egoEMOTION: Egocentric Vision and Physiological Signals for Emotion and Personality Recognition in Real-World Tasks

Matthias Jammot*, **Björn Braun***, **Paul Streli**, **Rafael Wampfler**, **Christian Holz**
Department of Computer Science
ETH Zurich, Switzerland

https://siplab.org/projects/egoEMOTION

Abstract

Understanding affect is central to anticipating human behavior, yet current egocentric vision benchmarks largely ignore the person's emotional states that shape their decisions and actions. Existing tasks in egocentric perception focus on physical activities, hand-object interactions, and attention modeling—assuming neutral affect and uniform personality. This limits the ability of vision systems to capture key internal drivers of behavior. In this paper, we present *egoEMOTION*, the first dataset that couples egocentric visual and physiological signals with dense self-reports of emotion and personality across controlled and real-world scenarios. Our dataset includes over 50 hours of recordings from 43 participants, captured using Meta's Project Aria glasses. Each session provides synchronized eye-tracking video, head-mounted photoplethysmography, inertial motion data, and physiological baselines for reference. Participants completed emotion-elicitation tasks and naturalistic activities while self-reporting their affective state using the Circumplex Model and Mikels' Wheel as well as their personality via the Big Five model. We define three benchmark tasks: (1) continuous affect classification (valence, arousal, dominance); (2) discrete emotion classification; and (3) trait-level personality inference. We show that a classical learning-based method, as a simple baseline in real-world affect prediction, produces better estimates from signals captured on egocentric vision systems than processing physiological signals. Our dataset establishes emotion and personality as core dimensions in egocentric perception and opens new directions in affect-driven modeling of behavior, intent, and interaction.

1 Introduction

Egocentric vision systems are well positioned to capture the signals for modeling human attention, interaction, and behavior in real-world environments. Benchmarks in this area have driven advances in action recognition [15], object manipulation [31, 48], gaze prediction [25], and interaction understanding [17, 18]. These tasks focus on what people do and attend to, using first-person visual input to model external behavior [17, 18, 38]. Such progress has expanded the scope of perception systems, in domains such as Mixed Reality [23, 40], front-line productivity work [9], and context-aware interaction [5, 19]. However, current benchmarks overlook internal states like emotion and personality that shape these behaviors, implicitly assuming affect-neutral and behaviorally uniform participants [18], ignoring individual differences. This limits how egocentric systems can model behavior that depends on mood, arousal, or personality traits [17]. Tasks involving decisions [45], social interaction [17], and memory [47] require grounding in affect. We argue that without such affective modeling, emerging egocentric platforms cannot fully understand human behavior.

*Equal contribution

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

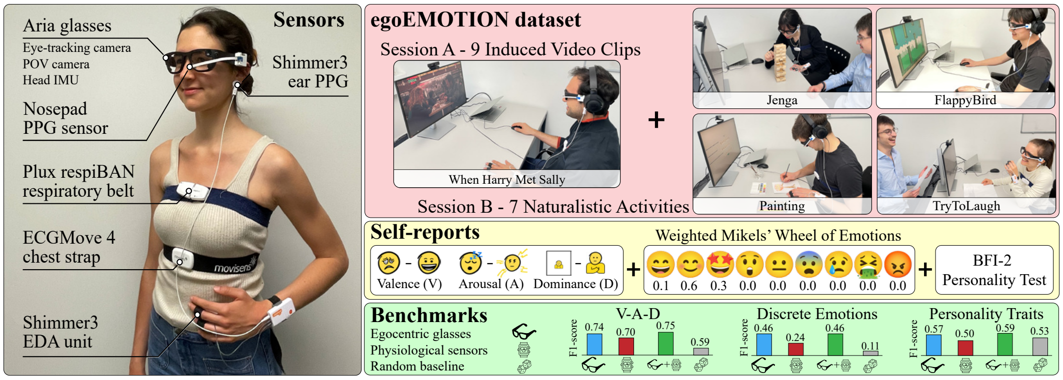


Figure 1: **egoEMOTION is a multimodal emotion and personality recognition dataset that captures participants' facial, eye-tracking, egocentric, and physiological signals during induced video stimuli and naturalistic real-world activities**. Participants reported their emotions via emoti-SAM [20] and a weighted Mikels' Wheel [41], and their personality using the Big Five model [10].

In this paper, we present egoEMOTION, a dataset for affect and personality recognition from egocentric visual and physiological signals. Our dataset thus addresses the current gap in egocentric vision by supplying emotional and trait labels grounded in self-reports. egoEMOTION comprises synchronized multimodal data during both emotion-elicitation protocols and naturalistic tasks, such as watching video clips, painting, playing social and video games. Each session captures a participant's eye-tracking video, inertial motion (IMU), outward point-of-view (POV) camera, and photoplethysmogram (PPG) to gauge cardiac activity from Meta's Project Aria glasses [13], as well as physiological baseline measurements, including electrocardiograms (ECG), respiratory rates (RSP), and electrodermal activity (EDA)—all suitable to extract indicators of a person's affective state [42, 62]. Participants reported their affect using the Circumplex Model [54] and Mikels' Wheel [41] and assessed their personality using the Big Five model [10]. In total, our dataset spans 50 hours of recordings from 43 participants across varied emotional and social contexts.

We then define three prediction benchmarks—continuous affect regression, discrete emotion classification, and personality inference— and provide baselines showing that egocentric signals, particularly eye-tracking features, outperform traditional physiological baselines in real-world emotion prediction. This highlights the promise of affective modeling from egocentric vision systems and establishes egoEMOTION as a foundation for future research in this direction.

Collectively, we contribute:

1. the first multimodal dataset that uses an egocentric vision system for emotion and personality recognition. Our dataset comprises both induced and naturalistic tasks that cover a wide range of elicited emotions, while offering nuanced mixed-emotions self-reporting.
2. three benchmark tasks and associated baseline: valence-arousal-dominance, discrete emotion, and personality recognition. Our results show that using features solely from egocentric vision systems outperforms estimates from physiological signals.
3. an open-source release of our ethics-approved dataset and baseline implementations (23 `ETHICS-008`).

2 Related Work

Emotion elicitation can be either induced, using predefined stimuli like videos or sounds, or naturalistic, arising spontaneously in real-life contexts. The terms *in-the-wild*, *real-world*, or *naturalistic* data have been denoted to describe data collection when the experimenters do not control the emotion elicitation nor constrain the data acquisition [33]. These emotional responses may occur in either static environments, where participants remain still (workplace, car, cinema), or ambulatory environments, where data is collected during everyday activities [33].

Induced. Due to the challenges of collecting physiological data in real-world settings [43, 51, 59], many emotion recognition studies have been conducted in controlled laboratory settings and have

Table 1: **Comparison of public multimodal affective datasets.**

			Elicitation				**Sensing Modalities**										**Annotation**			
Dataset (year)	No. Subjects	Mobile Sensors	Induced	Natural	Individual	Social	BVP	ECG	EDA	EEG	Eye Tracking	Face	IMU (head)	IMU (wrist)	POV camera	RSP	A-V (-D)	Emotion Tags	Weighted Tags	Big-5
egoEMOTION (2025)	43	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
EmoPairCompete [12] (2024)	28	✓		✓		✓	✓		✓					✓				✓	✓	
G-REx [3] (2024)	191	✓		✓		✓	✓		✓								✓			
eSEE-d [58] (2023)	48		✓		✓						✓							✓	✓	
BIRAFFE2 [29] (2022)	102	✓		✓	✓			✓	✓					✓			✓			✓
PPB-Emo [36] (2022)	40			✓	✓			✓		✓		✓					✓	✓		
K-EMOCON [49] (2020)	21	✓		✓		✓	✓	✓	✓	✓		✓		✓			✓	✓		
AMIGOS [42] (2018)	40	✓	✓		✓	✓		✓	✓	✓		✓					✓	✓		✓
ASCERTAIN [62] (2016)	58	✓	✓		✓			✓	✓	✓		✓					✓			✓
DEAP [26] (2012)	32		✓		✓		✓		✓	✓						✓	✓			
MAHNOB-HCI [60] (2012)	27		✓		✓			✓	✓	✓	✓	✓		✓		✓	✓	✓		

Datasets where participants were shown videos are classified as 'induced' elicitation.
A: Arousal, V: Valence, D: Dominance. BVP: Blood Volume Pressure (from PPG sensor).

used pre-selected video clips as emotional stimuli, as shown in Table 1. DEAP [26] collected electroencephalogram (EEG), facial, and physiological data from 32 participants who self-reported their emotions using valence-arousal-dominance (V-A-D) ratings after viewing 40 1-minute-long music videos. MAHNHOB-HCI [60] collected signals similar to those of DEAP with the addition of audio and eye gaze data. Their study gathered 27 participants who, after watching 20 short videos in a first experiment, followed by 28 images and 14 videos in a second experiment, annotated their emotions using V-A-D rating scales and emotional tags. ASCERTAIN [62] extended these studies by using wireless physiological sensors and facial features from 58 participants, while also capturing personality traits through the Big Five model [10]. AMIGOS [42] further advanced the field by introducing group-based video viewing and assessing mood in parallel to emotions and personality.

Naturalistic. While controlled lab settings are useful for isolating variables and evaluating specific emotions, their ecological validity is limited, raising concerns about real-world applicability [33, 57]. The G-REx [3], EmoPairCompete [12], and K-EmoCon [49] datasets naturally induced emotions in participants through group movie sessions, solving puzzles in pairs, and paired debates, respectively. BIRAFFE2 [29] exposed its participants to IAPS [32] visual and IADS [4] audio stimuli, followed by three mini-quest video games. PPB-Emo [36] recorded participants in a driving simulator. While these datasets have advanced emotion recognition in static real-world tasks, they remain limited in the array of sensors used and their range of emotionally-diverse activities. To vary naturalistic emotions recorded, emotion recognition has been investigated in ambulatory real-world settings, using mobile phones to self-report emotions [27, 56, 57, 64] and personality [28]. However, these in-the-wild studies face key limitations: self-reports are often infrequent and intrusive, the lack of known stimuli hinders interpretation of physiological responses, and signal quality is affected by motion artifacts and inconsistent sensor use.

Egocentric. The rise of mobile egocentric systems has enabled large-scale [21], in-the-wild datasets such as EPIC-KITCHENS [11], Ego4D [17], Ego-Exo4D [18], and Nymeria [38], supporting tasks like activity recognition and social behavior modeling. However, these datasets assume neutral affect and lack emotional context, limiting their use for modeling user intent or emotion-driven behaviors. Integrating emotion recognition could enable more affect-aware activity analysis and adaptive human-AI interaction. In this context, the eye-tracking videos recorded with our glasses offer a valuable modality for capturing users' intrinsic emotional states during diverse real-world tasks. MAHNOB-HCI [60] was one of the first datasets to introduce eye-tracking as a modality. While emotion recognition using *mobile* eye-tracking systems has been explored [30, 46, 63], their datasets were not made public and gathered few participants. To date, eSEE-d [58] is the only public dataset

Table 2: **Summary of emotional elicitation tasks**: induced (`1-9`) and naturalistic (`10-16`).

Activity	ID	Description	Duration
*Video Clips**	`1-9`	AnimalCruelty, AuroraBorealis, BearGrylls, CollegeAcceptance, HarrySally, JoJoRabbit, LoveActually, MovingShapes, Psycho	9×48 s
Flappy Bird	`10`	Click to keep a bird flying through pipes. Restart upon failure.	4 min
Jelly Bean	`11`	Eat three unpleasant-tasting jelly beans.	2 min
Jenga	`12`	Remove blocks from a tower without collapsing it with experimenter.	5 min
Painting	`13`	Paint with brushes and crayons, listening to *Your Song* (Elton John).	4 min
Sad Letter	`14`	Write a letter to someone lost, listening to *Adagio for Strings*.	4 min
Slenderman	`15`	Find eight pages in dark woods while escaping the Slenderman.	6 min
Try to Laugh	`16`	Take turns with experimenter telling pre-written jokes.	4 min

*A detailed description of the emotion-inducing video clips is presented Table 12 of Appendix A.

for emotion recognition using mobile eye-tracking. However, its limited four-emotion questionnaire, absence of physiological signals, and controlled setup (e.g., chin rest) reduce its validity for real-world applications. While heart rate can be estimated from facial videos [8, 52, 68], other physiological signals, such as EDA, remain challenging to estimate [6, 7] but are significant for judging a person's emotional response [37]. Personality recognition from mobile eye-tracking systems has been explored by Hoppe et al. [22] with participants walking on a university campus and Berkovsky et al. [2], in which participants watched images from the IAPS dataset [32] in laboratory settings. Neither of these studies recorded physiological signals, nor released their dataset.

