

Pain Detection Through Facial Expressions Analysis: A Validation Study

Luisina Gregoret, Divya Seernani, Jessica Wilson,
Serena Pang, Mohammad Mavadati, Isabel Marin Garcia, Mina Bishay
iMotions A/S, Copenhagen, Denmark

Assessing pain objectively is challenging, as people vary considerably in how they experience and express pain. Automated facial expression analysis offers an approach for assessing pain-related facial behavior from video. Expressed pain can be accurately detected using the AFFDEX Pain metric. The model achieves a ROC-AUC of 96.1% on an internal dataset comprising both a short-lasting pain paradigm based on the cold pressor test and a longer-lasting pain paradigm based on delayed onset muscle soreness, exceeding several recently reported ROC-AUC results on the popular UNBC dataset and other pain datasets evaluated in recent scientific publications. However, facial expressions show intersubject variability. Thus, adding physiological measures such as GSR or ECG allows for cross-validation, mitigating potential mis-identifications that may arise from facial masking and low facial expressivity. Combining physiological measurements with facial expression analysis can give unique and novel insights into the subjective perception and variable expression of pain.

1 Introduction

The International Association for the Study of Pain (IASP) defines pain as “an unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage” [27]. This definition is accompanied by additional context emphasizing that pain is a personal experience influenced by biological, psychological, and social factors.

Pain is one of the most important bodily signals our body produces, and is a complex multidimensional experience modulating various aspects of the autonomic nervous system, while also affecting the subjective perception based on current environment, as well as past experiences. Measuring pain can therefore be equally complex given the natural intersubject variability in pain responses, and challenges in rating pain perception within a sample exposed to the same experience. The use of multimodal assessment helps to overcome the challenges in connecting the experience of pain with its objective measurement.

Numerous studies have investigated the impact of pain on the sympathetic and parasympathetic nervous system using electrophysiological measures, primarily electrocardiography (ECG) and galvanic skin response (GSR), also referred to as electrodermal activity (EDA). These evaluations range from clinical cases to ecologically valid, well-established experimental models, such as the Cold Pressor Test (CPT) and a controlled model to investigate musculoskeletal pain called delayed onset muscle soreness (DOMS) [17]. For instance, when healthy young adults immersed their hand in ice cold water in accordance with the CPT, studies have reported higher heart rate, lower time-based heart rate variability, higher pupil dilation [4, 29] as well as higher GSR peak amplitude [8], relative to pain-free stages. Similarly, DOMS, normally induced in healthy participants through repetitive eccentric exercises, has elicited higher heart rate and lower time-based heart rate variability, as compared with baseline pain-free stages [17].

Alongside these electrophysiological approaches, pain is also assessed through a range of pain scales. Some rely on self-report, while others draw on external observation, including facial expression analysis [3, 13, 24]. The gold-standard in facial expression-based pain analysis is the Prkachin and Solomon Pain Intensity (PSPI) scale [25, 26], which estimates pain expression using six facial action units derived from Paul Ekman’s Facial Action Coding System (FACS). Whereas traditionally researchers relied on trained human coders to manually label video footage, computer vision-assisted facial expression analysis has substantially streamlined the detection and analysis of expressed pain from facial expressions. This approach enables webcam-based detection of expressed pain, both in real time and in post-processing, and offers high temporal resolution, non-invasiveness, and the ability to cross-validate findings against established physiological measures such as ECG and GSR. The present study a) proposes a PSPI-based pain detection method using standard webcam video during the CPT and DOMS, b) assesses discriminant validity for both pain models and evaluates its performance against state-of-the-art machine learning and deep learning approaches, and c) characterizes physiological changes during pain and pain-free stages using primarily ECG and GSR measures.

2 Methods

Thirty-eight healthy adults were recruited for this study (mean age: 36.2 ± 10.35 years; 14 females), with 32 participants recruited in Copenhagen, Denmark, and 6 in Boston, United States of America. The inclusion criteria comprised healthy, pain-free volunteers aged 21–60 years. Exclusion criteria included the presence of acute pain at the time of the experiment or chronic pain, cardiorespiratory disease, visual or motor disabilities, consumption of coffee or energy drinks within 3 hours prior to the experiment, medication intake, and recent surgery within the preceding 6 months.

2.1 Experimental procedure

The study protocol comprised two different days separated by 48 hours (Figure 1). On Day 1, following the introduction and consent process, participants went through three phases: a pain-free Baseline (3 minutes), the CPT stage (3 minutes; for more details see Section 2.1.1), and a pain-free Recovery period (3 minutes). During the CPT stage, every participant continuously rated their pain using an electronic Visual Analog Scale (VAS) (Gazepoint Dial, Gazepoint, Canada) anchored at 0 (“no pain”) and 10 (“worst imaginable pain”) on Day 1, and rated their average self-reported pain once at the end of the CPT stage on Day 1 using a digital survey (iMotions 11.12, iMotions A/S, Denmark).

The CPT phase was followed by a series of repetitive exercises designed to induce Delayed Onset Muscle Soreness (DOMS); for further details, see Section 2.1.2. On Day 3, participants returned to the laboratory to complete a second pain-free baseline (approx. 75 seconds), a weight lifting stage causing movement-evoked pain due to DOMS (approx. 75 seconds), and a final recovery stage (approx. 75 seconds). After the DOMS-related pain assessment, participants rated their self-reported pain once at the end of the DOMS stage on Day 3 using a digital survey (iMotions 11.12, iMotions A/S, Denmark).

2.1.1 Cold Pressor Test (CPT)

Before the experiment began, the examiner provided a detailed explanation of the protocol to each participant. During the procedure, subjects immersed their non-dominant hand in a container of ice-cold water. Throughout the session, they recorded their perceived pain levels using one of two methods: 32 participants utilized a digital dial for continuous rating on a scale from 0 to 10,

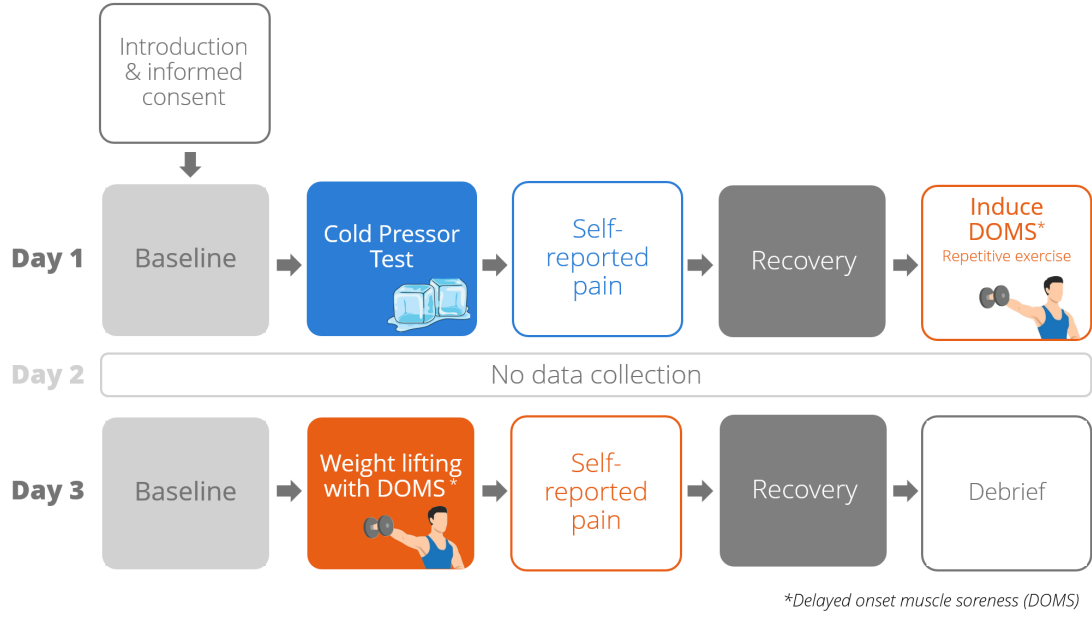


Figure 1: Diagram illustrating the main stages of the experiment. On Day 1, following the introduction and consent process, participants went through three phases: a pain-free Baseline, the Cold Pressor Test (CPT), and a pain-free Recovery period. This was followed by repetitive exercises designed to induce Delayed Onset Muscle Soreness (DOMS). On Day 3, participants returned to the lab to complete a second baseline, a weight-lifting task with lighter dumbbells as a DOMS-related pain assessment, and a final recovery stage.

while 6 participants provided discrete ratings on the same 0 to 10 scale via a numerical keypad. Participants were reminded of their right to withdraw their hand from the water at any point. To ensure consistency, the examiner verified that the water temperature remained under 3°C for the duration of the 3-minute recording period using a digital thermometer.

2.1.2 Delayed Onset Muscle Soreness (DOMS)

To induce DOMS on the upper arm (deltoid muscle) at the end of Day 1, the instructor led each participant through a specific protocol. Initially, the concentric 1 Repetition Maximum (1RM) was determined to identify the heaviest weight a participant could lift for a single repetition using their non-dominant arm [6]. Then, participants were instructed to perform eccentric repetitive exercises to ultimately induce DOMS [21] using the non-dominant arm in a scapular plane position. The load was set at 80–100% of the estimated 1RM. The protocol for DOMS induction on the upper arms consisted of a minimum of 5 bouts of 10 repetitions each, wherein the participant lowered the arm from 90° to 0° over 4–5 seconds. In between every bout, a 90-second break was set. If a participant had not reached exhaustion after the initial protocol of 5 bouts of 10 repetitions, supplementary blocks consisting of 10 repetitions were added until exhaustion, all separated by 90-second breaks in between.

