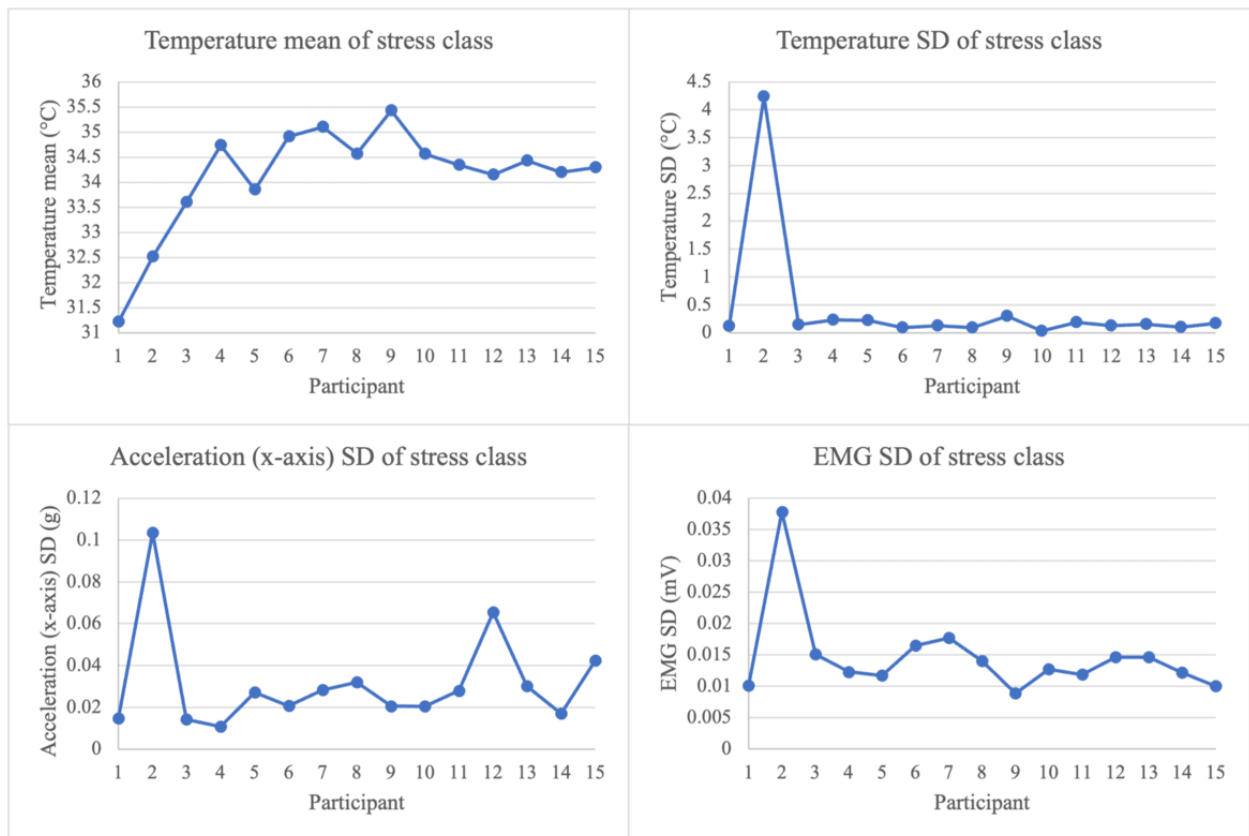
As shown in Tables 2 and 3, the performance of our personalized model deteriorates for participants 1 and 2. To analyze the lack of performance improvement of the personalized model for these 2 participants, we visualized the means and SDs of the different modalities for each emotion class. In Figures 3-5, we illustrate notable deviations in modality means and SDs for participants 1 and 2 compared to other participants. While the analysis of these modalities reveals important information about the nature of the WESAD data set, it still remains difficult to pinpoint the exact data set features that caused the performance decline in the personalized model for these 2 participants. This is another limitation: since we do not use a feature extraction model, we cannot assign a feature importance (eg, Gini importance) to individual features like Schmidt et al [18] do. We also analyzed the emotion class balances for each participant, which are included in Table 6, to see if anomalies existed in the class distributions for certain participants.