3 egoEMOTION Dataset

Prior work on emotion and personality recognition using physiological sensors has typically focused on affect, valence, and personality recognition—often in controlled lab settings with specialized equipment. We go further by collecting detailed emotion self-reports alongside affect, valence, and personality, across both induced and naturalistic tasks (see Figure 1). While our setup includes both standard physiological sensors and an egocentric vision system, we show that egocentric video alone is sufficient to enable practical, real-world applicability beyond traditional sensor-based approaches.

3.1 Dataset Design

3.1.1 Experimental Protocol

Upon arrival, we explained the study protocol and self-report questionnaires to the participants, asked them to sign a consent form, and then equipped them with the sensors. The experimental protocol (see Figure 1) consisted of two sessions (A and B) with a total of 16 different tasks. We conducted the experiment in a regular office next to a window, with the experimenter seated behind a curtain to avoid affecting participants' emotional reactions. Before starting session A, each participant performed an eye-tracking calibration. In session A, participants watched nine video clips ($\mu = 48$ s, see Table 2) corresponding to the eight emotions from Mikels' Wheel [41], plus a ninth neutral emotion. All videos were extensively validated by previous work to elicit target emotions [60, 62]. Before each video clip, participants had to watch a 5-second video of a fixation cross to refocus their gaze [42, 62]. In session B, participants conducted seven activities (see Table 2) that we selected to reflect spontaneous everyday activities to further the study's ecological validity [53]. We designed the activities to minimize physical effort to avoid activity-induced variations in the recorded signals. After each task in sessions A and B, participants self-reported their perceived emotions. They were instructed to report their true emotion, not the one they perceived as being the 'correct' one. The questionnaires completed, they watched a neutral video of clouds to mitigate any carry-over effect of the previous emotional stimulus. The task order in both sessions was randomized for each participant.

3.1.2 Data Annotation

The experiment was performed on a graphical user interface coded using the `PyQt5` Python library, which would successively show washout, emotional stimulus, and self-report. For session B, the

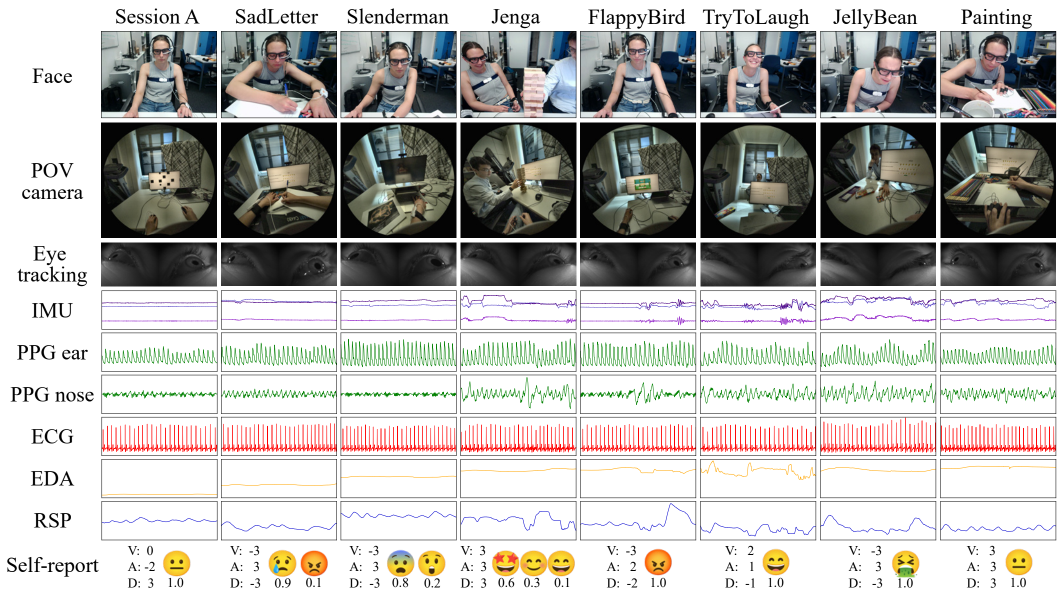


Figure 2: **Data collection** from the egocentric glasses and physiological sensors during each task with their associated self-reports. Further information about the study protocol is available in Appendix A.

experimenter would explain and setup the upcoming activity once the washout video was completed. Participants used their dominant hand to navigate through two questionnaires using a serial mouse. The first questionnaire, built on the Circumplex Model of Affect [54], comprised a self-assessment manikin (SAM) where participants reported the arousal (A), valence (V), and dominance (D) of their emotion. We used the 7-point emoji-based emoti-SAM [20], which balanced response granularity with cognitive load and was intuitive given the ubiquity of emojis. If SAM ratings are standard in emotion recognition studies, they are usually reduced to binary labels (e.g., *high* / *low*), which oversimplifies emotions. As exemplified in Figure 2, the distinct emotions of fear, sadness, disgust, and anger all fall into the low-valence / high-arousal quadrant, making them difficult to distinguish. Some studies have introduced binary emotional tags to address this [16, 42, 60], but these lack nuance, particularly when mixed emotions are present, since each emotion carries equal weight in the analysis.

To gain nuance, we used as second questionnaire a weighted version of emotional tags. The participants distributed 100% across nine emotions (eight from Mikels' wheel [41] plus a neutral option) in increments of 10%, ensuring the weights sum to unity for every report. The emotions were Amusement (Amu), Content (Con), Excitement (Exc), Awe, Fear (Fea), Sadness (Sad), Disgust (Dis), Anger (Ang), and Neutral (Neu). This captured the relative emotional strength perceived by the user, distinguishing between stimuli that have one dominating emotion and others where emotions are more homogeneous. This annotation allowed complex emotions to be represented as vectors with attributes such as polarity, type, intensity, similarity, and additivity, following Yang et al. [67]. It also enabled identifying the dominant emotion, leading to a more precise 9-class classification.

Finally, participants filled in the Big Five Inventory-2 (BFI-2) personality questionnaire online [61] before the experiment. The BFI-2 assesses five major personality traits: Extraversion (*Ex*), Agreeableness (*Ag*), Conscientiousness (*Co*), Negative Emotionality (*NE*), and Open-Mindedness (*OM*).

3.1.3 Sensors

Our study used mobile wearable sensors to capture participants emotional responses, as shown in Figure 1. We used the Project Aria glasses [13] for their significant promise in capturing ecologically valid data that does not inhibit natural activities and behavior of the participants. Using the device's 'Profile 16', we recorded eye-tracking (ET) videos with a 640 $\times$ 480 pixel resolution per eye at 90 fps, egocentric vision through a 1408 $\times$ 1408 POV RGB camera at 10 fps, and head movements through two IMUs sampling at 1000 Hz and 800 Hz. We supplemented the egocentric glasses with an in-house nosepad PPG sensor sampling at 128 Hz. A Shimmer3 unit recorded PPG and EDA signals at the ear and fingers, respectively, at 256 Hz. A 1024 Hz Movisens ECG4Move4 chest belt measured the participant's ECG data while a plux respiBAN respiratory belt measured their

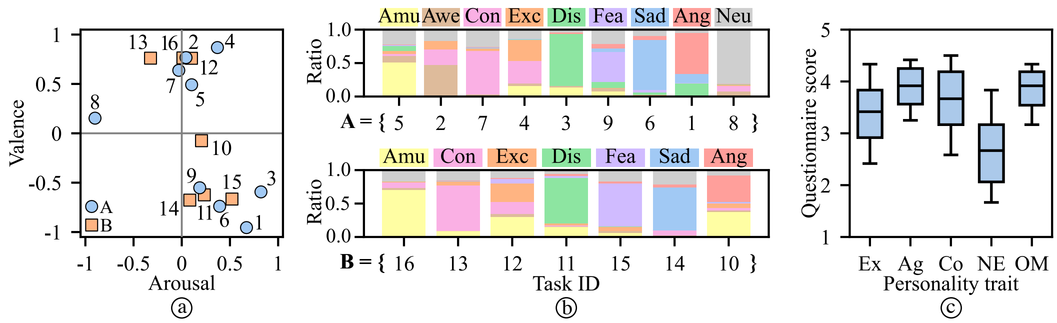


Figure 3: **Participant self-reports across tasks**. **(a)** Mean arousal-valence ratings. **(b)** Proportions of discrete emotions reported in Sessions A and B. **(c)** Boxplots of Big Five personality trait scores.

respiration pattern (RSP) at 400 Hz. We recorded participants' facial expressions with a 60 fps 1280×720 webcam for external labeling of emotions.

3.2 Recruitment and Recording

We recruited 43 healthy participants, mostly students, voluntarily with a CHF 30 compensation. The 24 female and 19 male participants were between 19 and 29 years old (μ = 26, σ = 2). Based on the Fitzpatrick scale [14], 3 participants had skin type I, 19 had skin type II, 9 had skin type III, 8 had skin type IV, and 5 had skin type V. Each participant was recorded in a single session that lasted approximately 105 minutes. They had to confirm that they were not taking tranquilizers, psychotropic drugs, or narcotics, and were not diagnosed with any cardiovascular disease. They were also informed that they would have to carry two belts (ECG and RSP) on their chest. We ensured that all participants had sufficient English proficiency to understand the videos that they were shown.

3.3 Dataset Composition

Participant recordings were cut to the duration of session A (μ = 20 min) and session B (μ = 49 min). The dataset contains all raw sensor streams presented in Figure 2, each preserved at its sampling frequency. Since all sensors were equipped with IMUs, they were synchronized at the start and end of each experiment by simultaneously shaking them. This yields egoEMOTION, a dataset composed of 43 participant recordings, each completing 16 emotional tasks. In total, the dataset offers over 50 hours of synchronized (90 Hz) multimodal data of egocentric and physiological signals. The dataset is structured by participant, with each folder containing the corresponding sensor streams. The start and end of each task were manually labelled to enable task-specific analyses.

4 Dataset Descriptives

4.1 Analysis of self-reports

Figure 3 shows the self-reports across all participants of (a) mean arousal-valence ratings for each video clip and task, (b) discrete emotions (dominant emotion indicated above task), and (c) personality traits. While participants reported a wide range of valence, arousal ratings were less varied and consistently high across tasks. The video clips and naturalistic activities elicited a diverse range of emotions (see Figure 3b). The naturalistic activities elicited stronger emotional responses, evidenced by a lower proportion of neutral tags. More detailed information is provided in Appendix B.

4.2 Correlations

Figure 4 presents the Pearson correlation matrices between continuous affect self-ratings (A-V-D), discrete emotions, and personality traits. Focusing on significant correlations, *Co* was negatively correlated with both V and D. *Ag* showed a moderate positive correlation with A. The strongest relationships for discrete emotions and personality were between *Ag* and Sadness and Disgust. The correlations between discrete emotional states and the continuous affective dimensions were the following: V was positively associated with Amusement and Content, while negatively associated

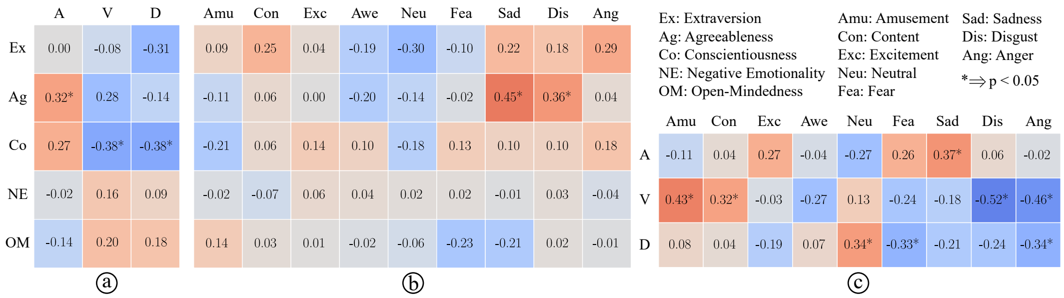


Figure 4: **Pearson correlations between self-reports.** **(a)** continuous self-ratings and personality scores **(b)** discrete emotions and personality scores **(c)** discrete emotions and continuous self-ratings.

with Disgust and Anger, aligning with expectations. D showed significant negative correlations with Disgust and Anger, while A was positively linked to Excitement and Sadness. Appendix B contains more details on the individual correlations between the self-reported annotations in Sessions A and B.