2.2 Signal processing

Facial expression analysis (FEA). Participants were recorded using a Logitech C920 webcam operating at 30 fps, with a variable resolution ranging from 640 × 480 to 1920 × 1080 pixels. The webcam was positioned close to the participant, either above or below the display, to obtain a

clear view of the face. Facial expression analysis was performed using AFFDEX 2.0 [1], integrated within the iMotions software platform (iMotions 11.12, iMotions A/S, Denmark).

Facial Action Units (AUs) extracted by AFFDEX were processed by a newly developed pain-detection model based on the Prkachin and Solomon Pain Intensity (PSPI) framework that describes pain facial activity using six AUs: brow lowering (AU4), cheek raising and lid tightening (AU6 and AU7), nose wrinkling and upper-lip raising (AU9 and AU10), and eye closure (AU43), as illustrated in Figure 2.

The new pain model is implemented as a higher-level classifier that takes as input the facial AUs estimated by AFFDEX, and for each video frame, the model outputs a pain probability score from 0 to 100 reflecting no expressed pain (0) to intense pain (100) expressed via action units. A threshold of 50, as implemented in the present study, binarizes the pain score corresponding to a clear expression of pain versus no expressed pain.



Figure 2: Facial action units that are used to detect pain, in accordance with the PSPI scale: brow lowering (also called brow furrow, AU4), cheek raising and lid tightening (AU6 and AU7), nose wrinkling and upper-lip raising (AU9 and AU10), and eye closure (AU43).

GSR. GSR was recorded via a portable device (Shimmer3 GSR+, Shimmer Sensing, Ireland), positioning the two electrodes on the volar side of the palm of the dominant hand using pregelled Ag/AgCl disposable electrodes (Arbo/Kendall model H124SG 24 mm diameter, Cardinal Health, Canada). The phasic signals were computed by applying a median filter to the raw GSR signal. A low-pass Butterworth filter with a cut-off frequency of 5 Hz was then applied. Thresholds of $0.01 \mu\text{S}$ and $0 \mu\text{S}$ were used to detect the onset and offset of every GSR peak, respectively. Increases in the phasic GSR signal were labeled as GSR peaks based on two criteria: i) if the amplitude between the onset and offset of every possible peak surpassed a minimum threshold of $0.005 \mu\text{S}$; and ii) the time difference between peaks was greater than 500 ms [18]. After the labeling of GSR peaks, a GSR peaks/minute metric was computed. [18].

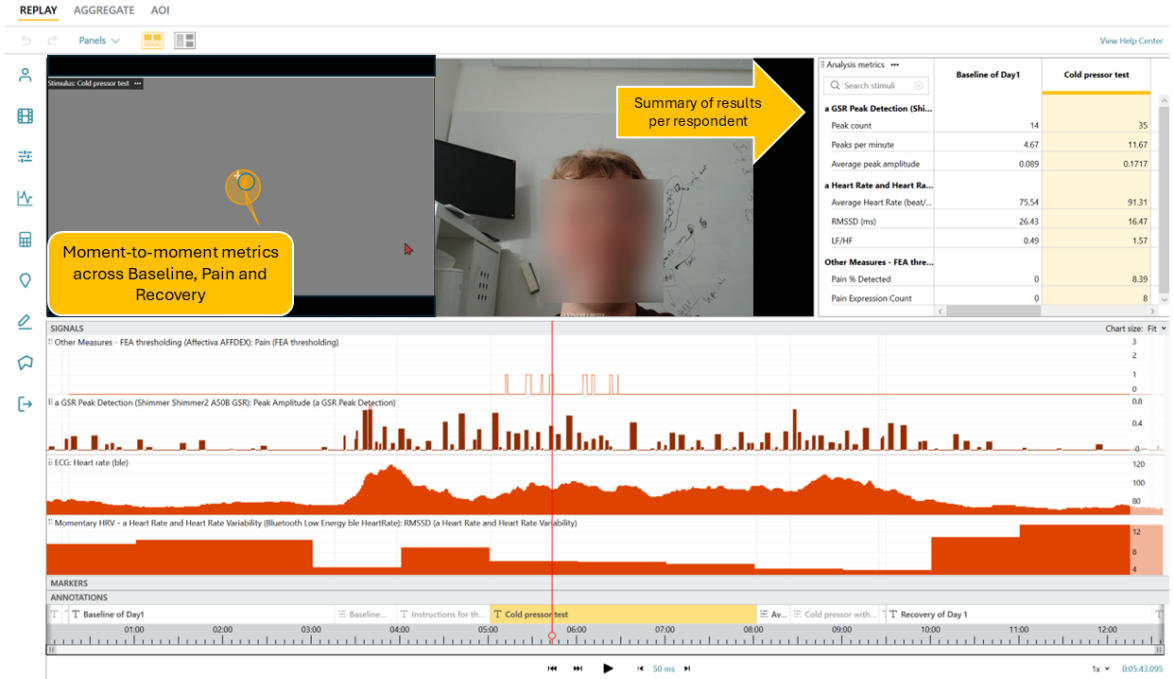


Figure 3: Automatic synchronization of eye-tracking, automated facial expression analysis, GSR, ECG and blink rate inside the iMotions software platform. At the top center, the respondent camera shows the respondent while doing the Cold Pressor Test. At the top left, the stimulus is shown with the overlaid fixations in orange. The bottom (Signals section) shows the automated synchronization of biosensors and the computation of metrics (in this specific example the Pain metric, the GSR peak amplitude, heart rate values as well as the momentary time-based heart rate variability are displayed over Baseline of Day 1, the Cold Pressor Test and Recovery of Day 1).

ECG. ECG activity was obtained through a wireless band (Polar Belt H10, Polar, Switzerland) applied on the chest. Heart rate was calculated from the R-R interval signal and HRV was computed from the R-R intervals, wherein an advanced detrending algorithm was applied based on the smoothness priors approach to obtain stationary signals. The selected smoothing parameter λ was set to 500 which corresponds to a cut-off frequency of 0.035 Hz. HRV was expressed using the root mean square of successive differences (RMSSD) in the time domain [18], and for recordings longer than 64 seconds, also using the low-frequency to high-frequency (LF/HF) ratio in the frequency domain. The synchronization of the ECG signals alongside the Pain metric, the GSR peak amplitude metric and eye tracking, is displayed in Figure 3.

Eye-tracking. Eye-tracking was performed using a professional infrared eye-tracker positioned below a 24-inch monitor (Aurora, Smart Eye, Sweden) to ensure participants gazed at the fixation crosses during the various stages of the experiment and to compute blink rate. Fixations were classified using a velocity-based I-VT filter (iMotions 11.12, iMotions A/S, Denmark) set at $30^\circ/\text{s}$. No additional filtering or cleaning parameters were added by the experimenters. Blinking rate was flagged whenever eyelid opening dropped below 50% of the median eyelid opening for the entire recording. Eyelid samples with a quality score below 0.1 were excluded from the analysis beforehand, while blinks lasting longer than 500 ms were discarded, given that genuine blinks are usually below that upper limit.

2.3 Statistics

Signal processing for facial expressions analysis, GSR, ECG and blink rate was conducted using the signal processing R-library in iMotions (iMotions 11.12, iMotions A/S, Denmark). Statistical analyses were performed in Python using the SciPy library (Python Software Foundation, Python Version 3.14, <https://www.python.org/>). The threshold for statistical significance was set at $p < 0.05$. All data were assessed for normality by the Shapiro-Wilk test. Linear mixed models were applied separately for Day 1 analysis (with the factor TIME: Day1 Baseline, Day1 Cold Pain, Day1 Recovery) and for Day 3 (with the factor TIME: Day3 Baseline, DOMS, Day3 Recovery), with p-values adjusted by Bonferroni correction for multiple comparisons. Additionally, using only heart rate, GSR peaks/minute, GSR peak amplitude and blink rate, a temporal analysis was conducted using linear mixed models on the 3-minute Cold Pressor Test (with the factor TIME: first minute, second minute and third minute) and during the DOMS weight-lifting task (with the factor TIME: first third, second third and last third), with p-values also adjusted by Bonferroni correction. That temporal analysis excluded time-based and frequency-based heart rate variability metrics because they often require longer time windows. The temporal evolution of expressed pain using the Pain metric was analyzed using the non-parametric Friedman test on repeated measures because of the zero-inflated dataset, which violates the residual normality assumption, followed by Wilcoxon signed-rank tests and Nemenyi post-hoc comparisons. Results are reported as mean (SD, standard deviation of the mean) in tables and mean (SEM, standard error of the mean) in plots. Confusion matrices were produced to illustrate the classification performance of the Pain metric, relative to the ground truth labeling (human facial action coding).