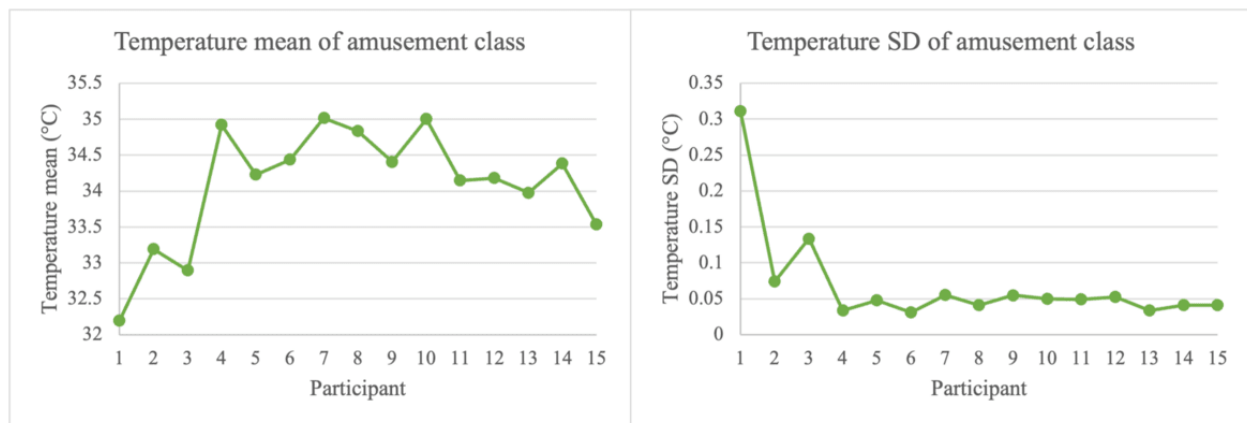
However, based on the ranges of the class distributions, class balance likely had minimal effect on the performance decline.

**Figure 3.** Deviations of mean and SD for participants 1 and 2 for neutral class modalities.



**Figure 4.** Deviations of mean and SD for subjects 1 and 2 for stress class modalities. EMG: electromyogram.



**Figure 5.** Deviations of mean and SD for subjects 1 and 2 for amusement class modalities.**Table 6.** Ranges of emotion class distributions per participant.

Emotion class	Range, %
Neutral	51.8-54.0
Stress	29.0-31.8
Amusement	16.3-17.4

Our participant-inclusive and participant-exclusive generalized models do not outperform previously published generalized models on the WESAD data set (eg, Schmidt et al [18] achieved up to 80% accuracy while we achieved 66.95% accuracy with our participant-inclusive model). This discrepancy can be attributed to a deliberate choice in our methodology: instead of maximizing our generalized models' performance with hyperparameter tuning, we simply opted for a consistent set of hyperparameters across the personalized and generalized models because our primary objective was to evaluate their relative performance. While hyperparameter tuning might yield higher results in practice, differing hyperparameters between our models would introduce additional variables that make it difficult to determine the role that personalization and generalization play in model performance.

Given the variations between participants, one approach to improving generalized model performance is adding embedding representations for each participant or participant-specific demographic data as additional features as a method of distinguishing individual participants in generalized models. However, to prevent overfitting to participant-specific features like demographic data, data sets with significantly more participants would need to be created, given the small sample size of the WESAD data set.

One limitation that personalized models may encounter during training is the cold start problem, given that personalized models receive less data than generalized models. Moreover, despite the accuracy improvement in personalized models, developing a model for each participant may be costly and unscalable: data must be labeled specifically per participant, and enough data must be provided to the model to overcome the cold start problem (notably, however, even though the cold start problem should theoretically put our personalized model at a disadvantage, the WESAD data set provided enough data for

our personalized model to outperform our generalized model). Both of these limitations can be addressed by a self-supervised learning approach to emotion recognition.

A self-supervised learning approach follows a framework used by natural language processing models such as the Bidirectional Encoder Representations from Transformers (BERT) model [43]. A model first pretrains on a large set of unlabeled data across numerous participants. Then, the pretrained model is fine-tuned to a small amount of labeled, participant-specific data. The pretraining phase eliminates the burden of manual labeling because all data are unlabeled, as well as the cold start problem because large amounts of data can be provided. The fine-tuning phase requires only a small amount of user-specific labeled data to perform accurately, and studies have already begun exploring the tradeoffs between the number of labels and model accuracy in WESAD using self-supervised or semisupervised approaches [44,45].