5 Baselines

To demonstrate the benefits of egoEMOTION and motivate follow-up research, we propose three benchmarking tasks: predicting a participant's self-reported *affective state*, *discrete emotions*, and *personality traits*. We evaluate classical machine learning methods using data from wearable sensors (ECG, EDA, RSP) and the Aria glasses (accelerometers, eye-tracking, nosepad PPG). We contrast these classical methods with deep learning methods to highlight potential future research directions.

5.1 Feature Extraction

We extracted a total of 612 features from all data modalities (see Appendix C.1) across the duration of each video clip in session A and each activity in session B . Following prior work, we extracted 77 features for ECG and PPG (green channel nose-pad PPG) [42], 31 features for EDA [42, 62], and 14 features for respiratory rate [26].

Pupil size was inferred over time from the eye-tracking video footage using an open-source eye-tracking algorithm [24]. We also computed the mean pixel intensity of each eye for each video frame as a basic visual descriptor. Additionally, we trained a Fisherface model (PCA followed by LDA) [1] on each training split for each target variable (affective state, emotion, and personality), and used it to project each video frame into a one-dimensional space. The resulting per-frame projections (Fisherface features) were included in our analysis. We used the open-source eye gaze extraction for the Project Aria glasses from Meta [39] to obtain the eye gaze (yaw and pitch). As there is no publicly available model yet for blinking detections from Project Aria glasses, we implemented a signal-processing-based approach using the variance map of the eye tracking videos to detect blinks [44]. To detect micro-expressions from the eye-tracking videos, we extracted features using LBP-TOP [69] with a window size of 10 frames (i.e., 111 ms) similar to previous work on facial videos [66]. For the acceleration signal from the Aria glasses, we calculated the magnitude across all three axis. All computations were run on AMD EPYC CPUs.

For each of the resulting time-series signals—pupil size, eye gaze, video pixel intensity, Fisherface features, and acceleration magnitude—we computed 15 statistical descriptors: mean, minimum, maximum, standard deviation, median, 5th percentile, 95th percentile, range, interquartile range (IQR), sum, energy, skewness, kurtosis, root mean square (RMS), and line integral. For the micro-expressions, we averaged each of the LBP-TOP features. The pupil detection and eye-tracking video preprocessing took 2 hours and about 50 GB of RAM per participant, with the micro-expressions taking 10 minutes. The rest of the feature extraction took under 1 minute per participant.

5.2 Continuous Affect Recognition

For continuous affect recognition, we focused on predicting a participant's self-reported *arousal*, *valence* and *dominance* levels. To enable classification, we binarized these continuous ratings into *low* and *high* categories using the median value across the training set, following prior work [42, 62].

Table 3: **Predictions for continuous affect ratings, discrete emotions, and personality traits.**

		Wearable devices				Egocentric glasses									All	Baseline
						ET video										
Benchmark	**Domain**	ECG	EDA	RSP	$\bowtie$	Pup.	Int.	F.f.	Gaze	Blink	μ-E.	PPG	IMU	$\bowtie$	$\bowtie$	Random
Continuous Affect	Arousal	0.76	0.76	0.75	0.76	0.77	0.76	0.76	**0.78**	0.76	0.76	0.76	0.75	**0.78**	**0.78**	0.64
	Valence	0.67	0.64	0.69	0.69	0.73	0.72	0.69	0.63	0.66	0.68	0.66	0.75	0.76	**0.77**	0.55
	Dominance	0.63	0.66	0.66	0.66	0.67	0.66	0.65	0.65	**0.69**	0.66	0.67	0.67	0.68	0.68	0.57
	Mean	0.69	0.69	0.70	0.70	0.72	0.71	0.70	0.69	0.70	0.69	0.70	0.72	0.74	**0.75**	0.59
Discrete Emotions	Amused	0.37	0.44	0.45	0.45	0.39	0.50	0.43	0.36	0.23	0.32	0.31	**0.58**	0.50	0.52	0.21
	Content	0.28	0.20	0.29	0.29	0.37	0.49	0.31	0.31	0.23	0.20	0.20	**0.54**	0.52	0.50	0.16
	Excited	0.00	0.05	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.00	0.00	**0.12**	**0.12**	0.08	0.05
	Awe	0.00	0.00	0.00	0.00	0.24	0.00	0.00	0.06	0.05	0.00	0.00	0.23	**0.31**	0.28	0.04
	Neutral	0.18	0.29	0.22	0.15	0.36	0.34	0.34	0.36	0.16	0.17	0.17	0.37	**0.41**	0.40	0.17
	Fear	0.06	0.14	0.17	0.28	0.48	0.40	0.08	0.24	0.04	0.20	0.10	0.42	0.55	**0.59**	0.08
	Sad	0.15	0.42	0.17	0.46	0.45	0.52	0.32	0.37	0.11	0.12	0.10	**0.60**	0.57	0.57	0.10
	Disgust	0.08	0.40	0.27	0.39	0.40	0.61	0.34	0.40	0.08	0.20	0.18	0.60	**0.65**	0.61	0.12
	Anger	0.03	0.05	0.11	0.11	0.26	0.17	0.17	0.12	0.09	0.03	0.00	0.48	**0.53**	0.50	0.08
	Mean	0.13	0.22	0.19	0.24	0.34	0.34	0.22	0.25	0.11	0.14	0.12	0.44	**0.46**	**0.46**	0.11
Personality Traits	Ex	0.28	0.52	0.32	0.22	0.40	**0.60**	0.50	0.58	0.43	0.48	0.30	0.55	0.55	0.48	0.55
	Ag	0.38	0.42	0.48	0.55	0.45	0.40	**0.60**	**0.60**	0.35	0.43	0.40	0.57	0.30	0.55	0.52
	Co	0.55	0.55	0.30	0.50	0.55	**0.65**	0.45	0.48	0.58	0.55	0.57	0.40	0.55	**0.65**	0.55
	NE	0.52	0.50	0.65	0.68	0.68	0.60	0.55	0.60	0.50	0.58	0.42	0.30	0.68	**0.70**	0.52
	OM	0.32	0.55	0.50	0.57	0.48	0.30	0.62	0.53	0.38	0.33	0.60	0.57	**0.70**	0.58	0.52
	Mean	0.41	0.51	0.45	0.50	0.51	0.51	0.54	0.56	0.45	0.47	0.46	0.48	0.57	**0.59**	0.53

$\bowtie$ = fusion of modalities, Pup. = Pupil size, Int. = Pixel Intensity, F.f. = Fisherface features, μ-E. = micro-expressions. The error bars are not displayed in this table for clarity purposes. They are available in Appendix C.

We trained a separate Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel (default settings [50]) for each target (arousal and valence). Features were standardized to the $[0, 1]$ range on a per-participant basis. No feature selection was applied for this task, as the SVM model was shown to perform robustly with the full set of features. Classification was performed using a leave-one-subject-out (LOSO) cross-validation strategy to ensure generalization across participants. We report the mean F1 score across all participants, averaged over the two binary classification tasks (see Table 3 and Appendix Table 6).

5.3 Discrete Emotion Recognition

In the discrete emotion recognition task, we aimed to classify one of nine basic emotions as reported by participants: *amusement*, *content*, *excitement*, *awe*, *neutral*, *fear*, *sadness*, *disgust*, and *anger*. For each participant and task, the ground-truth label corresponds to the strongest self-reported emotion. We used a Random Forest classifier with standardized features. To reduce dimensionality and focus on the most relevant inputs, we applied SelectKBest [50] feature selection using mutual information, retaining the top 10 features from the training set. As with the affect recognition task, we employed leave-one-subject-out cross-validation to assess generalization. We trained a single multi-class classifier and evaluated performance using the mean F1 score across participants (see Table 3 and Table 8). The 9-class classification task had a random baseline F1 score of 0.11.

5.4 Personality Prediction

The personality prediction task involves estimating each participant's Big Five personality traits: *open-mindedness*, *conscientiousness*, *extraversion*, *agreeableness*, and *negative emotionality*. For each trait, we binarized the self-reported score into *low* and *high* categories using the median across the training set. We trained a separate Random Forest classifier for each trait. Unlike the other tasks, we did not apply feature standardization, as the absolute magnitude of certain features was found to be more informative for personality prediction. For each classifier, we applied SelectKBest feature selection using mutual information and retained the top 10 features from the training set. Feature vectors were constructed by averaging the features for all samples belonging to a participant. We evaluated the model using leave-one-subject-out cross-validation and report the mean F1 score averaged across all five traits (see Table 3 and Appendix Table 10).

Table 4: **Performance comparison between classical and deep learning approaches.**

Benchmark	Model	Wearable devices	Egocentric glasses	All
Continuous Affect	Classical	0.70±0.14	0.74±0.13	0.75±0.13
	CNN [55]	0.63±0.05	0.68±0.05	0.68±0.07
	WER [65]	0.49±0.21	0.65±0.11	0.60±0.16
Discrete Emotions	Classical	0.24±0.08	0.46±0.18	0.46±0.17
	CNN [55]	0.12±0.01	0.23±0.03	0.22±0.02
	WER [65]	0.13±0.02	0.22±0.03	0.21±0.04
Personality Traits	Classical	0.50±0.48	0.57±0.49	0.59±0.49
	CNN [55]	0.43±0.26	0.42±0.20	0.41±0.25
	WER [65]	0.38±0.28	0.47±0.24	0.44±0.28

5.5 Use of deep learning models

To ground the above results in contemporary approaches, we implemented two deep learning-based models from previous works for wearable emotion recognition: one classical convolutional neural network (CNN) [55] and one state-of-the-art transformer-based architecture (WER) [65]. As input, we use the filtered continuous signals (without Fisherfaces), similar to previous work [55, 65]. We trained all models using a five-fold cross-validation approach. The training was conducted with a batch size of 128 for 30 epochs, a learning rate of 0.0001, and cross-entropy loss as the loss function. Each model was trained on a NVIDIA H200 with a total runtime of about 8 hours for the CNN and 30 hours for the transformer-based architecture.

For all proposed benchmark tasks, our implemented classical methods perform better than the deep learning-based approaches (see Table 4). Using the classical methods, we obtain maximum F1 scores of 0.75 (continuous affect), 0.46 (emotion prediction) and 0.59 (personality prediction) compared to maximum F1 scores of 0.68, 0.23 and 0.47 using the deep learning-based approaches, respectively.

5.6 Discussion

The results in Table 3 highlight the value of incorporating data from the Aria headset alongside traditional physiological modalities such as ECG, EDA, and RSP. For continuous affect recognition, the SVM model achieves a mean F1 score of 0.75 when using all modalities. Signals captured exclusively from the headset reach a comparable mean F1 score of 0.74, slightly outperforming traditional wearable signals ($F_1 = 0.70$), with pupil size features contributing strongly. In the more challenging discrete emotion recognition task, head-mounted signals alone yield a mean F1 score of 0.46, which is substantially above the random baseline of 0.11. Notably, acceleration magnitude from the Aria IMU achieves $F_1 = 0.44$ on its own, and pupil intensity reaches 0.34, while wearable signals perform significantly lower (e.g., $F_1 = 0.13$ for ECG, $F_1 = 0.22$ for EDA). For personality prediction, combining all modalities results in a mean F1 score of 0.59 versus 0.53 for the baseline. Signals from the egocentric glasses alone yield $F_1 = 0.57$, outperforming wearable-only inputs ($F_1 = 0.50$), with eye gaze being the best-performing individual modality.

These findings suggest that egocentric signals from head-mounted devices, such as eye-tracking video and head motion, capture rich behavioral information beyond traditional physiological sensors. While such modalities were once impractical in mobile settings, the growing availability of smartglasses and augmented reality headsets makes their use increasingly practical. Coupled with more expressive models, such as temporal neural networks or multimodal foundation models, these data sources offer promising directions for real-time user state inference and next-generation human-centered systems.

6 Limitations

While egoEMOTION offers a rich multimodal dataset for emotion and personality recognition in induced and naturalistic settings, several limitations must be acknowledged. First, the ground truth labels rely on retrospective self-reports after each task, which may be affected by recall bias and do not capture the dynamic nature of emotional responses over time [3, 34]. More fine-grained labeling (e.g., via facial expression analysis from our facial recordings) could further improve the temporal resolution of the annotations. Second, the dataset lacks longitudinal recordings, which may limit the study of emotional and personality state changes over extended periods. While our study was designed for identifying distinct emotions rather than mapping them to the arousal-valence scale in order to get finer emotion labels, we acknowledge that it has limited representation in the low-arousal, low-valence quadrant. We also recognize that some modalities like IMU may primarily capture task-related motor activity rather than affective states due to inherent coupling between behavior and emotion, which could confound emotion recognition with task classification. However, our results show IMU-based prediction performs better in Session A (identical participant behavior across emotions) than in Session B (different participant behavior across emotions), suggesting the IMU captures more than just overt behavioral differences, and is informative for emotion prediction even when behavior is held constant.