3 Results

3.1 Primary analysis: Facially expressive participants

For ground-truth analysis, all video recordings were visually inspected and manually reviewed to identify the onset and offset of pain-expressive (facially expressive) events based on PSPI-associated facial Action Units. The main goal of this validation study was to distinguish moderate-to-intense pain expressions from no-pain or subtle expressions; an event was labeled as pain when two to four components of the PSPI rule were simultaneously activated. The four PSPI-rule components are based on the 6 facial action units mentioned in the Methods section of the current work and in multiple scientific publications: AU4, **max**(AU6, AU7), **max**(AU9, AU10), and eye closure (AU43). Activation of only a single component was considered a subtle expression on its own, as these patterns can also occur in many non-pain expressions. In contrast, moderate-to-high intensity pain expressions were typically characterized by the simultaneous activation of multiple PSPI components. During the **Cold Pressor Test**, 51.3% of the participants (N=19 out of 37 participants) exhibited pain-related facial expressions and hence were classified as “**facially expressive participants**”, while the remaining 48.7% (N=18) did not show pain-related facial reactions at all and hence were classified as “**facially inexpressive participants**”. Similarly, during **DOMS**, 41.9% of the participants (N=13 out of 31 participants) showed pain-related facial expressions, while the remaining 58.1% (N=18 out of 31) did not show any pain-related facial expressions.

3.1.1 Performance of the classification

Confusion matrices and overall performance are reported at two levels of analysis: a macro level, where a detection counts as correct if the model identifies at least one genuine pain facial reaction

for that person within a recording, and a per-activation level, where each individual facial reaction is evaluated on its own.

Ground Truth	Actual Positive	18	1	10	3	28	4
	Actual Negative	1	17	2	16	3	33
		Predicted pain		Predicted not painful		Predicted pain	
		Predicted		Predicted		Predicted	

Figure 4: At a macro level, confusion matrices of a) the Cold Pressor Test, b) DOMS (muscle soreness), and c) both Cold Pressor Test and DOMS together.

At a macro level: Relative to all manually labeled sessions, the Cold Pressor Test shows an accuracy of 0.946, a precision of 0.947, a recall of 0.947, and an F1 of 0.947 (Figure 4a), while DOMS shows an accuracy of 0.839, a precision of 0.833, a recall of 0.769 and an F1 of 0.800 (Figure 4b). The pooled analysis of both pain models shows an accuracy of 0.897, a precision of 0.903, a recall of 0.875 and an F1 of 0.889 (Figure 4c).

At a per activation level: Relative to all manually labeled pain sessions, the CPT shows an accuracy of 0.984, a precision of 0.746, a recall of 0.707, and an F1 of 0.726 (Figure 5a), while DOMS shows an accuracy of 0.977, a precision of 0.870, a recall of 0.674 and an F1 of 0.760 (Figure 5b).

Ground Truth	Actual Positive	3912	1623	2371	1145	6283	2768
	Actual Negative	1332	182842	354	61797	1686	244639
		Predicted pain		Predicted not painful		Predicted pain	
		(a)		(b)		(c)	

Figure 5: At a per activation level: Including all video frames, confusion matrices of a) the Cold Pressor Test, b) DOMS (muscle soreness), and c) both Cold Pressor Test and DOMS together.

The pooled analysis of both pain models yields an accuracy of 0.983, F1 of 0.738, precision of 0.788, recall of 0.694, and an ROC-AUC of 0.961 (Figure 5c), demonstrating consistency with findings from other facial expression analyses.

Collectively, these findings demonstrate that the pain metric accurately identifies pain in expressive participants and reveals that pain expression is not continuous throughout the entire pain exposure.

3.1.2 Self reported pain of facially expressive respondents

Immediately after completion of the Cold Pressor Test and DOMS assessments, respondents were prompted to rate their perceived pain using a continuous 0–10 numerical rating scale (NRS) using

the Advanced Surveys in iMotions.

Table 1: Mean ($\pm$ SD) self-reported pain during the Cold Pressor Test and DOMS in the facially expressive respondents.

	Cold Pressor Test		DOMS	
	N	Self-reported Pain NRS scale	N	Self-reported Pain NRS scale
Facially expressive respondents (Ground truth)	19 out of 37	5.87 ± 2.23	13 out of 31	6.29 ± 1.54

3.1.3 Physiological findings of facially expressive respondents

Main findings in the Cold Pressor Test of facially expressive respondents Linear Mixed Effects Models analysis revealed that the average heart rate, frequency-based heart rate variability (LF/HF), GSR peaks/minute, average GSR peak amplitude and blink rate significantly increased during the Cold Pressor Test, compared with the pain-free stages (Figure 6). These results suggest an upregulation of autonomic responses and provide comprehensive insights into the evolution of a diverse array of responses to an acute noxious stressor.

Temporal evolution during the Cold Pressor Test of facially expressive respondents

The temporal evolution of the physiological responses and the self-reported pain during the Cold Pressor Test was divided into three 1-minute intervals (first minute, second minute and third minute). Linear mixed effect models revealed reduced heart rate during the third minute, relative to both the second and first minutes, reduced GSR peaks/minute during the third minute relative to the first minute and reduced GSR peak amplitude during the second and the third minute relative to the first minute (Figure 7), consistent with sympathetic habituation.

Given that the temporal analysis assesses short time windows, both time-based and frequency-based heart rate variability metrics were not considered a priori for the analysis because these metrics normally require longer recordings.

As mentioned in the Methods section, the current work selected a Pain metric threshold = 50 in order to binarize it. In this context, the iMotions Lab software can output binarized facial expressions metrics (including the Pain metric) either as the number of frames in which the continuous signal surpassed that threshold (e.g. Pain frames $\geq$ threshold) or as a Time Percentage (e.g. Pain Time Percentage). Friedman tests indicated a decrease in duration of expressed pain as indexed by both Pain frames $\geq$ threshold and as Pain Time Percentage during the third minute of the Cold Pressor Test, relative to the second minute. This is consistent with the physiological findings.

Linear mixed models analysis on self-reported pain intensity exhibited no significant changes across the 3-minute Cold Pressor Test (Figure 8).

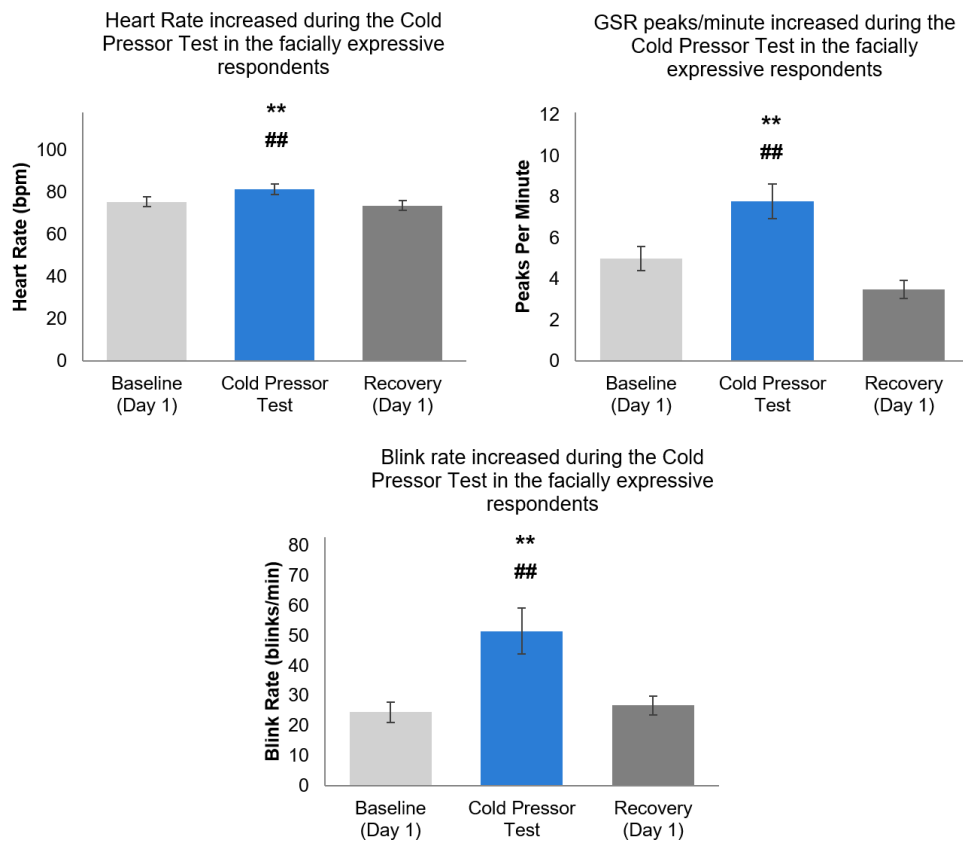


Figure 6: Physiological and blink rate outcomes that were significantly modulated across the Baseline, the Cold Pressor Test, and Recovery sessions in the facially expressive respondents: Heart rate, GSR peaks/minute, and blink rate significantly increased during the Cold Pressor Test, as opposed to both the Baseline (**, $p < 0.01$) and Recovery (##, $p < 0.01$) stages.