Finally, to expand beyond the WESAD data set, it is valuable to reproduce results on additional physiological signal data sets for emotion analysis, such as the Database for Emotion Analysis using Physiological Signals (DEAP) [46] and Cognitive Load, Affect, and Stress (CLAS) [47]. Data from WESAD were collected under controlled laboratory environments, which may not generalize to the real world. Therefore, analyzing emotions in a real-world context through data sets such as K-EmoCon [48], which contain physiological data collected in naturalistic conversations, may be useful. Emotions in the K-EmoCon data set were categorized into 18 different classes, so exploring this data set could also help us better assess the benefits of personalization for a broader range of emotions. A major goal of this approach is to provide support for personalized digital interventions for neuropsychiatry, which could benefit a variety of applications, such as video-based digital therapeutics for



children with autism to predict the child's affective state as part of the therapeutic process [49-52].

## Acknowledgments

The project described was supported by grant U54GM138062 from the National Institute of General Medical Sciences (NIGMS), a component of the National Institutes of Health (NIH), and its contents are solely the responsibility of the author and do not necessarily represent the official view of NIGMS or NIH. The project was also supported by a grant from the Medical Research Award fund of the Hawai'i Community Foundation (grant MedRes\_2023\_00002689).

## Conflicts of Interest

None declared.