Additionally, despite recording eye-tracking and facial video data, we only extracted pupil diameter, pixel intensity, Fisherface features, gaze fixations, blink rate, and micro-expressions. Designing emotion-specific features (e.g., to recognize narrowed eyes when smiling, or teary eyes when sad) would further enhance the performance of the models. Moreover, while we leveraged end-to-end deep learning networks [35] in the CNN [55] and WER [65] models we used, we expect that incorporating pre-training on large-scale wearable physiological datasets, as well as future advances in model design and training, will improve the results. We believe our dataset will motivate future research in these directions. Finally, our participant pool was primarily composed of young adults. While this may support training stability, it introduces some demographic bias and potentially limits generalization to more diverse populations.

7 Ethical Considerations and Data Accessibility

The collection of the egoEMOTION dataset was approved by the ETH Zürich Ethics Commission (no. `23 ETHICS-008`). All participants provided informed consent for the recording of their sessions, the creation of the dataset, and its use for research purposes. To protect participants' privacy, all personally identifiable information (e.g., age, sex, skin type) and physiological data were anonymized using a numeric participant ID. However, given the inherently identifiable nature of egocentric, eye-tracking, and external video data, this information is treated as sensitive. While emotion and personality recognition can improve mental health monitoring, adaptive interfaces, and user-centric technologies, our dataset could be misused for behavioral profiling or targeted advertising. As such, access to this dataset requires users to be permanent staff members of an academic research institution and sign a Data Transfer and Use Agreement to adhere to the terms and conditions of the usage of this dataset. The dataset is hosted on servers from ETH Zurich for long-term availability and will be transferred using *sett* (the secure encryption and transfer tool) to minimize the risk of compromised data. Code to analyze the dataset is released under an open-source license.

8 Conclusion

We introduce egoEMOTION, the first publicly available dataset combining egocentric vision and physiological signals for emotion and personality recognition across both induced and naturalistic tasks. Capturing over 50 hours of synchronized multimodal recordings from 43 participants engaged in 16 emotionally diverse activities, egoEMOTION sets itself apart by covering a broad spectrum of real-world individual and social scenarios. It proposes three benchmark tasks: continuous affect regression, discrete emotion classification, and personality inference. Our results demonstrate that signals from egocentric devices—particularly eye-tracking features and head motion—outperform traditional physiological baselines in emotion and personality recognition tasks. These findings highlight the potential of egocentric vision systems to move beyond modeling observable behavior and towards capturing the underlying affective and dispositional states that shape human interaction. We envision egoEMOTION as a foundation for advancing affect-aware human-computer interaction and real-time user state estimation in the wild, enabling more personalized and emotionally intelligent systems across domains such as healthcare, education, and immersive computing.

References

[1] Peter N. Belhumeur, Joao P Hespanha, and David J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7):711–720, 1997. 7

[2] Shlomo Berkovsky, Ronnie Taib, Irena Koprinska, Eileen Wang, Yucheng Zeng, Jingjie Li, and Sabina Kleitman. Detecting personality traits using eye-tracking data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–12, New York, NY, USA, 2019. Association for Computing Machinery. 4

[3] Patricia Bota, João Brito, Ana Fred, et al. A real-world dataset of group emotion experiences based on physiological data. *Scientific Data*, 11:116, 2024. 3, 9

[4] Margaret Bradley and Peter J Lang. *The International affective digitized sounds (IADS): stimuli, instruction manual and affective ratings*. NIMH Center for the Study of Emotion and Attention, 1999. 3

[5] Björn Braun, Rayan Armani, Manuel Meier, Max Moebus, and Christian Holz. egoPPG: Heart Rate Estimation from Eye-tracking Cameras in Egocentric Systems to Benefit Downstream Vision Tasks. *arXiv preprint arXiv:2502.20879*, 2025. 1

[6] Björn Braun, Daniel McDuff, Tadas Baltrusaitis, and Christian Holz. Video-based sympathetic arousal assessment via peripheral blood flow estimation. *Biomedical Optics Express*, 14(12):6607–6628, 2023. 4

[7] Björn Braun, Daniel McDuff, Tadas Baltrusaitis, Paul Streli, Max Moebus, and Christian Holz. Sympcam: Remote optical measurement of sympathetic arousal. In *2024 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, pages 1–8. IEEE, 2024. 4

[8] Björn Braun, Daniel McDuff, and Christian Holz. How suboptimal is training rppg models with videos and targets from different body sites? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 410–418, 2024. 4

[9] Lili Cheng. 3 ways mixed reality empowers frontline workers, August 2023. Microsoft Industry Blog. 1

[10] P. T. J. Costa and R. R. McCrae. *NEO-PI-R Professional Manual: Revised NEO Personality and NEO Five-Factor Inventory (NEO-FFI)*. Psychological Assessment Resources, Odessa, Florida, 1992. 2, 3

[11] Dima Damen, Hazel Doughty, Giovanni Maria Farinella, Antonino Furnari, Jian Ma, Evangelos Kazakos, Davide Moltisanti, Jonathan Munro, Toby Perrett, Will Price, and Michael Wray. Rescaling egocentric vision: Collection, pipeline and challenges for epic-kitchens-100. *International Journal of Computer Vision (IJCV)*, 130:33–55, 2022. 3

[12] Sneha Das, Nicklas Leander Lund, Carlos Ramos González, and Line H Clemmensen. Emopaircompete - physiological signals dataset for emotion and frustration assessment under team and competitive behaviors. In *ICLR 2024 Workshop on Learning from Time Series For Health*, 2024. 3

[13] Jakob Engel, Kiran Somasundaram, Michael Goesele, Albert Sun, Alexander Gamino, Andrew Turner, Arjang Talattof, Arnie Yuan, Bilal Souti, Brighid Meredith, et al. Project aria: A new tool for egocentric multi-modal ai research. *arXiv preprint*, arXiv:2308.13561, 2023. 2, 5

[14] Thomas B Fitzpatrick. The validity and practicality of sun-reactive skin types i through vi. *Archives of Dermatology*, 124(6):869–871, Jun 1988. 6

[15] Guillermo Garcia-Hernando, Shanxin Yuan, Seungryul Baek, and Tae-Kyun Kim. First-person hand action benchmark with rgb-d videos and 3d hand pose annotations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 409–419, 2018. 1

[16] Christoph Gebhardt, Andreas Brombach, Tiffany Luong, Otmar Hilliges, and Christian Holz. Detecting users' emotional states during passive social media use. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 8(2), May 2024. 5

[17] Kristen Grauman, Andrew Westbury, Eugene Byrne, Zachary Chavis, Antonino Furnari, Rohit Girdhar, Jackson Hamburger, Hao Jiang, Miao Liu, Xingyu Liu, et al. Ego4d: Around the world in 3,000 hours of egocentric video. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 18995–19012, 2022. 1, 3

[18] Kristen Grauman, Andrew Westbury, Lorenzo Torresani, Kris Kitani, Jitendra Malik, Triantafyllos Afouras, Kumar Ashutosh, Vijay Baiyya, Siddhant Bansal, and Bikram Boote et al. Ego-exo4d: Understanding skilled human activity from first- and third-person perspectives, 2024. 1, 3

[19] Pascal Hartig. Building multimodal ai for ray-ban meta glasses, March 2025. Meta Engineering Blog. 1

[20] Elaine C. S. Hayashi, Julián E. Gutiérrez Posada, Vanessa R. M. L. Maike, and M. Cecília C. Baranauskas. Exploring new formats of the self-assessment manikin in the design with children. In *Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems*, IHC '16, New York, NY, USA, 2016. Association for Computing Machinery. 2, 5, 22

[21] Christian Holz and Edward J. Wang. Glabella: Continuously sensing blood pressure behavior using an unobtrusive wearable device. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(3), September 2017. 3

[22] Sabrina Hoppe, Tobias Loetscher, Stephanie A. Morey, and Andreas Bulling. Eye movements during everyday behavior predict personality traits. *Frontiers in Human Neuroscience*, 12:105, 2018. 4

[23] Apple Inc. Apple vision pro. `https://www.apple.com/apple-vision-pro/`, 2024. Accessed: 2025-04-13. 1

[24] JEOresearch. Eyetracker: A lightweight and robust python eye tracker. `https://github.com/JEOresearch/EyeTracker`, 2025. Accessed: 2025-03-14. 7

[25] Petr Kellnhofer, Adria Recasens, Simon Stent, Wojciech Matusik, and Antonio Torralba. Gaze360: Physically unconstrained gaze estimation in the wild. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6912–6921, 2019. 1

[26] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis using physiological signals. *IEEE Transactions on Affective Computing*, 3(1):18–31, Jan.-March 2012. 3, 7

[27] Nikola Kovacevic, Christian Holz, Markus Gross, and Rafael Wampfler. On multimodal emotion recognition for human-chatbot interaction in the wild. In *Proceedings of the 26th International Conference on Multimodal Interaction*, ICMI '24, page 12–21, New York, NY, USA, 2024. Association for Computing Machinery. 3

[28] Nikola Kovačević, Christian Holz, Tobias Günther, Markus Gross, and Rafael Wampfler. Personality trait recognition based on smartphone typing characteristics in the wild. *IEEE Transactions on Affective Computing*, 14(4):3207–3217, 2023. 3

[29] Krzysztof Kutt, Dominik Drążyk, Laura Żuchowska, Michał Szelążek, Szymon Bobek, and Grzegorz J Nalepa. Biraffe2, a multimodal dataset for emotion-based personalization in rich affective game environments. *Scientific Data*, 9:274, 2022. 3

[30] J. Kwon, J. Ha, D.-H. Kim, J. W. Choi, and L. Kim. Emotion recognition using a glasses-type wearable device via multi-channel facial responses. *IEEE Access*, 9:146392–146403, 2021. 3

[31] Taein Kwon, Bugra Tekin, Jan Stühmer, Federica Bogo, and Marc Pollefeys. H2o: Two hands manipulating objects for first person interaction recognition. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10138–10148, 2021. 1

[32] Peter J. Lang, Margaret M. Bradley, and Bruce N. Cuthbert. International affective picture system (iaps): Technical manual and affective ratings. Technical Report 1, NIMH Center for the Study of Emotion and Attention, 1997. 3, 4

[33] François Larradet, Rafał Niewiadomski, Giovanni Barresi, Darwin G. Caldwell, and Leonardo S. Mattos. Toward emotion recognition from physiological signals in the wild: Approaching the methodological issues in real-life data collection. *Frontiers in Psychology*, 11, 2020. 2, 3

[34] Robert W. Levenson. Emotion and the autonomic nervous system: A prospectus for research on autonomic specificity. In H. Wagner, editor, *Social Psychophysiology: Perspectives on Theory and Clinical Applications*, pages 17–42. Wiley, London, 1988. 9

[35] Shan Li and Weihong Deng. Deep facial expression recognition: A survey. *IEEE transactions on affective computing*, 13(3):1195–1215, 2020. 10

[36] Weijian Li, Runze Tan, Yiming Xing, et al. A multimodal psychological, physiological and behavioural dataset for human emotions in driving tasks. *Scientific Data*, 9:481, 2022. 3

[37] Tiffany Luong and Christian Holz. Characterizing physiological responses to fear, frustration, and insight in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3917–3927, 2022. 4

[38] Lingni Ma, Yuting Ye, Fangzhou Hong, Vladimir Guzov, Yifeng Jiang, Rowan Postyeni, Luis Pesqueira, Alexander Gamino, Vijay Baiyya, Hyo Jin Kim, et al. Nymeria: A massive collection of multimodal egocentric daily motion in the wild. In *European Conference on Computer Vision*, pages 445–465. Springer, 2024. 1, 3

[39] Yusuf Mansour, Ajoy Savio Fernandes, Kiran Somasundaram, Tarek Hefny, Mahsa Shakeri, Oleg V. Komogortsev, Abhishek Sharma, and Michael J. Proulx. Enabling eye tracking for crowd-sourced data collection with project aria. *IEEE Access*, 13:114736–114745, 2025. 7

[40] Meta. Meta quest 3: Expand your world with meta quest 3. https://www.oculus.com/quest-3/, 2024. Accessed: 2025-03-13. 1