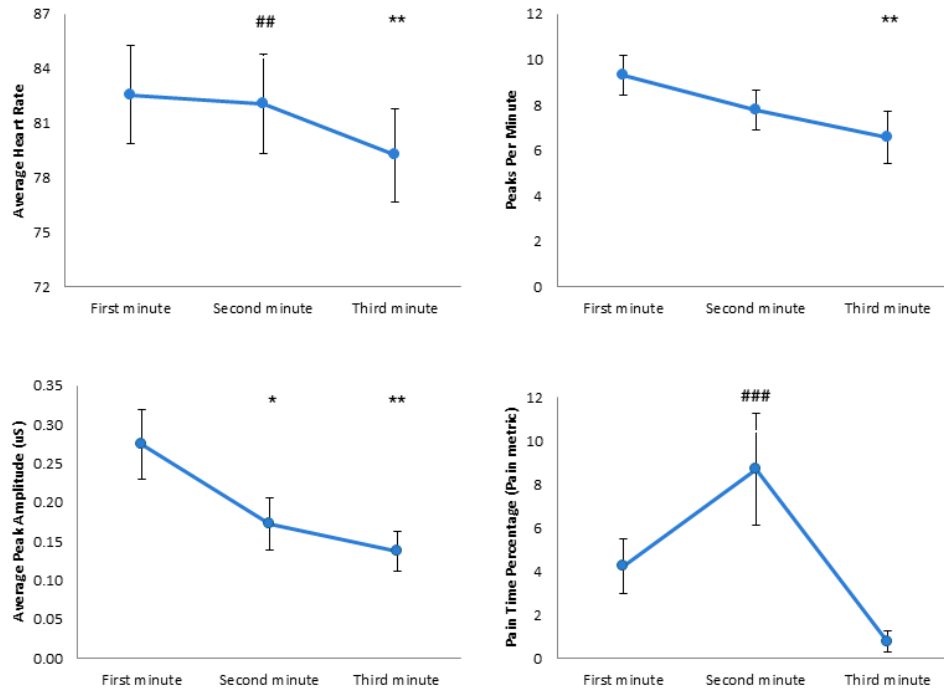


Figure 7: The 3-minute Cold Pressor Test significantly reduced autonomic responses in the facially expressive group as indexed through heart rate, GSR peaks/min and GSR peak amplitude, relative to the first minute (*, $p < 0.05$; **, $p < 0.01$) and to the third minute (##, $p < 0.01$; ###, $p < 0.001$). The Pain time percentage (Pain metric) significantly decreased during the third minute, relative to the second minute (###, $p < 0.001$).

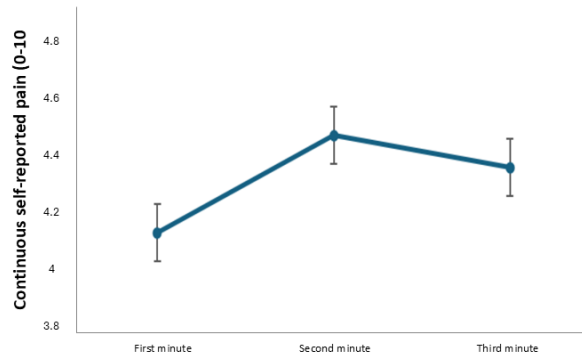


Figure 8: No significant changes in self-reported pain across the 3-minute Cold Pressor Test.

DOMS in the facially expressive respondents Linear mixed effect models revealed that heart rate ($p < 0.001$), GSR peaks/minute ($p < 0.001$), GSR peak amplitude ($p < 0.01$) and blink rate ($p < 0.01$) significantly increased during DOMS, relative to both Baseline and Recovery, while RMSSD (heart rate variability) decreased ($p < 0.05$) during DOMS, relative to Recovery. These findings suggest that movement-evoked pain produces a consistent, convergent response across all modalities in the facially expressive respondents.

Temporal evolution during DOMS of facially expressive respondents The temporal evolution of DOMS was split into 3 substages: first, second and final third, each lasting approxi-

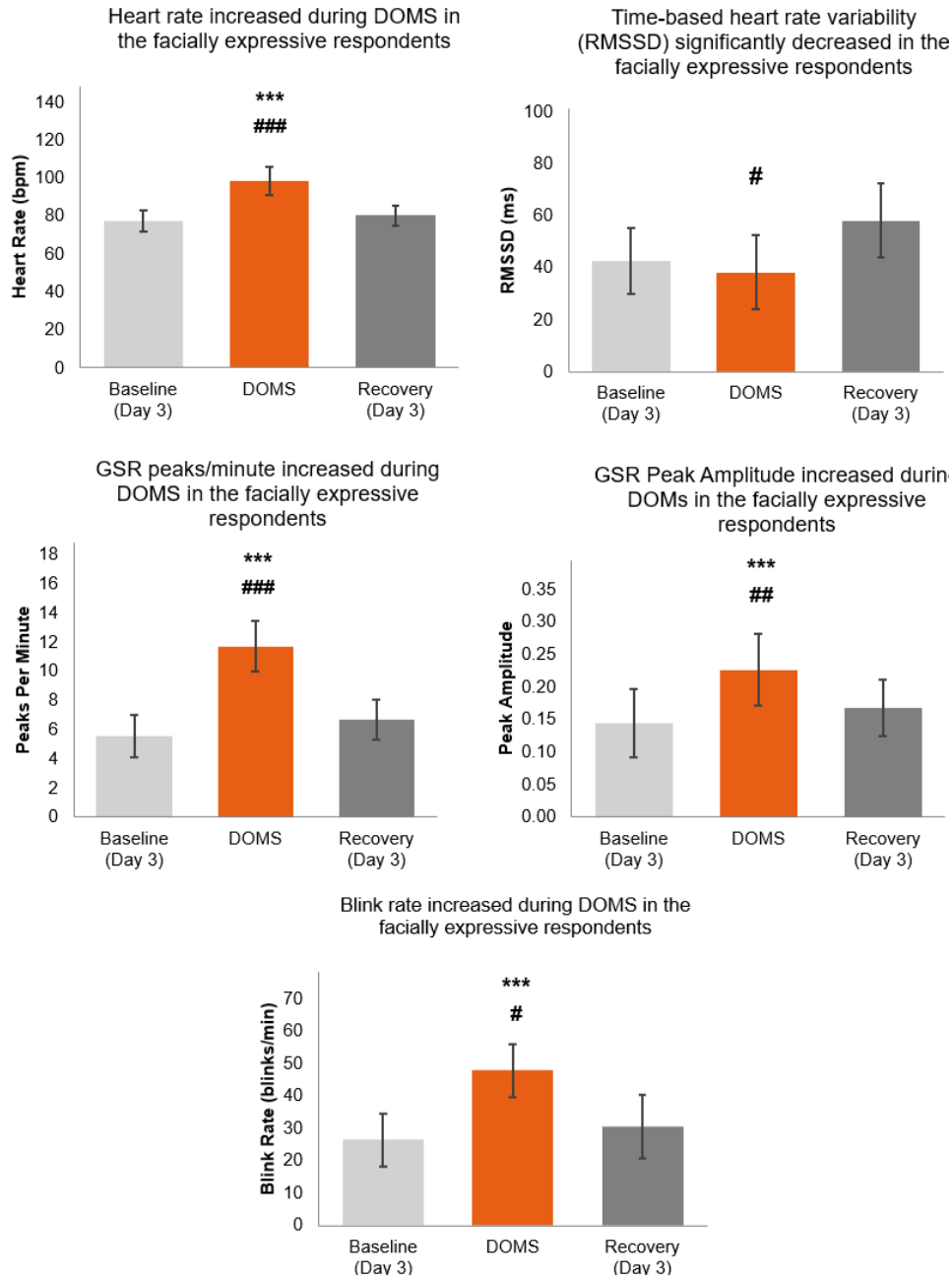


Figure 9: A significant increase in several physiological outcomes while weight lifting during DOMS relative to both Baseline (***, $p < 0.01$) and Recovery (#, $p < 0.05$; ##, $p < 0.01$; ###, $p < 0.001$): Heart rate, GSR peaks/minute, GSR peak amplitude and blink rate. As expected, RMSSD (heart rate variability) significantly decreased during movement-evoked pain.

mately 25 seconds.

In this sample size, linear mixed effect models showed that heart rate significantly increased during the second third of DOMS ($p = 0.011$), compared with the first one, while GSR peaks/minute marginally increased ($p = 0.071$) during the second third too, relative to the first one. Friedman tests revealed a significant increase of the Pain time percentage (%) during the last third, relative to the first one.

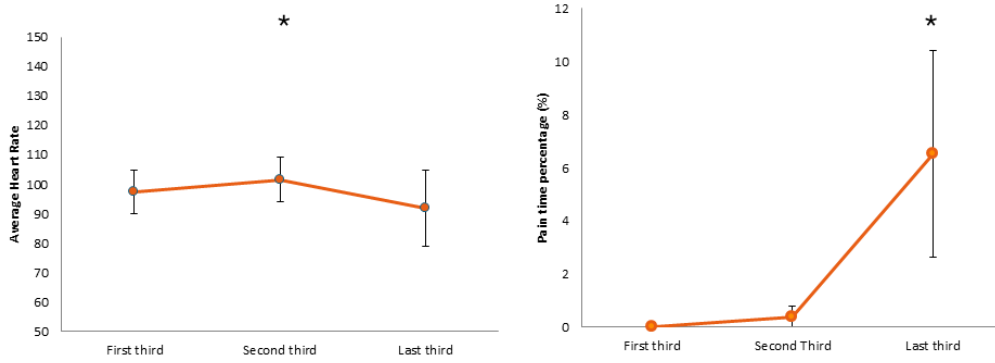


Figure 10: Temporal evolution during movement-evoked pain due to DOMS: Relative to the first third, heart rate significantly increased during the second third and the Pain time percentage (%) significantly increased during the last third (*, $p < 0.05$) in the facially expressive respondents.