## References

1. Kendler KS, Karkowski LM, Prescott CA. Causal relationship between stressful life events and the onset of major depression. *Am J Psychiatry*. 1999;156(6):837-841. [FREE Full text] [doi: [10.1176/ajp.156.6.837](https://doi.org/10.1176/ajp.156.6.837)] [Medline: [10360120](https://pubmed.ncbi.nlm.nih.gov/10360120/)]
2. Chiang JJ, Turiano NA, Mroczek DK, Miller GE. Affective reactivity to daily stress and 20-year mortality risk in adults with chronic illness: findings from the national study of daily experiences. *Health Psychol*. 2018;37(2):170-178. [FREE Full text] [doi: [10.1037/hea0000567](https://doi.org/10.1037/hea0000567)] [Medline: [29154603](https://pubmed.ncbi.nlm.nih.gov/29154603/)]
3. Leger KA, Charles ST, Almeida DM. Let it go: lingering negative affect in response to daily stressors is associated with physical health years later. *Psychol Sci*. 2018;29(8):1283-1290. [FREE Full text] [doi: [10.1177/0956797618763097](https://doi.org/10.1177/0956797618763097)] [Medline: [29553880](https://pubmed.ncbi.nlm.nih.gov/29553880/)]
4. Jorm AF. Mental health literacy: empowering the community to take action for better mental health. *Am Psychol*. 2012;67(3):231-243. [doi: [10.1037/a0025957](https://doi.org/10.1037/a0025957)] [Medline: [22040221](https://pubmed.ncbi.nlm.nih.gov/22040221/)]
5. Mauss IB, Cook CL, Cheng JYJ, Gross JJ. Individual differences in cognitive reappraisal: experiential and physiological responses to an anger provocation. *Int J Psychophysiol*. 2007;66(2):116-124. [doi: [10.1016/j.ijpsycho.2007.03.017](https://doi.org/10.1016/j.ijpsycho.2007.03.017)] [Medline: [17543404](https://pubmed.ncbi.nlm.nih.gov/17543404/)]
6. Jordan AH, Monin B, Dweck CS, Lovett BJ, John OP, Gross JJ. Misery has more company than people think: underestimating the prevalence of others' negative emotions. *Pers Soc Psychol Bull*. 2011;37(1):120-135. [FREE Full text] [doi: [10.1177/0146167210390822](https://doi.org/10.1177/0146167210390822)] [Medline: [21177878](https://pubmed.ncbi.nlm.nih.gov/21177878/)]
7. Lane RD, Smith R. Levels of emotional awareness: theory and measurement of a socio-emotional skill. *J Intell*. 2021;9(3):42. [FREE Full text] [doi: [10.3390/jintelligence9030042](https://doi.org/10.3390/jintelligence9030042)] [Medline: [34449662](https://pubmed.ncbi.nlm.nih.gov/34449662/)]
8. el Kaliouby R, Picard R, Baron-Cohen S. Affective computing and autism. *Ann N Y Acad Sci*. 2006;1093:228-248. [doi: [10.1196/annals.1382.016](https://doi.org/10.1196/annals.1382.016)] [Medline: [17312261](https://pubmed.ncbi.nlm.nih.gov/17312261/)]
9. D'Alfonso S, Lederman R, Bucci S, Berry K. The digital therapeutic alliance and human-computer interaction. *JMIR Ment Health*. 2020;7(12):e21895. [FREE Full text] [doi: [10.2196/21895](https://doi.org/10.2196/21895)] [Medline: [33372897](https://pubmed.ncbi.nlm.nih.gov/33372897/)]
10. Washington P, Wall DP. A review of and roadmap for data science and machine learning for the neuropsychiatric phenotype of autism. *Annu Rev Biomed Data Sci*. 2023;6:211-228. [FREE Full text] [doi: [10.1146/annurev-biodatasci-020722-125454](https://doi.org/10.1146/annurev-biodatasci-020722-125454)] [Medline: [37137169](https://pubmed.ncbi.nlm.nih.gov/37137169/)]
11. Washington P, Park N, Srivastava P, Voss C, Kline A, Varma M, et al. Data-driven diagnostics and the potential of mobile artificial intelligence for digital therapeutic phenotyping in computational psychiatry. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2020;5(8):759-769. [FREE Full text] [doi: [10.1016/j.bpsc.2019.11.015](https://doi.org/10.1016/j.bpsc.2019.11.015)] [Medline: [32085921](https://pubmed.ncbi.nlm.nih.gov/32085921/)]
12. Voss C, Schwartz J, Daniels J, Kline A, Haber N, Washington P, et al. Effect of wearable digital intervention for improving socialization in children with autism spectrum disorder: a randomized clinical trial. *JAMA Pediatr*. 2019;173(5):446-454. [FREE Full text] [doi: [10.1001/jamapediatrics.2019.0285](https://doi.org/10.1001/jamapediatrics.2019.0285)] [Medline: [30907929](https://pubmed.ncbi.nlm.nih.gov/30907929/)]
13. Washington P, Voss C, Kline A, Haber N, Daniels J, Fazel A, et al. SuperpowerGlass: a wearable aid for the at-home therapy of children with autism. *Proc ACM Interact Mob Wearable Ubiquitous Technol*. 2017;1(3):1-22. [doi: [10.1145/3130977](https://doi.org/10.1145/3130977)]
14. Rainville P, Bechara A, Naqvi N, Damasio AR. Basic emotions are associated with distinct patterns of cardiorespiratory activity. *Int J Psychophysiol*. 2006;61(1):5-18. [doi: [10.1016/j.ijpsycho.2005.10.024](https://doi.org/10.1016/j.ijpsycho.2005.10.024)] [Medline: [16439033](https://pubmed.ncbi.nlm.nih.gov/16439033/)]
15. Nummenmaa L, Glerean E, Hari R, Hietanen JK. Bodily maps of emotions. *Proc Natl Acad Sci U S A*. 2014;111(2):646-651. [FREE Full text] [doi: [10.1073/pnas.1321664111](https://doi.org/10.1073/pnas.1321664111)] [Medline: [24379370](https://pubmed.ncbi.nlm.nih.gov/24379370/)]
16. Jang EH, Park BJ, Park MS, Kim SH, Sohn JH. Analysis of physiological signals for recognition of boredom, pain, and surprise emotions. *J Physiol Anthropol*. 2015;34(1):25. [FREE Full text] [doi: [10.1186/s40101-015-0063-5](https://doi.org/10.1186/s40101-015-0063-5)] [Medline: [26084816](https://pubmed.ncbi.nlm.nih.gov/26084816/)]
17. Domínguez-Jiménez JA, Campo-Landines KC, Martínez-Santos JC, Delahoz EJ, Contreras-Ortiz SH. A machine learning model for emotion recognition from physiological signals. *Biomed Signal Process Control*. 2020;55:101646. [doi: [10.1016/j.bpsc.2019.101646](https://doi.org/10.1016/j.bpsc.2019.101646)]