[41] Joseph A. Mikels, Barbara L. Fredrickson, Grace R. Larkin, Claire M. Lindberg, Stephen J. Maglio, and Patricia A. Reuter-Lorenz. Emotional category data on images from the international affective picture system. *Behavior Research Methods*, 37(4):626–630, 2005. 2, 4, 5, 22

[42] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras. Amigos: A dataset for affect, personality and mood research on individuals and groups. *IEEE Transactions on Affective Computing*, 12(2):479–493, April-June 2021. 2, 3, 4, 5, 7

[43] Max Moebus, Lars Hauptmann, Nicolas Kopp, Berken Demirel, Björn Braun, and Christian Holz. Nightbeat: Heart rate estimation from a wrist-worn accelerometer during sleep. *IEEE Journal of Biomedical and Health Informatics*, 2024. 2

[44] T. Morris et al. Blink detection for real-time eye tracking. *Journal of Network and Computer Applications*, 2002. 7

[45] Mikhail Mozikov, Nikita Severin, Valeria Bodishtianu, Maria Glushanina, Ivan Nasonov, Daniil Orekhov, Pekhotin Vladislav, Ivan Makovetskiy, Mikhail Baklashkin, Vasily Lavrentyev, et al. Eai: Emotional decision-making of llms in strategic games and ethical dilemmas. *Advances in Neural Information Processing Systems*, 37:53969–54002, 2024. 1

[46] Jingping Nie, Yanchen Liu, Yigong Hu, Yuanyuting Wang, Stephen Xia, Matthias Preindl, and Xiaofan Jiang. Spiders+: A light-weight, wireless, and low-cost glasses-based wearable platform for emotion sensing and bio-signal acquisition. *Pervasive and Mobile Computing*, 75:101424, 2021. 3

[47] Kevin N Ochsner and Daniel L Schacter. A social cognitive neuroscience approach to emotion and memory. *The neuropsychology of emotion*, pages 163–193, 2000. 1

[48] Takehiko Ohkawa, Kun He, Fadime Sener, Tomas Hodan, Luan Tran, and Cem Keskin. Assemblyhands: Towards egocentric activity understanding via 3d hand pose estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12999–13008, 2023. 1

[49] Chanjoo Y. Park, Nayoung Cha, Sunwoo Kang, Hyeoncheol Hwang, Yanghwa Cho, Joonwon Lee, Jaewoo Lee, and Jong-Seok Lee. K-emocon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations. *Scientific Data*, 7:293, 2020. 3

[50] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. 8

[51] Rosalind W. Picard and Jennifer A. Healey. Affective wearables. *Personal Technologies*, 1(4):231–240, 1997. 2

[52] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics express*, 18(10):10762–10774, 2010. 4

[53] Jonathan Rottenberg, Richard D. Ray, and James J. Gross. Emotion elicitation using films. In *Handbook of Emotion Elicitation and Assessment*, Series in Affective Science, pages 12 – 28. Oxford University Press, 2007. 4

[54] James A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980. 2, 5

[55] Luz Santamaria-Granados, Mario Munoz-Organero, Gustavo Ramirez-Gonzalez, Enas Abdulhay, and NJIA Arunkumar. Using deep convolutional neural network for emotion detection on a physiological signals dataset (amigos). *IEEE Access*, 7:57–67, 2018. 9, 10

[56] Philip Schmidt, Robert Dürichen, Attila Reiss, Kristof Van Laerhoven, and Thomas Plötz. Multi-target affect detection in the wild: an exploratory study. In *Proceedings of the 2019 ACM International Symposium on Wearable Computers*, ISWC '19, page 211–219, New York, NY, USA, 2019. Association for Computing Machinery. 3

[57] Xuan Shui, Meng Zhang, Zhihao Li, et al. A dataset of daily ambulatory psychological and physiological recording for emotion research. *Scientific Data*, 8:161, 2021. 3

[58] Vasileios Skaramagkas, Emmanouil Ktistakis, Dimitrios Manousos, Eleni Kazantzaki, Nikolaos S. Tachos, Evanthia Tripoliti, Dimitrios I. Fotiadis, and Manolis Tsiknakis. esee-d: Emotional state estimation based on eye-tracking dataset. *Brain Sciences*, 13(4):589, 2023. 3

[59] E. Smets, W. De Raedt, and C. Van Hoof. Into the wild: The challenges of physiological stress detection in laboratory and ambulatory settings. *IEEE Journal of Biomedical and Health Informatics*, 23(2):463–473, March 2019. 2

[60] Mohammad Soleymani, Johan Lichtenauer, Thierry Pun, and Maja Pantic. A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing*, 3(1):42–55, Jan.-March 2012. 3, 4, 5

[61] Christopher J. Soto and Oliver P. John. The next big five inventory (bfi-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, 113(1):117–143, 2017. 5

[62] R. Subramanian, J. Wache, M. K. Abadi, R. L. Vieriu, S. Winkler, and N. Sebe. Ascertain: Emotion and personality recognition using commercial sensors. *IEEE Transactions on Affective Computing*, 9(2):147–160, April-June 2018. 2, 3, 4, 7

[63] Piotr Tarnowski, Maciej Kołodziej, Andrzej Majkowski, and Ryszard J. Rak. Eye-tracking analysis for emotion recognition. *Computational Intelligence and Neuroscience*, 2020(1):2909267, 2020. 3

[64] Rafael Wampfler, Severin Klingler, Barbara Solenthaler, Victor R. Schinazi, Markus Gross, and Christian Holz. Affective state prediction from smartphone touch and sensor data in the wild. CHI '22, New York, NY, USA, 2022. Association for Computing Machinery. 3

[65] Yujin Wu, Mohamed Daoudi, and Ali Amad. Transformer-based self-supervised multimodal representation learning for wearable emotion recognition. *IEEE Transactions on Affective Computing*, 15(1):157–172, 2023. 9, 10

[66] Wen-Jing Yan, Xiaobai Li, Su-Jing Wang, Guoying Zhao, Yong-Jin Liu, Yu-Hsin Chen, and Xiaolan Fu. Casme ii: An improved spontaneous micro-expression database and the baseline evaluation. *PloS one*, 9(1):e86041, 2014. 7

[67] Jingyuan Yang, Jie Li, Leida Li, Xiumei Wang, and Xinbo Gao. A circular-structured representation for visual emotion distribution learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4237–4246. IEEE, 2021. 5

[68] Zitong Yu, Yuming Shen, Jingang Shi, Hengshuang Zhao, Philip HS Torr, and Guoying Zhao. Physformer: Facial video-based physiological measurement with temporal difference transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4186–4196, 2022. 4

[69] Guoying Zhao and Matti Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):915–928, 2007. 7

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Our claims in the abstract and introduction accurately reflect the paper's contributions and scope.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: see Section 6.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

Justification: The paper does not include theoretical results.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The study protocol is described in section 3 and in further detail in Appendix A. Additionally, we provide the code to run the baselines.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: Our paper provides open access to the data and code (both linked to in the abstract), with sufficient instructions to faithfully reproduce the main experimental results described in Sections 4 and 5.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: We specify all details of our employed approach in Section 5. We use a leave-one-subject-out cross-validation approach, ensuring no data leakage between participants, and use the default settings for all deployed SciPy classifiers. Furthermore, we provide the entire preprocessing, feature calculation, and training and testing pipeline in our code, which can be accessed with the link in the abstract.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The paper reports error bars and other statistical significance for the experiments we conducted. We have included such information in the Supplementary Material (Appendix C) to improve clarity in the paper. Additionally, we provide an extensive analysis of the correlations between self-reports and sensor modalities in Appendix B.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   We provide all compute details in Section 5. All computations were run on AMD EPYC CPUs. One CPU is enough to process each participant individually, with about 50 GB of RAM necessary. Processing of the pupils and video data took about 2 hours per participant. The rest of the feature extraction and training/testing took under 1 minute per participant and feature.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

Justification: The paper discuss both potential positive societal impacts and negative societal impacts of the work performed in section 7.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [Yes]

    Justification: The paper describes in section 7 the safeguards that have been put in place for responsible release of data that have a high risk for misuse.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The original owners of code, dataset and model used in our paper are properly cited and credited. The license and terms of use explicitly mentioned in the supplemental material and properly respected.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.

- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: All new assets introduced in our paper including, data, code, and models are well documented with documents included in the paper (links available in abstract).

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [Yes]

    Justification: The paper includes the full set of instructions given to participants as well as the compensation received by participants (see section 3 and Supplementary Material).

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [Yes]

Justification: The study protocol was approved by the ETH Zürich Ethics committee (`no. 23 ETHICS-008`), as described in section 7.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (https://neurips.cc/Conferences/2025/LLM) for what should or should not be described.

Appendix

A Study Protocol

Before starting the experiment, an explanation of the study protocol, illustrated in Figure 5, was given to each participant. The experiment consisted in participants watching 9 videos and performing 7 tasks, as listed in Table 2. Details of the videos are provided in Table 12. We informed participants that each target emotion would be experienced only once during session A. This ensured that, after viewing a disturbing video, such as one eliciting disgust or fear, they would not anticipate encountering a similar emotional stimulus in the remaining videos. Between each stimulus, a washout video of clouds was shown to mitigate any emotional carry-over effect. Washouts lasted 40 seconds in session A and 1 minute in session B.

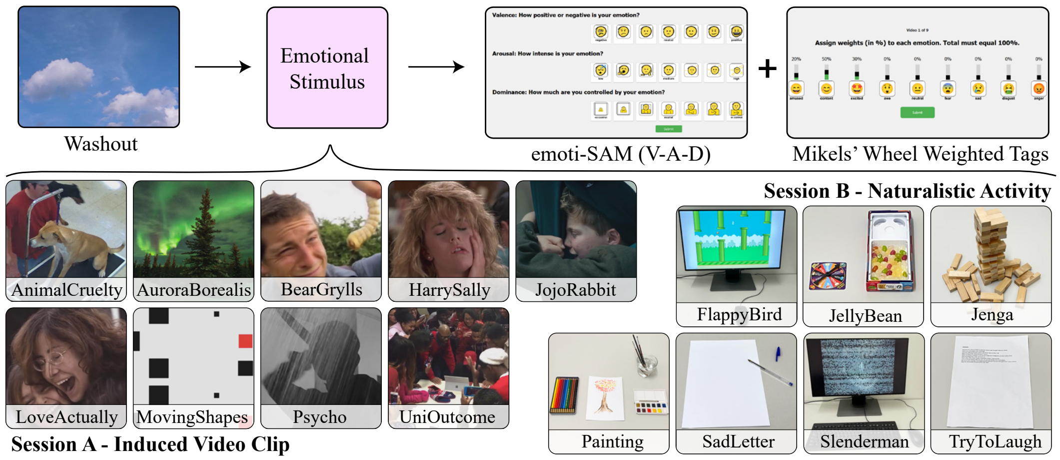


Figure 5: **Overview of the experimental protocol**. The experiment consisted of two sessions. In session A, participants watched 9 video clips, with a 40 s washout between clips and a 5 s video of a cross preceding each clip. In session B, participants performed 7 real-world tasks. Each task was spaced by a 1-min washout clip. Two questionnaires, corresponding to the emoti-SAM [20] and a weighted Mikels' Wheel [41] were answered after each emotional stimulus.