3.2 Secondary analysis: inexpressive respondents

This section shows the results of self-reported pain and physiological findings considering only the facially inexpressive respondents.

3.2.1 Self-reported pain of inexpressive respondents

Table 2: Mean ($\pm$ SD; standard deviation of the mean) self-reported pain of the Cold Pressor Test and the DOMS of the inexpressive respondents.

	Cold Pressor Test		DOMS	
	N	Self-reported pain (NRS) scale	N	Self-reported pain in NRS scale
Inexpressive respondents (Ground truth)	18 out of 37 respondents	5.41 ± 2.21	18 out of 31 respondents	5.54 ± 2.56

The average self-reported pain of the Cold Pressor Test and DOMS is shown in Table 2.

3.2.2 Physiological findings in inexpressive respondents

Considering only the facially inexpressive respondents during the Cold Pressor Test (N=18), heart rate significantly increased during the Cold Pressor Test, relative to both Day1 Baseline and Day1 Recovery. The frequency-based heart rate variability (LF/HF) and average GSR peak amplitude results were marginally significant in this sample size after Bonferroni correction.

Considering solely the facially inexpressive respondents during DOMS (N=18), heart rate, LF/HF, blink rate and GSR peaks/minute significantly increased during DOMS, relative to Day 3 Baseline and/or Day 3 Recovery.

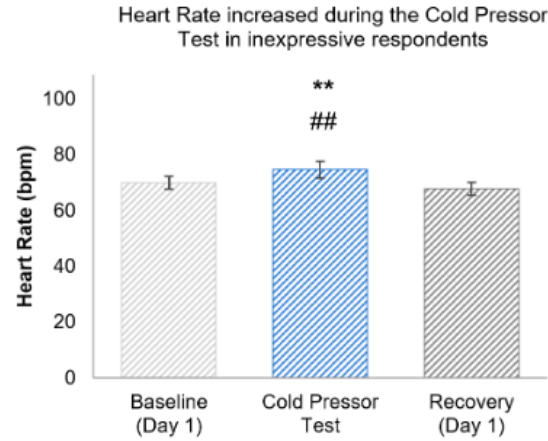


Figure 11: When considering only those who were facially inexpressive, heart rate significantly increased during the Cold Pressor Test, relative to both Baseline (**, $p < 0.01$) and Recovery (##, $p < 0.01$). Bonferroni correction applied.

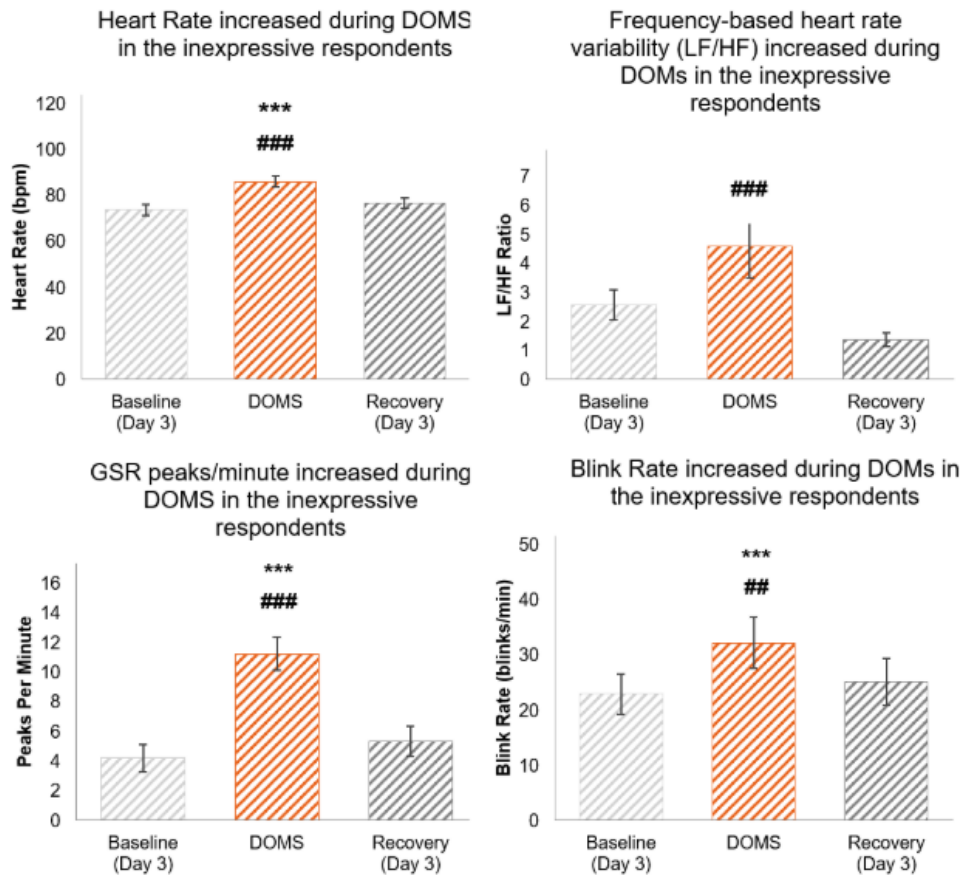


Figure 12: Statistical findings when considering only those who were facially inexpressive during DOMS. Heart rate, GSR peaks/minute and blink rate significantly increased during DOMS, relative to both Baseline (***, $p < 0.001$) and Recovery (##, $p < 0.01$; ###, $p < 0.001$). Bonferroni correction applied.

4 Discussion

The present work validated the performance of the Pain metric in the iMotions platform to detect expressed pain using computer vision algorithms through established and scientifically valid pain

models such as the Cold Pressor Test (CPT) and Delayed Onset Muscle Soreness (DOMS) on upper-limb muscles.

4.1 Key Findings & Convergence

The notable observations include:

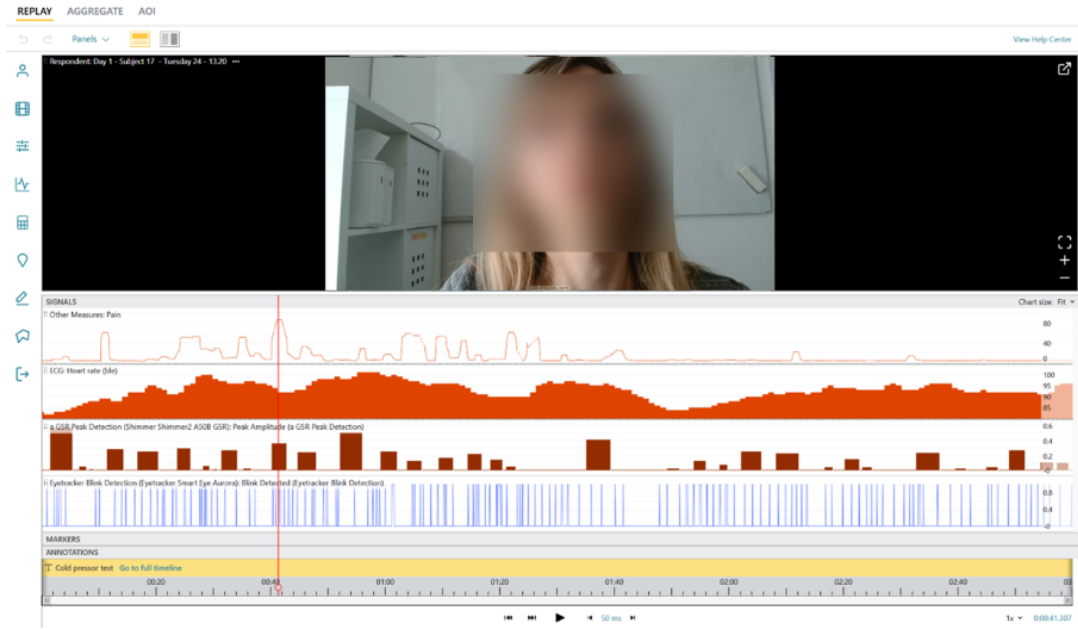
- **Facial expression-based pain detection:** The current facial expressions based pain metric detects pain-related facial expressions in accordance with the 6 facial action units that are part of the PSPI scale: brow lowerer (also called brow furrow, AU4), cheek raiser and lid tightener (AU6 and AU7), nose wrinkler and upper-lip raiser (AU9 and AU10), and eyes closed (AU43). This offers an easy to use, high temporally resolved and non-invasive method to detect pain in various settings.
- **Classification performance:** The classification performance is comparable to or higher than other approaches using internal datasets comprising various types of chronic pain and acute pain [32], as well as post-surgical pain [23, 7], and public datasets [28, 31, 14, 15, 16].
- **Convergence:** The analysis revealed a convergence across all modalities, with most physiological measures being altered during CPT and DOMS relative to the baseline and/or recovery stages.
- **Cross-validation:** Specifically in purely pain-based studies, the facial expressions based pain metric can cross-validate GSR and ECG findings, given that general sympathetic arousal measures cannot unambiguously isolate or distinguish pain from stress, effort, or cognitive workload, among other constructs without additional context.