18. Schmidt P, Reiss A, Duerichen R, Marberger C, Van Laerhoven K. Introducing WESAD, a multimodal dataset for wearable stress and affect detection. 2018. Presented at: ICMI '18: Proceedings of the 20th ACM International Conference on Multimodal Interaction; October 16-20, 2018;400-408; Boulder, CO. [doi: [10.1145/3242969.3242985](https://doi.org/10.1145/3242969.3242985)]
19. He C, Yao YJ, Ye XS. An emotion recognition system based on physiological signals obtained by wearable sensors. In: Yang C, Virk GS, Yang H, editors. *Wearable Sensors and Robots: Proceedings of International Conference on Wearable Sensors and Robots 2015*. Singapore. Springer; 2017;15-25.
20. Ramzan M, Dawn S. Fused CNN-LSTM deep learning emotion recognition model using electroencephalography signals. *Int J Neurosci*. 2023;133(6):587-597. [doi: [10.1080/00207454.2021.1941947](https://doi.org/10.1080/00207454.2021.1941947)] [Medline: [34121598](https://pubmed.ncbi.nlm.nih.gov/34121598/)]
21. Vijayakumar S, Flynn R, Murray N. A comparative study of machine learning techniques for emotion recognition from peripheral physiological signals. 2020. Presented at: 2020 31st Irish Signals and Systems Conference (ISSC); June 11-12, 2020; Letterkenny, Ireland. [doi: [10.1109/issc49989.2020.9180193](https://doi.org/10.1109/issc49989.2020.9180193)]
22. Huynh L, Nguyen T, Nguyen T, Pirttikangas S, Siirtola P. StressNAS: affect state and stress detection using neural architecture search. 2021. Presented at: UbiComp/ISWC '21 Adjunct: Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers; September 21-26, 2021;121-125; Virtual. [doi: [10.1145/3460418.3479320](https://doi.org/10.1145/3460418.3479320)]
23. Hsieh CP, Chen YT, Beh WK, Wu AYA. Feature selection framework for XGBoost based on electrodermal activity in stress detection. 2019. Presented at: 2019 IEEE International Workshop on Signal Processing Systems (SiPS); October 20-23, 2019; Nanjing, China. [doi: [10.1109/sips47522.2019.9020321](https://doi.org/10.1109/sips47522.2019.9020321)]
24. Garg P, Santhosh J, Dengel A, Ishimaru S. Stress detection by machine learning and wearable sensors. 2021. Presented at: IUI '21 Companion: 26th International Conference on Intelligent User Interfaces - Companion; April 14-17, 2021;43-45; College Station, TX. [doi: [10.1145/3397482.3450732](https://doi.org/10.1145/3397482.3450732)]
25. Lai K, Yanushkevich SN, Shmerko VP. Intelligent stress monitoring assistant for first responders. *IEEE Access*. 2021;9:25314-25329. [FREE Full text] [doi: [10.1109/access.2021.3057578](https://doi.org/10.1109/access.2021.3057578)]
26. Siirtola P. Continuous stress detection using the sensors of commercial smartwatch. 2019. Presented at: UbiComp/ISWC '19 Adjunct: Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers; September 9-13, 2019;1198-1201; London, United Kingdom. [doi: [10.1145/3341162.3344831](https://doi.org/10.1145/3341162.3344831)]
27. Bobade P, Vani M. Stress detection with machine learning and deep learning using multimodal physiological data. 2020. Presented at: 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA); July 15-17, 2020; Coimbatore, India. [doi: [10.1109/icirca48905.2020.9183244](https://doi.org/10.1109/icirca48905.2020.9183244)]
28. Kumar A, Sharma K, Sharma A. Hierarchical deep neural network for mental stress state detection using IoT based biomarkers. *Pattern Recognit Lett*. 2021;145:81-87. [doi: [10.1016/j.patrec.2021.01.030](https://doi.org/10.1016/j.patrec.2021.01.030)]
29. Schmidt P, Reiss A, Dürichen R, Van Laerhoven K. Wearable-based affect recognition: a review. *Sensors (Basel)*. 2019;19(19):4079. [FREE Full text] [doi: [10.3390/s19194079](https://doi.org/10.3390/s19194079)] [Medline: [31547220](https://pubmed.ncbi.nlm.nih.gov/31547220/)]
30. Sah RK, Ghasemzadeh H. Stress classification and personalization: getting the most out of the least. *ArXiv*. Preprint posted online on July 12 2021. [doi: [10.48550/arXiv.2107.05666](https://doi.org/10.48550/arXiv.2107.05666)]
31. Indikawati FI, Winiarti S. Stress detection from multimodal wearable sensor data. *IOP Conf Ser Mater Sci Eng*. 2020;771(1):012028. [FREE Full text] [doi: [10.1088/1757-899X/771/1/012028](https://doi.org/10.1088/1757-899X/771/1/012028)]
32. Liu JC, Goetz J, Sen S, Tewari A. Learning from others without sacrificing privacy: simulation comparing centralized and federated machine learning on mobile health data. *JMIR Mhealth Uhealth*. 2021;9(3):e23728. [FREE Full text] [doi: [10.2196/23728](https://doi.org/10.2196/23728)] [Medline: [33783362](https://pubmed.ncbi.nlm.nih.gov/33783362/)]
33. Ahmad Z, Khan N. A survey on physiological signal-based emotion recognition. *Bioengineering (Basel)*. 2022;9(11):688. [FREE Full text] [doi: [10.3390/bioengineering9110688](https://doi.org/10.3390/bioengineering9110688)] [Medline: [36421089](https://pubmed.ncbi.nlm.nih.gov/36421089/)]
34. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III*. Cham, Switzerland. Springer; 2015;234-241.
35. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: an imperative style, high-performance deep learning library. 2019. Presented at: 33rd Conference on Neural Information Processing Systems (NeurIPS 2019); December 8–14, 2019; Vancouver, BC.
36. Albaladejo-González M, Ruipérez-Valiente JA, Gómez Mármol F. Evaluating different configurations of machine learning models and their transfer learning capabilities for stress detection using heart rate. *J Ambient Intell Humaniz Comput*. 2023;14(8):11011-11021. [FREE Full text] [doi: [10.1007/s12652-022-04365-z](https://doi.org/10.1007/s12652-022-04365-z)]
37. Koldijk S, Sappelli M, Verberne S, Neerincx MA, Kraaij W. The SWELL knowledge work dataset for stress and user modeling research. 2014. Presented at: ICMI '14: Proceedings of the 16th International Conference on Multimodal Interaction; November 12-16, 2014;291-298; Istanbul, Turkey. [doi: [10.1145/2663204.2663257](https://doi.org/10.1145/2663204.2663257)]
38. Ghosh S, Kim S, Ijaz MF, Singh PK, Mahmud M. Classification of mental stress from wearable physiological sensors using image-encoding-based deep neural network. *Biosensors (Basel)*. 2022;12(12):1153. [FREE Full text] [doi: [10.3390/bios12121153](https://doi.org/10.3390/bios12121153)] [Medline: [36551120](https://pubmed.ncbi.nlm.nih.gov/36551120/)]