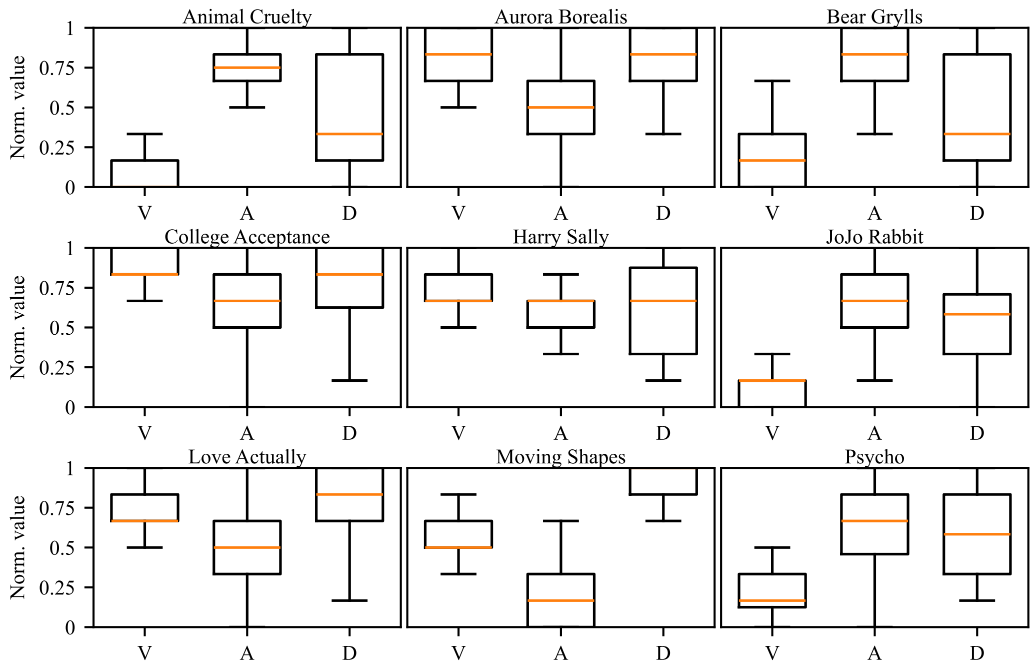
Following each emotional stimulus, participants rated their emotions using an emoti-SAM [20] and a weighted Mikels' Wheel [41], as shown in Figure 6. To familiarise the participant with each questionnaire, we explained what each term in the emoti-SAM meant i.e., *arousal*, *valence*, *dominance* and provided a definition for each emotion on Mikels' Wheel. In addition, we gave two examples of emotions and their associated self-reports. For the weighted Mikels' Wheel questionnaire, we indicated to the participant that they could gauge the intensity of their emotion using the neutral emotion. For example, if only feeling a single emotion but in low intensity, the participant could distribute the remaining weights in the neutral emotion. (e.g., 20% amused and 80% neutral indicates low amusement).

Figure 6: **Close-up view of the self-report questionnaires**. In the emoti-SAM [20], participants rated their arousal, valence and dominance using a 7-point scale. In the weighted Mikels' Wheel [41], participants distributed a 100% weight across emotions in 10% increments.

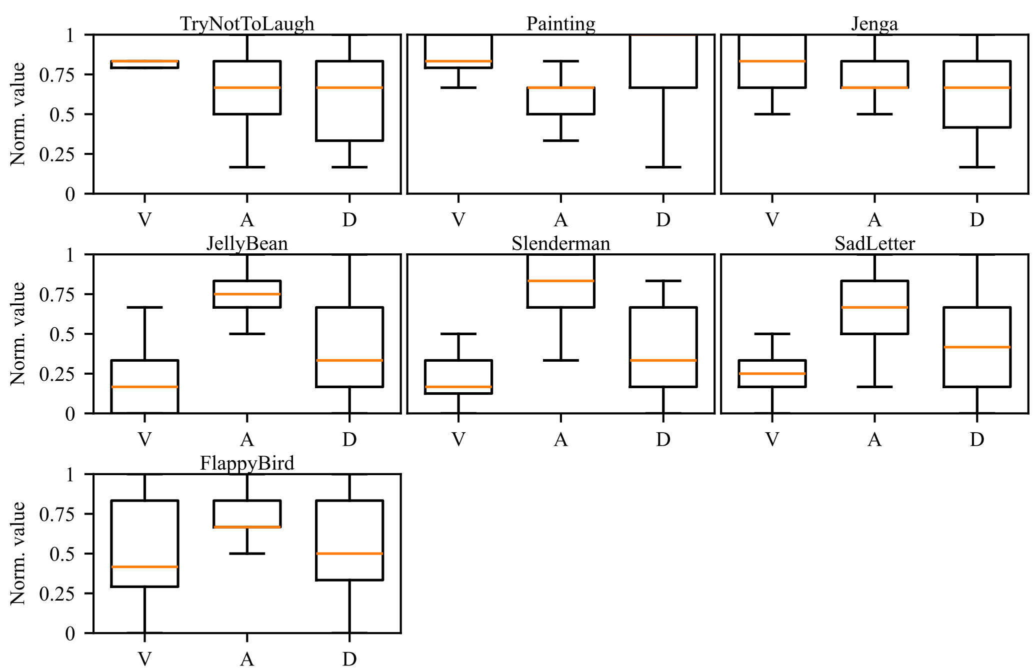
B Additional Dataset Descriptives

B.1 Mean self-ratings per task

The normalized continuous affect self-ratings for all video clips, averaged across participants, is displayed in Figure 7a. Similarly, the mean continuous affect self-ratings for the naturalistic activities of session B are displayed in Figure 7b.



(a)



(b)

Figure 7: Box plots of affect self-ratings **a)** per video clip **b)** per naturalistic task.

B.2 Self-correlations of continuous self-ratings

Figure 8 presents the Pearson correlation matrices between continuous affect self-ratings (A-V-D) across sessions A, B and A+B. Across all sessions, a strong negative relationship between arousal and dominance was observed, as well as a moderate positive relationship between valence and dominance. This indicated participants associated intense emotions with low dominance and vice-versa, while associating negative emotions to low dominance.

(a)

	Arousal	Valence	Dominance
Arousal	1.00	-0.24	-0.52*
Valence		1.00	0.40*
Dominance			1.00

(b)

	Arousal	Valence	Dominance
Arousal	1.00	-0.19	-0.53*
Valence		1.00	0.49*
Dominance			1.00

(c)

	Arousal	Valence	Dominance
Arousal	1.00	-0.20	-0.54*
Valence		1.00	0.42*
Dominance			1.00

Figure 8: Pearson correlations between continuous self-ratings in **a)** session A **b)** session B **c)** session A+B.

B.3 Self-correlations of discrete emotions

Figures 9, 10, 11 present the Pearson correlation matrices between discrete emotions across sessions A, B and A+B, respectively. The video clips of session A resulted in significant negative relationships between the neutral emotions and all other emotions excluding amused and disgust. Fear had a positive correlation with excitement and anger, while anger had a negative relationship with disgust. In session B, amusement had a strong negative correlation with anger and a moderate negative correlation with disgust.

	Amused	Content	Excited	Awe	Neutral	Fear	Sad	Disgust	Anger
Amused	1.00*	0.00	-0.18	-0.10	-0.19	-0.13	-0.07	-0.08	-0.09
Content		1.00*	-0.24	-0.23	-0.33*	0.02	0.07	-0.15	0.20
Excited			1.00*	0.07	-0.47*	0.33*	0.10	0.05	0.27
Awe				1.00*	-0.41*	0.29	0.03	0.11	0.10
Neutral					1.00*	-0.54*	-0.33*	-0.16	-0.43*
Fear						1.00*	0.00	-0.24	0.32*
Sad							1.00*	0.27	-0.14
Disgust								1.00*	-0.37*
Anger									1.00*

Figure 9: Pearson correlations between discrete emotions (session A).

	Amused	Content	Excited	Awe	Neutral	Fear	Sad	Disgust	Anger
Amused	1.00*	0.11	0.02	0.02	-0.29	-0.18	-0.02	-0.38*	-0.66*
Content		1.00*	-0.20	-0.28	-0.25	-0.33*	0.05	-0.10	-0.17
Excited			1.00*	-0.27	-0.26	-0.22	0.13	0.07	-0.02
Awe				1.00*	-0.11	0.37*	-0.06	-0.13	-0.09
Neutral					1.00*	-0.33*	-0.60*	-0.03	-0.05
Fear						1.00*	0.33*	-0.07	0.15
Sad							1.00*	-0.07	0.03
Disgust								1.00*	0.20
Anger									1.00*

Figure 10: Pearson correlations between discrete emotions (session B).

Fear had a moderate positive relationship with content and neutral, while having a negative relationship with awe and sadness. Finally, sadness had a strong positive relationship with the neutral emotion.

After combining the self-reports of discrete emotions from session A and B, amusement was negatively correlated with disgust and anger, while fear was positively correlated with awe. The neutral emotion was negatively correlated with excitement, awe, fear, sadness and anger.

	Amused	Content	Excited	Awe	Neutral	Fear	Sad	Disgust	Anger
Amused	1.00*	0.07	-0.21	-0.17	-0.14	-0.16	0.04	-0.39*	-0.36*
Content		1.00*	-0.29	-0.23	-0.29	-0.19	0.07	-0.09	-0.13
Excited			1.00*	-0.07	-0.33*	0.13	0.25	0.15	0.28
Awe				1.00*	-0.33*	0.35*	0.06	0.01	0.21
Neutral					1.00*	-0.48*	-0.57*	-0.17	-0.42*
Fear						1.00*	0.20	-0.11	0.26
Sad							1.00*	0.07	0.08
Disgust								1.00*	0.31
Anger									1.00*

Figure 11: Pearson correlations between discrete emotions (session A+B).

B.4 Self-correlations of personality traits

Figure 12 presents the Pearson correlations between personality traits. No significant correlation was found between personality traits.

	Ex	Ag	Co	NE	OM
Ex	1.00	0.20	0.17	-0.28	0.13
Ag		1.00	0.12	-0.09	-0.08
Co			1.00	-0.26	-0.02
NE				1.00	0.28
OM					1.00

Figure 12: Pearson correlations between personality traits.

B.5 Correlations between continuous self-ratings and personality traits.

Figures 13a and 13b display the correlations between the continuous self-ratings and personality traits in session A and session B, respectively. In session A, significant negative correlations are found between dominance and *Ex* and *Co*, as well as arousal and *OM*. In session B, a negative relationship between dominance and *Co* was observed.

	Ex	Ag	Co	NE	OM
Arousal	0.20	0.20	0.20	-0.30	-0.42*
Valence	-0.07	-0.13	-0.01	-0.01	0.03
Dominance	-0.31*	-0.12	-0.34*	0.17	0.22

(a)

	Ex	Ag	Co	NE	OM
Arousal	-0.01	0.10	-0.10	0.07	0.08
Valence	-0.10	-0.05	-0.30	0.20	0.30
Dominance	-0.24	-0.16	-0.37*	-0.05	0.09

(b)

Figure 13: Pearson correlations between continuous self-ratings and personality scores in **a)** session A **b)** session B.

B.6 Correlations between continuous self-ratings and discrete emotions.

Figures 14a and 14b present the Pearson correlations between continuous self-ratings and discrete emotions in session A and session B, respectively. In session A, arousal was positively correlated with excitement, fear, sadness and anger, while being negatively correlated with the neutral emotion. Valence was positively correlated with amusement and negatively correlated with disgust. In session B, arousal was positively correlated with excitement and negatively correlated with the neutral emotion. Valence was strongly positively correlated to amusement, while being strongly negatively correlated with fear and anger. Dominance was negatively correlated with fear, sadness and disgust, while being positively correlated with the neutral emotion.

	Amused	Content	Excited	Awe	Neutral	Fear	Sad	Disgust	Anger
Arousal	0.03	0.06	0.39*	0.18	-0.63*	0.32*	0.41*	0.28	0.32*
Valence	0.49*	0.27	0.01	-0.17	-0.02	-0.26	-0.24	-0.34*	-0.04
Dominance	-0.18	-0.03	0.00	-0.02	0.25	-0.18	0.03	-0.16	-0.24

(a)

	Amused	Content	Excited	Awe	Neutral	Fear	Sad	Disgust	Anger
Arousal	-0.06	-0.13	0.37*	0.07	-0.42*	0.22	0.25	0.22	0.10
Valence	0.65*	0.39*	0.18	-0.08	-0.05	-0.55*	-0.20	-0.27	-0.62*
Dominance	0.06	0.08	-0.30	0.09	0.44*	-0.32*	-0.32*	-0.32*	-0.04

(b)

Figure 14: Pearson correlations between continuous self-ratings and discrete emotions in **a)** session A **b)** session B.

B.7 Pearson correlations between personality scores and discrete emotions.

Figures 15a and 15b present the Pearson correlations between personality traits and discrete emotions in session A and session B, respectively. In session A, significant positive correlations were observed between *Ag* and sadness and disgust. In session B, *Ex* was positively correlated with content. *Ag* was positively correlated with sadness and negatively correlated with awe.