Table 3: ROC-AUC of various machine learning and deep learning approaches for pain detection.

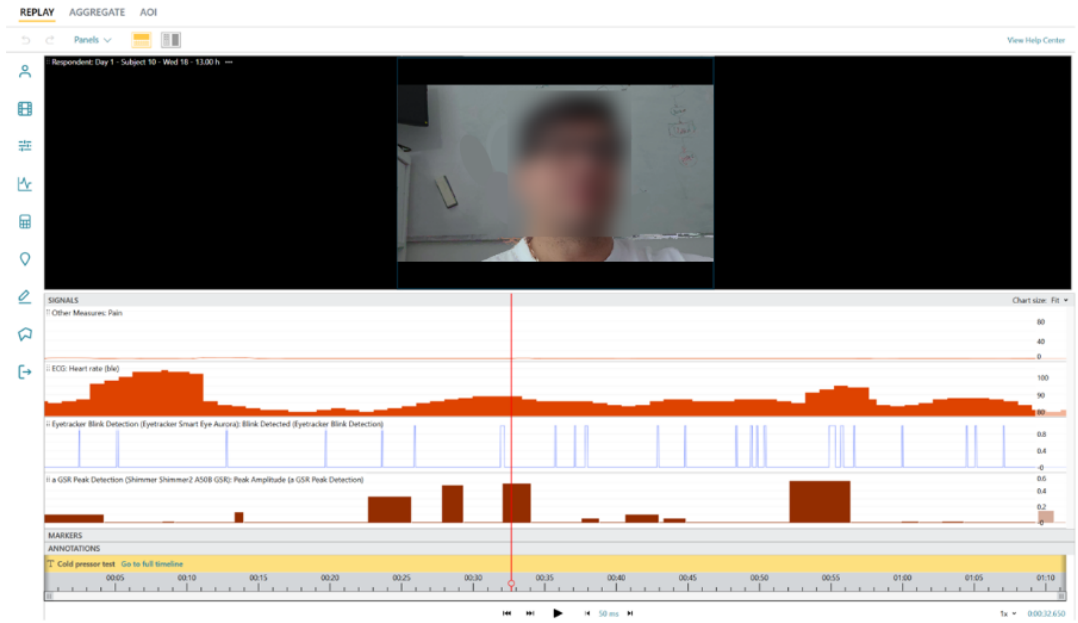
Scientific publications	Year of publication	Dataset	ROC-AUC
Lucey et al. [14]	2009	UNBC	78.4%
Lucey et al. [16]	2011		83.9%
Lucey et al. [15]	2011		84.7%
Zafar & Khan [31]	2014		87.4%
Rodriguez et al. [28]	2017		93.3%
Park et al. [23]	2024	Internal datasets	85–96%
Gutierrez et al. [7]	2024		94%
Zhang et al. [32]	2025		90%
iMotions Pain metric	2026		96.1%

Alongside facial expression analysis, the current work evaluated the impact of pain on physiological responses such as heart rate (ECG), heart rate variability, galvanic skin response (GSR), also referred to as electrodermal activity (EDA), and spontaneous blink rate during both short-lasting pain (Cold Pressor Test) and long-lasting delayed-onset muscle soreness paradigms.

The current findings of ECG, GSR, and blink rate add to the existing literature reporting increased heart rate, altered time-based and frequency-based heart rate variability, as well as increased spontaneous blink rate during pain. Specifically concerning the Cold Pressor Test,



(a) Facially expressive respondent while lifting a weight during pain.



(b) Facially inexpressive respondent while lifting a weight during pain.

Figure 13: Individual replay of a facially expressive respondent (top) and inexpressive respondent (bottom) while lifting a weight during pain in the iMotions software. In studies specifically focused on pain, ECG and GSR responses can help confirm a painful reaction occurred even in facially inexpressive respondents.

facially expressive respondents showed increased heart rate [2, 20, 29, 4], GSR peak amplitude and GSR peaks/minute [8], whereas inexpressive respondents showed primarily increased heart rate. The inexpressive respondents also showed comparable statistical effect sizes to those observed in the facially expressive group for the time-based and frequency-based heart rate variability metrics as well as the GSR metrics but the findings were marginally significant in this sample size. When

pooling the results of both facially expressive and inexpressive respondents together, the results are consistent with previous literature.

Regarding DOMS paradigms, both facially expressive and inexpressive respondents showed comparable results in terms of GSR, ECG and spontaneous blink rate responses. Not only higher heart rate and altered heart rate variability [12] was observed but also higher galvanic skin response, relative to the pain-free stages. These changes are essentially connected to an upregulated sympathetic response and/or reduced parasympathetic (vagal) activity in the 24 to 72 hour window after DOMS induction. Furthermore, these findings demonstrate that these outcomes remain consistent across both facially expressive and inexpressive individuals. This suggests that this physiological approach is suitable for all types of respondents, regardless of their level of facial expressivity.

Current spontaneous blink rate results during the Cold Pressor Test and movement-evoked pain after DOMS induction expand previous results in the scientific literature showing higher spontaneous blink rate during pain, relative to pain-free stages. Previous work reported increased spontaneous blink rate during both experimental pain e.g. noxious withdrawal reflex, Cold Pressor Test, as well as electrical [5] and laser stimulation [22] and clinical chronic pain [30]. Spontaneous blink rate is regarded as a proxy of central dopaminergic activity and that may in turn contribute to the modulation of nociceptive processing and integration. It has been suggested that such blink-related pain modulation may rely on dopaminergic receptors expressed in the periaqueductal gray [30], the topmost area of the brainstem, which is involved in pain modulation.

The temporal evolution analysis of the Cold Pressor Test exhibited a slowing heart rate and a lower GSR response in both amplitude (average GSR peak amplitude) and frequency (GSR peaks/minute) consistent with a de-arousal or habituation response throughout the 3-minute period. These significant effects indicate not only the absence of an escalating pain response but also confirm the steady lack of sensations or numbness reported by respondents. Future studies can investigate if the use of the water cooling system during the Cold Pressor Test can prevent this effect [2, 19]. Conversely, blink rate did not significantly change across those 3 minutes, suggesting that blink dynamics and cardiovascular and electrodermal arousal may index overlapping but distinct processes during sustained pain. On the other hand, the temporal evolution of DOMS showed a higher heart rate during the second third, relative to the first one, whereas the last third did not show statistically significant differences, relative to the other thirds. Please note that since the temporal analysis uses short time windows both time-based and frequency-based heart rate variability metrics were not considered for the analysis since it normally requires long recordings.

4.2 Application areas

The facial expression based pain metric is a validated, easy-to-use, non-invasive, remote tool to detect facially expressed pain. It enables researchers to accurately detect expressed pain from a camera feed of the respondent. This makes it a powerful tool for many high-value applications, particularly when combined with continuous physiological monitoring. Given the intersubject variability in how individuals express pain facially, it is important that researchers operationalize and define both the constructs under study and the relevant metrics connected to that. Some of the research areas that the Pain metric can be applied are:

1. **Occupational health and ergonomics:** The pain metric can be used to measure overexertion during physical tasks without workflow interruption.
2. **Human factors:** Expressed pain can add valuable insights to continuous monitoring in high stress positions and tasks.

3. **Sports science:** Tracking pain-linked events during rehabilitation and exercise protocols.
4. **Telehealth research:** Leveraging existing camera-based signals for remote monitoring in research settings.
5. **Data Fusion for pain assessment:** Deep learning or machine learning evaluations allow fusing multimodal data to detect pain by combining the Pain metric (derived from facial expressions) with ECG and GSR responses, to name a few.

As is evident from the inter-individual variability in expressed pain presented in the results above, researchers are advised to use caution when working with specific patient populations or using the metric as a screening tool in research projects. Inter individual differences can be altered in non-verbal or communication-impaired populations (e.g., ICU, dementia, or infants). These require further investigation with dedicated studies.

Research indicates that approximately 33% to 38% of people are naturally low or inexpressive even under controlled laboratory conditions [9, 11]. This trait level inexpressiveness is further compounded by cultural display rules, where social contexts [10] or the presence of the experimenter trigger the neutralization or masking of negative emotions. Additionally, gender linked expectations of inexpressiveness can create further variability in how pain is externalized. These individual differences in display rules, cultural context and training will all influence expressed pain and therefore the detection possible by the pain metric. Researchers should therefore use their discretion when drawing conclusions based on experienced pain levels, and ideally run multimodal studies combining implicit measurements from physiological measures such as ECG.

As seen in the present study, 50% of participants were inexpressive. Owing to the reasons discussed in this section, this can be expected from other real-world studies, with camera aware participants. This creates a significant challenge to facial classifiers, which rely on expressed emotion to observe and categorize it as pain. Adding autonomic measures such as cardiovascular and galvanic skin responses that are involuntary and continue to track pain even when expressed behavior is muted, can help researchers to have measurements of pain even from inexpressive respondents. Autonomic measures can also be used to distinguish participants that lack pain and those displaying intentional or trait based masking. As seen in the sections above, subgroup analysis can be carried out to explore and understand pain, a construct high on interindividual differences.

5 Summary and Guidance for End Users

The Pain metric is designed to be interpretable and adaptable across a wide range of research settings. It outputs a continuous probability score ranging from 0 to 100 for each video frame, reflecting the degree to which pain-related facial activity is present. Users can adjust the detection threshold to binarize its output.