39. Bajpai D, He L. Evaluating KNN performance on WESAD dataset. 2020. Presented at: 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN); September 25-26, 2020; Bhimtal, India. [doi: [10.1109/cicn49253.2020.9242568](https://doi.org/10.1109/cicn49253.2020.9242568)]
40. Almadhor A, Sampedor GA, Abisado M, Abbas S, Kim YJ, Khan MA, et al. Wrist-based electrodermal activity monitoring for stress detection using federated learning. *Sensors (Basel)*. 2023;23(8):3984. [FREE Full text] [doi: [10.3390/s23083984](https://doi.org/10.3390/s23083984)] [Medline: [37112323](https://pubmed.ncbi.nlm.nih.gov/37112323/)]
41. Behinaein B, Bhatti A, Rodenburg D, Hungler P, Etemad A. A transformer architecture for stress detection from ECG. 2021. Presented at: ISWC '21: Proceedings of the 2021 ACM International Symposium on Wearable Computers; September 21-26, 2021;132-134; Virtual. [doi: [10.1145/3460421.3480427](https://doi.org/10.1145/3460421.3480427)]
42. Nkurikiyeyezu K, Yokokubo A, Lopez G. The effect of person-specific biometrics in improving generic stress predictive models. *ArXiv*. Preprint posted online on December 31 2019. [doi: [10.48550/arXiv.1910.01770](https://doi.org/10.48550/arXiv.1910.01770)]
43. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. *ArXiv*. Preprint posted online on May 24 2019. [doi: [10.48550/arXiv.1810.04805](https://doi.org/10.48550/arXiv.1810.04805)]
44. Khan N, Sarkar N. Semi-supervised generative adversarial network for stress detection using partially labeled physiological data. *ArXiv*. Preprint posted online on October 27 2022. [doi: [10.48550/arXiv.2206.14976](https://doi.org/10.48550/arXiv.2206.14976)]
45. Islam T, Washington P. Personalized prediction of recurrent stress events using self-supervised learning on multimodal time-series data. *ArXiv*. Preprint posted online on July 07 2023. [doi: [10.48550/arXiv.2307.03337](https://doi.org/10.48550/arXiv.2307.03337)]
46. Koelstra S, Muhl C, Soleymani M, Lee JS, Yazdani A, Ebrahimi T, et al. DEAP: a database for emotion analysis using physiological signals. *IEEE Trans Affect Comput*. 2012;3(1):18-31. [doi: [10.1109/t-affc.2011.15](https://doi.org/10.1109/t-affc.2011.15)]
47. Markova V, Ganchev T, Kalinkov K. CLAS: a database for cognitive load, affect and stress recognition. 2019. Presented at: 2019 International Conference on Biomedical Innovations and Applications (BIA); November 8-9, 2019; Varna, Bulgaria. [doi: [10.1109/bia48344.2019.8967457](https://doi.org/10.1109/bia48344.2019.8967457)]
48. Park CY, Cha N, Kang S, Kim A, Khandoker AH, Hadjileontiadis L, et al. K-EmoCon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations. *Sci Data*. 2020;7(1):293. [FREE Full text] [doi: [10.1038/s41597-020-00630-y](https://doi.org/10.1038/s41597-020-00630-y)] [Medline: [32901038](https://pubmed.ncbi.nlm.nih.gov/32901038/)]
49. Daniels J, Schwartz JN, Voss C, Haber N, Fazel A, Kline A, et al. Exploratory study examining the at-home feasibility of a wearable tool for social-affective learning in children with autism. *NPJ Digit Med*. 2018;1(1):32. [FREE Full text] [doi: [10.1038/s41746-018-0035-3](https://doi.org/10.1038/s41746-018-0035-3)] [Medline: [31304314](https://pubmed.ncbi.nlm.nih.gov/31304314/)]
50. Daniels J, Haber N, Voss C, Schwartz J, Tamura S, Fazel A, et al. Feasibility testing of a wearable behavioral aid for social learning in children with autism. *Appl Clin Inform*. 2018;9(1):129-140. [FREE Full text] [doi: [10.1055/s-0038-1626727](https://doi.org/10.1055/s-0038-1626727)] [Medline: [29466819](https://pubmed.ncbi.nlm.nih.gov/29466819/)]
51. Kalantarian H, Jedoui K, Washington P, Tariq Q, Dunlap K, Schwartz J, et al. Labeling images with facial emotion and the potential for pediatric healthcare. *Artif Intell Med*. 2019;98:77-86. [FREE Full text] [doi: [10.1016/j.artmed.2019.06.004](https://doi.org/10.1016/j.artmed.2019.06.004)] [Medline: [31521254](https://pubmed.ncbi.nlm.nih.gov/31521254/)]
52. Kalantarian H, Jedoui K, Washington P, Wall DP. A mobile game for automatic emotion-labeling of images. *IEEE Trans Games*. 2020;12(2):213-218. [FREE Full text] [doi: [10.1109/tg.2018.2877325](https://doi.org/10.1109/tg.2018.2877325)] [Medline: [32551410](https://pubmed.ncbi.nlm.nih.gov/32551410/)]