	Amused	Content	Excited	Awe	Neutral	Fear	Sad	Disgust	Anger
Ex	0.23	0.13	0.09	-0.15	-0.27	-0.11	0.08	0.30	0.22
Ag	-0.05	-0.02	-0.05	-0.05	-0.12	-0.02	0.35*	0.42*	-0.02
Co	-0.11	0.06	0.13	0.14	-0.22	0.15	0.14	0.19	-0.06
NE	-0.10	-0.01	-0.09	-0.06	0.15	0.04	-0.04	-0.15	-0.01
OM	0.03	-0.07	-0.02	-0.13	0.15	-0.22	-0.10	0.15	-0.06

(a)

	Amused	Content	Excited	Awe	Neutral	Fear	Sad	Disgust	Anger
Ex	-0.05	0.32*	-0.04	-0.16	-0.24	-0.06	0.28	-0.08	0.17
Ag	-0.10	0.13	0.07	-0.35*	-0.12	-0.02	0.42*	0.08	0.06
Co	-0.21	0.06	0.06	-0.02	-0.05	0.08	0.03	-0.07	0.23
NE	0.05	-0.13	0.21	0.18	-0.20	-0.01	0.02	0.24	-0.04
OM	0.17	0.15	0.04	0.17	-0.10	-0.17	-0.25	-0.18	0.03

(b)

Figure 15: Pearson correlation between personality scores and discrete emotions in **a)** session A **b)** session B.

C Additional Descriptions

C.1 Detailed Description of Baseline Features

A detailed overview of the features extracted from each modality is presented in Table 5.

Table 5: **Overview of features extracted from the recorded physiological time-series signals.** Recorded physiological signals include respiration rate, ECG, EDA, PPG recorded using a nosepad sensor and acceleration magnitude from the Aria-integrated IMU sensor. We further compute statistical descriptors of the time-series signals inferred from the video data captured by the eye-tracking cameras, including pupil size, video pixel intensity, and Fisherface feature coefficients.

Time-series signal	**Features extracted**
Acceleration Magnitude (Aria IMU)	Mean, min, max, standard deviation, median, 5th and 95th percentiles, range, interquartile range, sum, energy, skewness, kurtosis, RMS, and line integral.
Blinking (ET camera)	Number of blinks of the left eye and the right eye.
ECG	Root mean square of the mean squared IBIs, mean IBI, 60 spectral power values in the [0–6] Hz band of the ECG signal, low-frequency [0.01–0.08] Hz, medium-frequency [0.08–0.15] Hz, and high-frequency [0.15–0.5] Hz components of HRV spectral power, HR and HRV statistics.
EDA	Mean skin resistance and mean of derivative, mean differential for negative values only (mean decrease rate during decay time), proportion of negative derivative samples, number of local minima in the GSR signal, average rising time of the GSR signal, spectral power in the [0–2.4] Hz band, zero crossing rate of skin conductance slow response (SCSR) [0–0.2] Hz, zero crossing rate of skin conductance very slow response (SCVSR) [0–0.08] Hz, mean SCSR and SCVSR peak magnitude.
Eye Gaze (Yaw and Pitch)	Mean, min, max, standard deviation, median, 5th and 95th percentiles, range, interquartile range, sum, energy, skewness, kurtosis, RMS, and line integral.
Fisherface Features (ET camera)	Mean, min, max, standard deviation, median, 5th and 95th percentiles, range, interquartile range, sum, energy, skewness, kurtosis, RMS, and line integral.
Micro-expressions (ET camera)	Mean of each LBP-TOP feature.
PPG (Nosepad Sensor)	Root mean square of the mean squared IBIs, mean IBI, 60 spectral power values in the [0–6] Hz band of the PPG signal, low-frequency [0.01–0.08] Hz, medium-frequency [0.08–0.15] Hz, and high-frequency [0.15–0.5] Hz components of HRV spectral power, HR and HRV statistics.
Pupil Size (ET camera)	Mean, min, max, standard deviation, median, 5th and 95th percentiles, range, interquartile range, sum, energy, skewness, kurtosis, RMS, and line integral.
RSP	Band energy ratio (difference between the logarithm of energy between the lower (0.05–0.25 Hz) and the higher (0.25–5 Hz) bands), average respiration signal, mean of derivative (variation of the respiration signal), standard deviation, range or greatest breath, breathing rhythm (spectral centroid), breathing rate, 10 spectral power values in the bands from 0 to 2.4 Hz, average peak-to-peak time, median peak-to-peak time.
Video Pixel Intensity (ET camera)	Mean, min, max, standard deviation, median, 5th and 95th percentiles, range, interquartile range, sum, energy, skewness, kurtosis, RMS, and line integral.

C.2 Continuous Affect Prediction

Table 6 presents the continuous affect domain prediction results for session A and session B. The egocentric glasses provided better predictions of the continuous affect self-reports than the physiological sensors. The glasses had a F_1 score of 0.72 and 0.73 in session A and B, respectively, while the physiological sensors reported an F_1 score of 0.68 in session A and session B. Notably, all sensors had a strong performance when predicting arousal in session B.

Table 7 presents the standard deviation results of the continuous affect domain predictions across all sessions. Session A displays greater variability, particularly within the arousal and dominance self-reports. In contrast, Session B demonstrates lower and more uniform standard deviations across all modalities. When aggregating session A with session B, the lowest standard deviations are achieved.

Table 6: **Continuous affect domain prediction results.**

		Wearable devices				**Egocentric glasses**									**All**	**Baseline**
						ET video										
Session	Domain	ECG	EDA	RSP	⋈	Pup.	Int.	F.f.	Gaze	Blink	μ-E.	PPG	IMU	⋈	⋈	Random
A	Arousal	0.64	0.65	0.60	0.63	0.67	0.62	0.64	0.69	0.66	0.64	0.64	0.65	0.68	**0.69**	0.56
	Valence	0.71	0.68	0.62	0.71	0.71	0.66	**0.78**	0.64	0.71	0.63	0.77	0.66	**0.78**	**0.78**	0.56
	Dominance	0.70	0.70	0.71	0.71	0.68	**0.72**	0.68	0.71	0.72	0.71	0.69	0.71	0.71	0.71	0.61
	Mean	0.68	0.68	0.65	0.68	0.69	0.67	0.70	0.68	0.70	0.66	0.71	0.68	0.72	**0.73**	0.58
B	Arousal	**0.83**	**0.83**	**0.83**	**0.83**	**0.83**	**0.83**	**0.83**	**0.83**	**0.83**	**0.83**	**0.83**	**0.83**	**0.83**	**0.83**	0.74
	Valence	0.64	0.57	0.62	0.68	0.72	0.66	**0.78**	0.66	0.70	0.67	0.71	0.71	0.77	0.72	0.53
	Dominance	0.51	0.50	0.55	0.54	0.56	0.55	0.58	0.44	0.56	0.60	0.58	0.46	**0.59**	0.57	0.50
	Mean	0.66	0.63	0.67	0.68	0.70	0.68	**0.73**	0.64	0.70	0.70	0.71	0.65	**0.73**	0.71	0.59

⋈ = fusion of modalities, Pup. = Pupil size, Int. = Pixel Intensity, F.f. = Fisherface features, μ-E. = micro-expressions.

Table 7: **Standard deviation of continuous affect domain prediction results.**

		Wearable devices				**Egocentric glasses**									**All**	**Baseline**
						ET video										
Session	Domain	ECG	EDA	RSP	⋈	Pup.	Int.	F.f.	Gaze	Blink	μ-E.	PPG	IMU	⋈	⋈	Random
A	Arousal	0.22	0.22	0.20	0.20	0.22	0.22	0.22	0.24	0.23	0.22	0.21	0.24	0.22	0.20	
	Valence	0.12	0.10	0.16	0.15	0.27	0.19	0.08	0.15	0.05	0.10	0.08	0.11	0.14	0.14	
	Dominance	0.23	0.24	0.24	0.24	0.22	0.23	0.22	0.24	0.24	0.24	0.23	0.24	0.24	0.24	
	Mean	0.19	0.19	0.20	0.20	0.24	0.21	0.17	0.21	0.17	0.19	0.17	0.20	0.20	0.20	
B	Arousal	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	
	Valence	0.16	0.16	0.16	0.15	0.14	0.13	0.14	0.16	0.19	0.18	0.15	0.18	0.14	0.14	
	Dominance	0.23	0.26	0.23	0.25	0.25	0.22	0.26	0.25	0.23	0.26	0.22	0.22	0.26	0.22	
	Mean	0.18	0.19	0.18	0.19	0.18	0.17	0.19	0.19	0.19	0.20	0.18	0.19	0.19	0.17	
A+B	Arousal	0.14	0.13	0.13	0.14	0.14	0.14	0.13	0.14	0.14	0.13	0.14	0.13	0.15	0.14	
	Valence	0.10	0.11	0.10	0.09	0.10	0.10	0.06	0.09	0.09	0.09	0.09	0.10	0.10	0.09	
	Dominance	0.18	0.18	0.17	0.18	0.17	0.19	0.16	0.19	0.19	0.20	0.17	0.18	0.18	0.17	
	Mean	0.14	0.14	0.13	0.14	0.14	0.14	0.12	0.14	0.14	0.14	0.14	0.14	0.14	0.13	

⋈ = fusion of modalities, Pup. = Pupil size, Int. = Pixel Intensity, F.f. = Fisherface features, μ-E. = micro-expressions.

C.3 Discrete Emotion Prediction

Table 8 presents the discrete emotion prediction results in session A and session B. The egocentric glasses significantly exceeded the physiological sensors in predicting 9 discrete emotions ($F_1 = 0.55$ vs. $F_1 = 0.25$) in session A. In the naturalistic tasks, the egocentric glasses and the wearable devices had comparable results, with $F_1 = 0.40$ and $F_1 = 0.33$, respectively. With the current feature extraction, emotions such as sadness (0.87) and anger (0.68) had high prediction scores in session A using the combined sensors from the egocentric glasses. The emotion of awe proved to be difficult to predict across experiments, with discrete emotions in session A achieving higher prediction results in comparison to session B. Disgust has a high F_1 score in session B (0.75) when predicting it from the egocentric glasses. The eye pupil size was highly informative for predicting fear in participants during session B ($F_1 = 0.66$).

Table 9 presents the standard deviations of the discrete emotion prediction results. Session A tends to be noisier, with larger fluctuations in certain cases (e.g., IMU and F.f.), while Session B looks more consistent overall.

Table 8: **Discrete emotion prediction results.**

		Wearable devices				**Egocentric glasses**									**All**	**Baseline**
						ET video										
Session	Domain	ECG	EDA	RSP	⋈	Pup.	Int.	F.f.	Gaze	Blink	μ-E.	PPG	IMU	⋈	⋈	Random
A	Amused	0.05	0.52	0.37	0.50	0.60	0.19	**0.62**	0.47	0.17	0.17	0.52	0.54	0.52	0.57	0.12
	Content	0.25	0.19	0.20	0.26	0.38	0.13	0.56	0.26	0.13	0.13	0.44	0.40	**0.61**	**0.61**	0.16
	Excited	0.00	0.00	0.00	0.00	0.09	0.00	0.25	0.00	0.06	0.00	0.07	0.00	**0.29**	0.24	0.05
	Awe	0.05	0.00	0.07	0.00	**0.38**	0.06	0.28	0.00	0.05	0.05	0.00	0.06	0.25	0.34	0.07
	Neutral	0.23	0.36	0.31	0.32	0.40	0.24	0.49	0.44	0.27	0.32	0.37	0.41	0.51	**0.52**	0.21
	Fear	0.00	0.24	0.00	0.06	0.48	0.08	0.52	0.11	0.00	0.00	**0.60**	0.12	0.57	0.53	0.06
	Sad	0.25	0.82	0.23	**0.88**	0.77	0.07	0.85	0.68	0.13	0.13	**0.88**	0.75	0.87	0.87	0.10
	Disgust	0.07	0.18	0.17	0.12	0.20	0.19	0.47	0.09	0.15	0.16	0.57	0.27	**0.63**	0.60	0.13
	Anger	0.12	0.00	0.12	0.14	0.32	0.08	0.66	0.06	0.16	0.05	0.14	0.12	**0.68**	**0.68**	0.09
	Mean	0.11	0.26	0.16	0.25	0.40	0.12	0.52	0.23	0.12	0.11	0.40	0.30	**0.55**	**0.55**	0.11
B	Amused	0.57	0.50	0.58	0.60	0.50	0.48	**0.62**	0.44	0.44	0.51	0.58	0.52	0.61	0.60	0.32
	Content	0.44	0.19	0.38	0.48	0.24	0.24	**0.58**	0.24	0.18	0.28	0.46	0.28	0.57	0.57	0.15
	Excited	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.13	0.00	0.00	0.00	**0.19**	0.10	0.05
	Awe	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.01**
	Neutral	0.04	**0.25**	0.13	0.05	0.10	0.14	0.00	0.32	0.11	0.04	0.04	0.09	0.00	0.00	0.11
	Fear	0.14	0.33	0.33	0.60	0.66	0.04	0.64	0.22	0.14	0.26	0.57	0.28	**0.67**	0.66	0.11
	Sad	0.16	0.26	0.34	0.39	0.19	0.26	**0.49**	0.47	0.20	0.29	0.37	0.05	0.44	0.42	0.10
	Disgust	0.27	0.70	0.50	0.74	0.61	0.26	0.73	0.25	0.15	0.55	0.75	0.72	0.75	**0.77**	0.10
	Anger	0.00	0.00	0.00	0.15	0.00	0.08	0.27	0.17	0.07	0.00	0.18	0.00	**0.34**	0.24	0.06
	Mean	0.18	0.25	0.25	0.33	0.26	0.17	0.38	0.12	0.16	0.21	0.33	0.22	**0.40**	0.37	0.11

⋈ = fusion of modalities, Pup. = Pupil size, Int. = Pixel Intensity, F.f. = Fisherface features, μ-E. = micro-expressions.