5.1 Adjustable detection threshold

By default, the Pain metric applies a binarization threshold of 50, classifying frames above this value as expressing clear pain and frames below it as not expressing pain. This threshold was selected in the present validation study because it can distinguish moderate-to-intense pain expressions from subtle or absent ones, and it underlies high classification performance as reported in Section 4.1.

Using iMotions Lab (iMotions 11.12 and onward), users can adjust the threshold to fit the sensitivity and specificity requirements of their specific use case:

- **Lowering the threshold** increases sensitivity, capturing more subtle facial reactions, but raises the likelihood of false positives from expressions that resemble but do not reflect genuine pain (e.g. concentration, effort, or squinting).
- **Raising the threshold** increases specificity and reduces false positives, at the cost of missing subtler or lower-intensity pain expressions, particularly in respondents with naturally low facial expressivity.

5.2 Addressing facial inexpressivity

Given that a substantial proportion of individuals - approximately 33–50% in both the general population and the present sample - show limited or no facial expression of pain even when physiological pain responses are present, the Pain metric should not be used as a sole indicator of pain absence. A low or zero Pain score reflects the absence of *expressed* facial pain behavior, not necessarily the absence of the *experience* of pain.

For applications where distinguishing true pain-free states from masked or inexpressive pain matters, users can choose to reduce the Pain metric threshold (available in the iMotions software 11 and onwards) and/or pair the Pain metric with concurrent physiological measures such as GSR and/or ECG. Specifically for pain-related studies, these measures remain sensitive to pain even in facially inexpressive respondents, providing a valuable cross-validation layer and reducing the risk of misclassifying inexpressive individuals as pain-free.

References

- [1] M. Bishay, K. Preston, M. Strafuss, G. Page, J. Turcot, and M. Mavadati. Affdex 2.0: A real-time facial expression analysis toolkit. In *2023 IEEE 17th international conference on automatic face and gesture recognition (FG)*, pages 1–8. IEEE, 2023.
- [2] T. Bullock, M. H. MacLean, T. Santander, A. P. Boone, V. Babenko, N. M. Dundon, A. Stuber, J. Jimmons, M. Raymer, N. Okafor, M. B. Miller, B. Giesbrecht, and S. T. Grafton. Habituation of the stress response multiplex to repeated cold pressor exposure. *Frontiers in Physiology*, 13:752900, 2023.
- [3] K. D. Craig, R. V. E. Grunau, and S. M. Branson. Age-related aspects of pain: Pain in children. In R. Dubner, G. F. Gebhart, and M. R. Bond, editors, *Pain Research and Clinical Management*, volume 3, pages 317–328. Elsevier Science, Amsterdam, 1988.
- [4] A. Daradas, S. Kulthinee, T. Promsrisuk, P. Kesornwanichwattana, P. Thaengkathok, S. Pongampai, P. Kongjaidee, N. Seeja, M. Poomvanicha, and P. Chotimol. Cardiovascular and ocular parameter alterations in response to cold pressor test in young adults. *Diagnostics*, 14(18):2010, 2024.
- [5] R. Duranti, R. Galletti, and T. Pantaleo. Relationships between characteristics of electrical stimulation, muscle pain and blink responses in man. *Electroencephalography and Clinical Neurophysiology*, 55(6):637–644, 1983.
- [6] A. García-Valverde, P. Asencio, J. L. Hernández-Davó, and R. Sabido. Effects of four consecutive one-repetition maximum testing days on neuromuscular performance, muscle soreness and perceived recovery. *Retos*, 52:565–570, 2024.
- [7] R. Gutierrez, J. Garcia-Ortiz, and W. Villegas-Ch. Multimodal ai techniques for pain detection: integrating facial gesture and paralinguistic analysis. *Frontiers in Computer Science*, 6:1424935, 2024.
- [8] R. K. Jha, M. S. Amatya, O. Nepal, M. M. Bade, and M. M. K. Jha. Effect of cold stimulation induced pain on galvanic skin response in medical undergraduates of kathmandu university school of medical sciences (kusms). *Int J Sci Healthcare Res*, 2:55–58, 2017.
- [9] A. J. Karmann, S. Lautenbacher, and M. Kunz. The role of inhibitory mechanisms in the regulation of facial expressiveness during pain. *Biological Psychology*, 104:82–89, 2015.

- [10] M. Kunz, N. Faltermeier, and S. Lautenbacher. Impact of visual learning on facial expressions of physical distress: A study on voluntary and evoked expressions of pain in congenitally blind and sighted individuals. *Biological Psychology*, 89(2):467–476, 2012.
- [11] M. Kunz, V. Mylius, K. Schepelmann, and S. Lautenbacher. On the relationship between self-report and facial expression of pain. *The Journal of Pain*, 5(7):368–376, 2004.
- [12] G. Langmár, T. Sümegi, B. Fülöp, L. Pozsgai, T. Mocsai, M. Tóth, L. Rácz, B. Kopper, A. Dér, A. Búzás, and B. Sonkodi. Heart rate variability alterations during delayed-onset muscle soreness-inducing exercise—with piezo2 interpretation. *Sports*, 13(8):262, 2025.
- [13] L. LeResche and S. F. Dworkin. Facial expression accompanying pain. *Social Science & Medicine*, 19(12):1325–1330, 1984.
- [14] P. Lucey, J. F. Cohn, S. Lucey, I. Matthews, S. Sridharan, and K. M. Prkachin. Automatically detecting pain using facial actions. In *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1–8. IEEE, 2009.
- [15] P. Lucey, J. F. Cohn, I. Matthews, S. Lucey, S. Sridharan, J. Howlett, and K. M. Prkachin. Automatically detecting pain in video through facial action units. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 41(3):664–674, 2011.
- [16] P. Lucey, J. F. Cohn, K. M. Prkachin, P. E. Solomon, and I. Matthews. Painful data: The unbc-mcmaster shoulder pain expression archive database. In *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pages 57–64. IEEE, 2011.
- [17] Y. Matsuda, S. Kan, H. Uematsu, M. Shibata, and Y. Fujino. Pain-related brain activity evoked by active and dynamic arm movement: delayed-onset muscle soreness as a promising model for studying movement-related pain in humans. *Pain medicine*, 16(8):1528–1539, 2015.
- [18] M. Minen, L. Gregoret, D. Seernani, and J. Wilson. Systematic evaluation of driver’s behavior: A multimodal biometric study. In *HCI International 2023 Posters*, pages 57–64, Cham, 2023. Springer.
- [19] N. Minkley, T. P. Schröder, O. T. Wolf, and W. H. Kirchner. The socially evaluated cold-pressor test (secpt) for groups: Effects of repeated administration of a combined physiological and psychological stressor. *Psychoneuroendocrinology*, 45:119–127, 2014.
- [20] L. Mourot, M. Bouhaddi, and J. Regnard. Effects of the cold pressor test on cardiac autonomic control in normal subjects. *Physiological Research*, 58(1):83–91, 2009.
- [21] H. Nie, A. Kawczynski, P. Madeleine, and L. Arendt-Nielsen. Delayed onset muscle soreness in neck/shoulder muscles. *European Journal of Pain*, 9(6):653–660, 2005.
- [22] G. Paparella, G. Di Stefano, A. Fasolino, G. Di Pietro, D. Colella, A. Truini, G. Cruccu, A. Berardelli, and M. Bologna. Painful stimulation increases spontaneous blink rate in healthy subjects. *Scientific Reports*, 10:20014, 2020.
- [23] I. Park, J. H. Park, J. Yoon, I.-A. Song, H.-S. Na, J.-H. Ryu, and A.-Y. Oh. Artificial intelligence model predicting postoperative pain using facial expressions: a pilot study. *Journal of Clinical Monitoring and Computing*, 38(2):261–270, 2024.
- [24] C. J. Patrick, K. D. Craig, and K. M. Prkachin. Observer judgements of acute pain: Facial action determinants. *Journal of Personality and Social Psychology*, 50(6):1291–1298, 1986.
- [25] K. M. Prkachin and K. D. Craig. Influencing non-verbal expressions of pain: Signal detection analyses. *Pain*, 21(4):399–409, 1985.
- [26] K. M. Prkachin and Z. Hammal. Computer mediated automatic detection of pain-related behavior: Prospect, progress, perils. *Frontiers in Pain Research*, 2:788606, 2021.
- [27] S. N. Raja, D. B. Carr, M. Cohen, N. B. Finnerup, H. Flor, S. Gibson, F. Keefe, J. S. Mogil, M. Ringkamp, K. A. Sluka, et al. The revised iasp definition of pain: concepts, challenges, and compromises. *Pain*, 161(9):1976, 2020.
- [28] P. Rodriguez, G. Cucurull, J. González, J. M. Gonfaus, K. Nasrollahi, T. B. Moeslund, and F. X. Roca. Deep pain: Exploiting long short-term memory networks for facial expression classification. *IEEE Transactions on Cybernetics*, 52(5):3314–3324, 2022.
- [29] R. Sebastião, A. Bento, and S. Brás. Analysis of physiological responses during pain induction. *Sensors*, 22(23):9276, 2022.