## Abbreviations

- BERT:** Bidirectional Encoder Representations from Transformers
- CLAS:** Cognitive Load, Affect, and Stress
- DEAP:** Database for Emotion Analysis using Physiological Signals
- ECG:** electrocardiography
- EDA:** electrodermal activity
- EMG:** electromyogram
- LOSO:** leave-one-subject-out
- PPG:** photoplethysmography
- SiLU:** sigmoid linear unit
- SWELL:** Smart Reasoning for Well-being at Home and at Work
- SWELL-KW:** SWELL knowledge work
- WESAD:** Wearable Stress and Affect Dataset

*Edited by K El Emam, B Malin; submitted 25.08.23; peer-reviewed by S Pandey, M Zhou, G Vos; comments to author 19.09.23; revised version received 19.02.24; accepted 23.03.24; published 10.05.24*

*Please cite as:*

*Li J, Washington P*

*A Comparison of Personalized and Generalized Approaches to Emotion Recognition Using Consumer Wearable Devices: Machine Learning Study*

*JMIR AI 2024;3:e52171*

*URL: <https://ai.jmir.org/2024/1/e52171>*

*doi: [10.2196/52171](https://doi.org/10.2196/52171)*

*PMID:*

©Joe Li, Peter Washington. Originally published in JMIR AI (<https://ai.jmir.org>), 10.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR AI, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.ai.jmir.org/>, as well as this copyright and license information must be included.