Table 9: **Standard deviation of discrete emotion prediction results.**

		Wearable devices				**Egocentric glasses**									**All**	**Baseline**
						ET video										
Session	Domain	ECG	EDA	RSP	⋈	Pup.	Int.	F.f.	Gaze	Blink	μ-E.	PPG	IMU	⋈	⋈	Random
A	Amused	0.16	0.46	0.40	0.45	0.44	0.32	0.46	0.44	0.25	0.25	0.44	0.45	0.46	0.47	
	Content	0.26	0.23	0.24	0.31	0.32	0.25	0.35	0.24	0.15	0.20	0.39	0.32	0.34	0.35	
	Excited	0.00	0.00	0.00	0.00	0.10	0.00	0.28	0.00	0.11	0.00	0.16	0.00	0.28	0.28	
	Awe	0.16	0.00	0.10	0.00	0.41	0.10	0.34	0.00	0.06	0.11	0.00	0.16	0.36	0.42	
	Neutral	0.26	0.32	0.25	0.26	0.28	0.23	0.33	0.32	0.26	0.28	0.32	0.29	0.32	0.33	
	Fear	0.00	0.28	0.00	0.16	0.43	0.16	0.47	0.22	0.00	0.00	0.47	0.22	0.47	0.46	
	Sad	0.32	0.42	0.33	0.38	0.45	0.19	0.40	0.47	0.17	0.28	0.36	0.44	0.38	0.38	
	Disgust	0.16	0.32	0.30	0.21	0.29	0.31	0.45	0.25	0.21	0.28	0.47	0.38	0.44	0.44	
	Anger	0.24	0.00	0.24	0.30	0.38	0.19	0.48	0.17	0.25	0.16	0.28	0.24	0.46	0.46	
	Mean	0.17	0.23	0.21	0.23	0.34	0.19	0.40	0.11	0.08	0.11	0.32	0.28	0.39	0.40	
B	Amused	0.25	0.28	0.29	0.28	0.25	0.27	0.28	0.26	0.28	0.28	0.25	0.22	0.31	0.30	
	Content	0.39	0.32	0.40	0.44	0.29	0.36	0.43	0.23	0.31	0.34	0.44	0.37	0.41	0.42	
	Excited	0.00	0.00	0.00	0.00	0.00	0.00	0.16	0.00	0.16	0.00	0.00	0.00	0.22	0.16	
	Awe	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	Neutral	0.00	0.31	0.18	0.10	0.11	0.21	0.00	0.00	0.16	0.11	0.08	0.13	0.00	0.00	
	Fear	0.30	0.39	0.36	0.49	0.49	0.08	0.50	0.46	0.28	0.30	0.48	0.36	0.49	0.49	
	Sad	0.28	0.33	0.41	0.42	0.26	0.33	0.46	0.34	0.32	0.37	0.41	0.10	0.43	0.43	
	Disgust	0.27	0.51	0.45	0.50	0.50	0.36	0.50	0.50	0.28	0.48	0.50	0.50	0.50	0.49	
	Anger	0.00	0.00	0.00	0.22	0.00	0.16	0.28	0.16	0.16	0.00	0.24	0.00	0.31	0.28	
	Mean	0.17	0.24	0.23	0.27	0.21	0.20	0.29	0.16	0.18	0.21	0.27	0.19	0.30	0.29	
A+B	Amused	0.22	0.23	0.22	0.22	0.22	0.18	0.21	0.22	0.19	0.32	0.18	0.23	0.21	0.22	
	Content	0.23	0.21	0.26	0.24	0.26	0.23	0.30	0.22	0.24	0.20	0.28	0.22	0.24	0.24	
	Excited	0.00	0.10	0.00	0.00	0.19	0.00	0.18	0.00	0.00	0.00	0.00	0.00	0.20	0.18	
	Awe	0.10	0.00	0.00	0.00	0.32	0.00	0.34	0.16	0.08	0.00	0.00	0.00	0.43	0.40	
	Neutral	0.20	0.25	0.21	0.17	0.24	0.19	0.28	0.26	0.21	0.17	0.24	0.24	0.29	0.28	
	Fear	0.18	0.24	0.26	0.38	0.39	0.25	0.38	0.36	0.13	0.20	0.41	0.16	0.38	0.40	
	Sad	0.30	0.35	0.27	0.35	0.34	0.21	0.35	0.31	0.17	0.12	0.38	0.33	0.32	0.34	
	Disgust	0.16	0.28	0.27	0.32	0.33	0.25	0.35	0.29	0.19	0.20	0.36	0.34	0.34	0.32	
	Anger	0.00	0.18	0.22	0.19	0.30	0.00	0.42	0.21	0.19	0.03	0.31	0.27	0.44	0.40	
	Mean	0.15	0.20	0.19	0.21	0.29	0.15	0.31	0.08	0.05	0.14	0.24	0.20	0.32	0.31	

⋈ = fusion of modalities, Pup. = Pupil size, Int. = Pixel Intensity, F.f. = Fisherface features, μ-E. = micro-expressions.

C.4 Personality Prediction

Table 10 presents the personality trait predictions for session A and session B. The pupil size and blink rate achieved the highest F_1 score in session A, while ECG performed best in session B ($F_1 = 0.60$). Table 11 presents the standard deviations of the personality prediction results.

Table 10: **Personality prediction results.**

		Wearable devices				**Egocentric glasses**									**All**	**Baseline**
						ET video										
Session	Domain	ECG	EDA	RSP	$\bowtie$	Pup.	Int.	F.f.	Gaze	Blink	μ-E.	PPG	IMU	$\bowtie$	$\bowtie$	Random
	Ex	0.48	0.48	0.42	0.30	0.48	0.48	0.45	0.48	**0.58**	0.38	0.45	0.48	0.38	0.25	**0.55**
	Ag	0.45	0.42	0.48	0.38	0.42	0.32	0.22	0.40	**0.58**	0.45	0.57	0.50	0.48	0.52	0.52
A	Co	0.42	0.52	0.28	0.42	**0.65**	0.48	0.52	0.58	0.63	0.63	0.52	0.40	0.35	0.38	0.55
	NE	0.50	0.52	0.50	0.52	**0.68**	0.45	0.30	0.58	0.45	0.53	0.50	0.57	0.62	0.48	0.52
	OM	0.30	0.60	**0.62**	0.48	0.52	0.55	0.42	0.60	0.53	0.45	0.28	0.60	0.60	0.45	0.52
	Mean	0.43	0.51	0.46	0.42	**0.55**	0.46	0.38	0.53	**0.55**	0.49	0.46	0.51	0.49	0.42	0.53
	Ex	0.45	0.50	0.48	0.55	0.45	**0.75**	0.53	0.45	0.73	0.52	0.55	**0.75**	0.48	0.55	
	Ag	0.48	0.60	0.48	0.48	0.48	0.60	0.45	**0.70**	0.45	0.63	0.38	0.32	0.65	0.35	0.52
B	Co	0.68	0.40	0.35	0.62	0.45	0.28	0.40	**0.70**	0.40	0.45	0.62	0.50	0.45	0.62	0.55
	NE	**0.78**	0.52	0.65	0.70	0.60	0.52	0.42	0.53	0.43	0.73	0.48	0.65	0.40	0.72	0.52
	OM	**0.62**	0.52	0.57	0.35	0.48	0.50	0.40	0.48	0.53	0.43	0.52	0.40	0.35	0.28	0.52
	Mean	**0.60**	0.51	0.51	0.54	0.49	0.53	0.44	0.59	0.45	0.59	0.50	0.48	0.52	0.49	0.53

$\bowtie$ = fusion of modalities, Pup. = Pupil size, Int. = Pixel Intensity, F.f. = Fisherface features, μ-E. = micro-expressions.

Table 11: **Standard deviation of personality prediction results.**

		Wearable devices				**Egocentric glasses**									**All**	**Baseline**
						ET video										
Session	Domain	ECG	EDA	RSP	$\bowtie$	Pup.	Int.	F.f.	Gaze	Blink	μ-E.	PPG	IMU	$\bowtie$	$\bowtie$	Random
	Ex	0.50	0.50	0.49	0.48	0.50	0.50	0.50	0.50	0.49	0.48	0.50	0.50	0.48	0.45	
	Ag	0.50	0.49	0.50	0.50	0.49	0.47	0.42	0.46	0.49	0.50	0.49	0.50	0.50	0.50	
A	Co	0.49	0.50	0.45	0.50	0.48	0.50	0.50	0.46	0.48	0.48	0.50	0.49	0.48	0.48	
	NE	0.50	0.50	0.50	0.50	0.47	0.50	0.46	0.50	0.50	0.50	0.50	0.49	0.48	0.50	
	OM	0.46	0.49	0.48	0.50	0.50	0.50	0.49	0.50	0.50	0.50	0.45	0.49	0.49	0.50	
	Mean	0.49	0.50	0.48	0.50	0.49	0.50	0.47	0.48	0.49	0.49	0.49	0.49	0.49	0.49	
	Ex	0.49	0.50	0.50	0.50	0.50	0.43	0.50	0.49	0.50	0.45	0.50	0.50	0.43	0.50	
	Ag	0.50	0.49	0.50	0.50	0.50	0.49	0.50	0.49	0.50	0.48	0.48	0.47	0.48	0.47	
B	Co	0.48	0.49	0.48	0.48	0.50	0.45	0.49	0.50	0.49	0.50	0.48	0.50	0.50	0.48	
	NE	0.47	0.50	0.48	0.46	0.49	0.50	0.49	0.49	0.49	0.45	0.50	0.48	0.49	0.47	
	OM	0.49	0.50	0.49	0.48	0.50	0.50	0.49	0.50	0.50	0.49	0.50	0.49	0.48	0.45	
	Mean	0.49	0.50	0.49	0.48	0.50	0.47	0.49	0.50	0.50	0.47	0.49	0.49	0.48	0.47	
	Ex	0.43	0.50	0.47	0.40	0.49	0.46	0.50	0.49	0.49	0.50	0.49	0.50	0.50	0.50	
	Ag	0.48	0.49	0.50	0.49	0.50	0.49	0.49	0.49	0.48	0.49	0.49	0.49	0.46	0.49	
A+B	Co	0.49	0.50	0.46	0.50	0.50	0.49	0.49	0.50	0.49	0.50	0.48	0.50	0.50	0.48	
	NE	0.50	0.50	0.48	0.47	0.47	0.49	0.46	0.49	0.50	0.49	0.49	0.50	0.48	0.45	
	OM	0.48	0.50	0.50	0.49	0.50	0.49	0.49	0.50	0.48	0.47	0.46	0.48	0.45	0.48	
	Mean	0.48	0.50	0.48	0.47	0.49	0.48	0.49	0.50	0.49	0.49	0.48	0.49	0.48	0.48	

$\bowtie$ = fusion of modalities, Pup. = Pupil size, Int. = Pixel Intensity, F.f. = Fisherface features, μ-E. = micro-expressions.

Table 12: **Detailed description of the emotion-inducing video clips.**

ID	**Video Label**	**Target Emotion**	**Description**	**Duration (s)**
1	AnimalCruelty	Anger	Televised news of a dog groomer abusing dogs.	40
2	AuroraBorealis	Awe	A timelapse of the northern lights.	40
3	BearGrylls	Disgust	A man eats a worm.	40
4	CollegeAcceptance	Excitement	A student gets accepted to his dream college.	40
5	HarrySally	Amusement	Sally shows Harry how women fake orgasms at a restaurant.	72
6	JojoRabbit	Sadness	A boy embraces his mother who has been hanged.	46
7	LoveActually	Content	Narrator purporting that "love is everywhere".	42
8	MovingShapes	Neutral	Shapes moving on a neutral background.	40
9	Psycho	Fear	A lady gets murdered in her bathtub by an intruder.	45