- [30] E. Stefanescu, N. Pancino, C. Graziani, V. Lachi, M. Sampoli, G. Dimitri, A. Bargagli, D. Zanca, M. Bianchini, D. Muresanu, and A. Rufa. Blinking rate comparison between patients with chronic pain and parkinson’s disease. In *Abstracts of the 8th Congress of the European Academy of Neurology*, volume 29, page 669. Wiley, 2022.
- [31] Z. Zafar and N. A. Khan. Pain intensity evaluation through facial action units. In *2014 22nd International Conference on Pattern Recognition*. IEEE, 2014.
- [32] J.-T. Zhang, X.-Y. Hu, W. Duan, M.-H. Ji, and J.-J. Yang. Application of deep learning-based facial pain recognition model for postoperative pain assessment. *Journal of Clinical Anesthesia*, 105:111898, 2025.

6 APPENDIX

The following section contains the statistical results when pooling both facially expressive and inexpressive respondents.

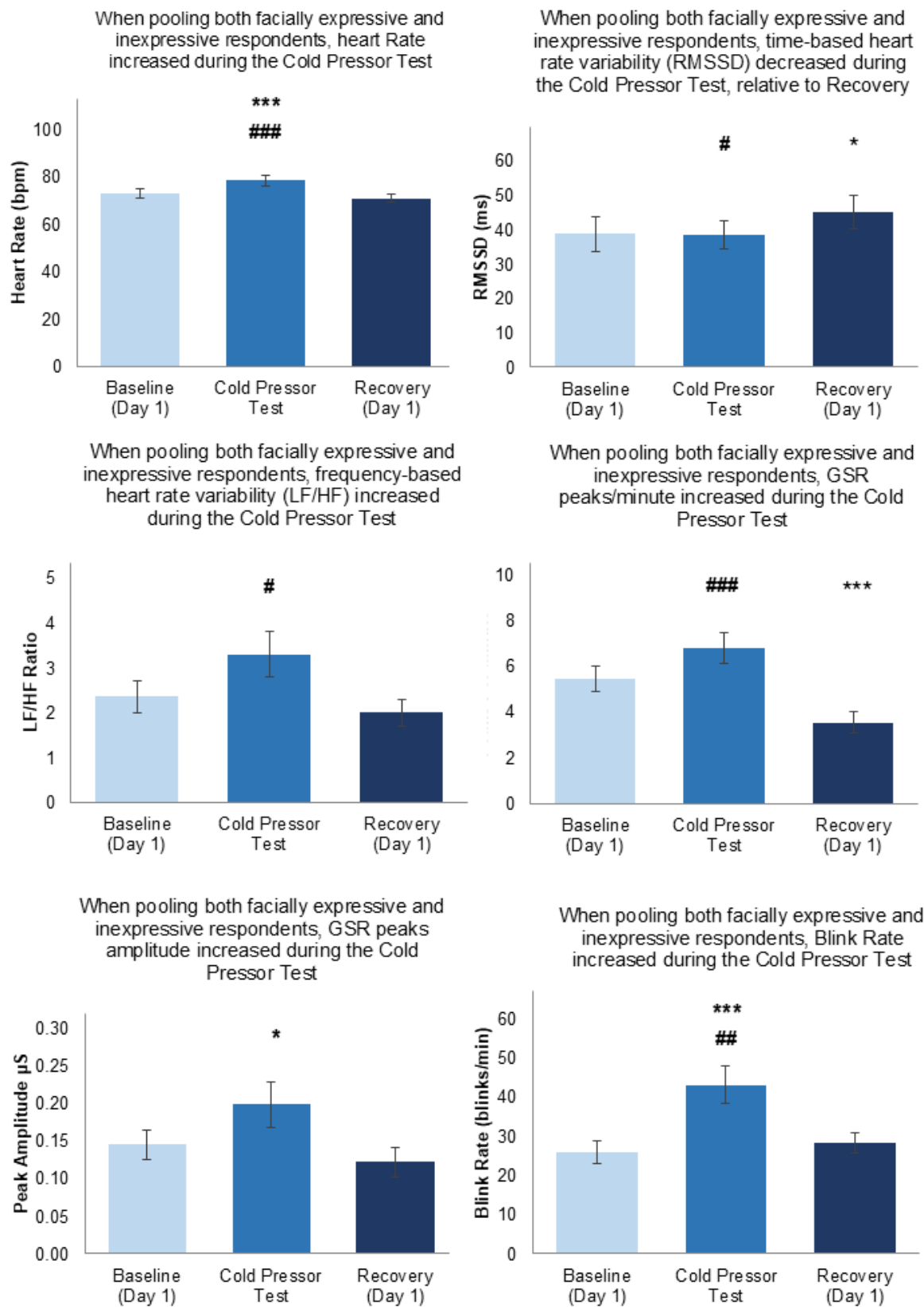


Figure 14: Appendix A: Statistical results of the Cold Pressor Test when pooling both facially expressive and inexpressive respondents together. Significant differences relative to Baseline of Day 1 are reported as *, $p < 0.05$; ***, $p < 0.001$ and relative to Recovery of Day 1 are reported as #, $p < 0.05$; ##, $p < 0.01$; ###, $p < 0.001$.

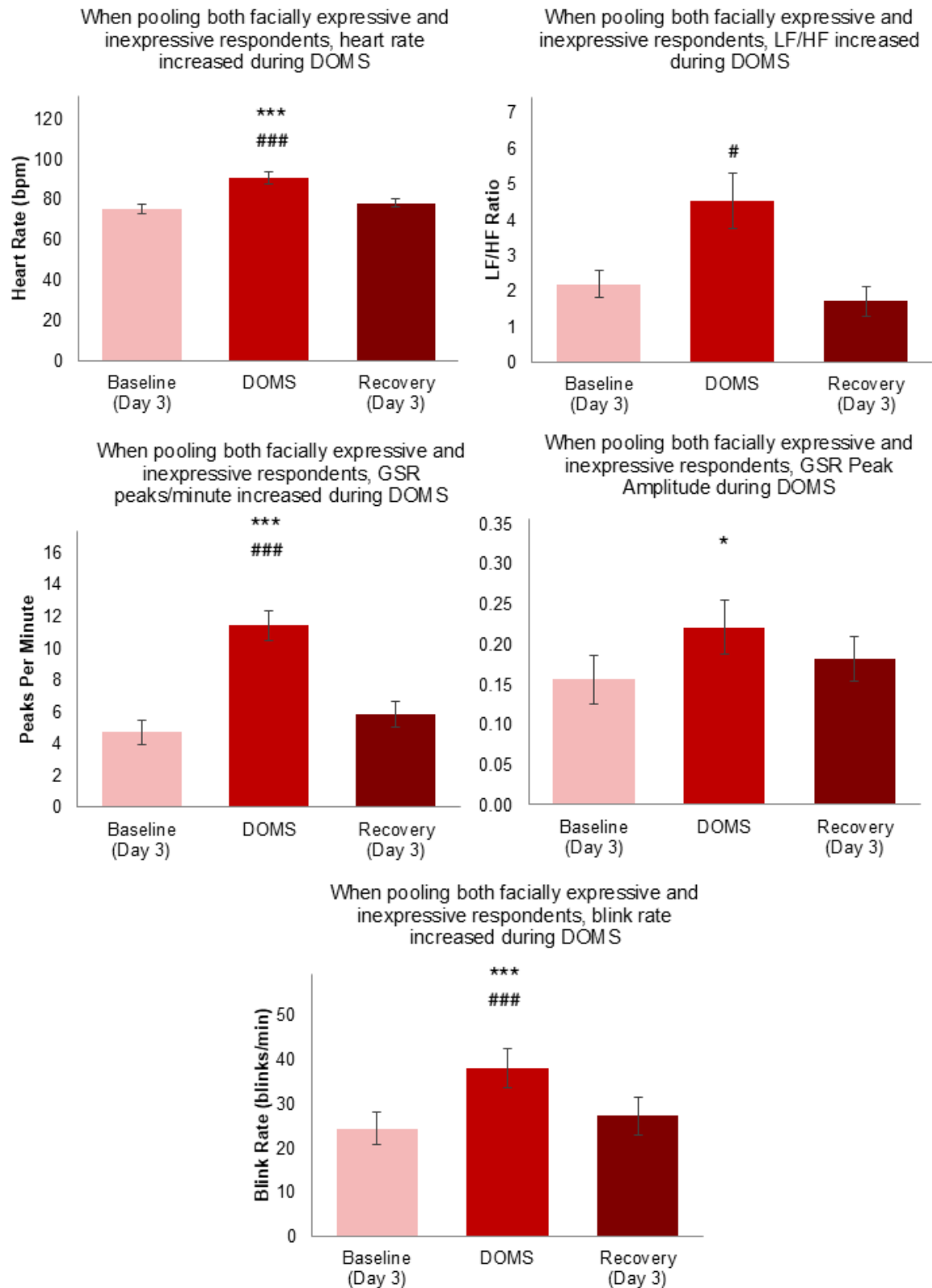


Figure 15: Appendix B: Statistical results of DOMS when pooling both facially expressive and inexpressive respondents together. Significant differences relative to Baseline of Day 1 are reported as *, $p < 0.05$; ***, $p < 0.001$ and relative to Recovery of Day 1 are reported as (#, $p < 0.05$; ###, $p < 0.001$).