Detecting Users' Emotional States during Passive Social Media Use

CHRISTOPH GEBHARDT*, ANDREAS BROMBACH*, TIFFANY LUONG, OTMAR HILLIGES, and CHRISTIAN HOLZ, Department of Computer Science, ETH Zürich, Switzerland

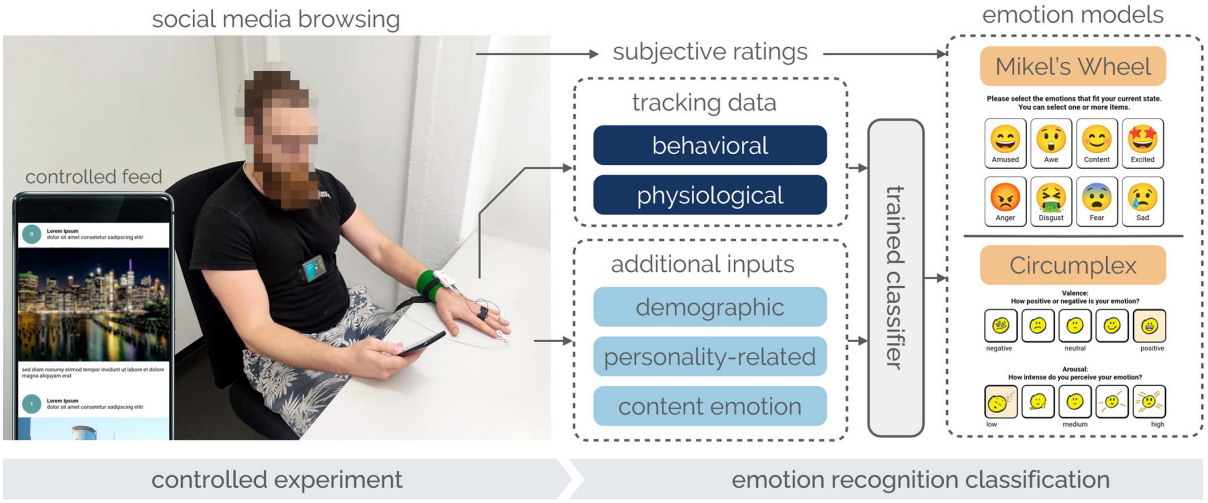


Fig. 1. We study the *passive* detection of users' emotions while they consume social media content. Our predictive model analyzes the behavior of phone use as well as physiological dynamics to estimate users' emotions. In our study, 29 participants interacted with a controlled social media feed that established a supervised learning setting. For our emotion detection model, we evaluated signals from behavioral and physiological sensors as well as demographic, personality-related, and content-related information. We also investigated several emotion models for the representation of subjective ratings.

The widespread use of social media significantly impacts users' emotions. Negative emotions, in particular, are frequently produced, which can drastically affect mental health. Recognizing these emotional states is essential for implementing effective warning systems for social networks. However, detecting emotions during *passive* social media use—the predominant mode of engagement—is challenging. We introduce the first predictive model that estimates user emotions during *passive* social media consumption alone. We conducted a study with 29 participants who interacted with a controlled social media feed. Our apparatus captured participants' behavior and their physiological signals while they browsed the feed and filled out self-reports from two validated emotion models. Using this data for supervised training, our emotion classifier robustly detected up to 8 emotional states and achieved 83% peak accuracy to classify affect. Our analysis shows that behavioral features

*These authors contributed equally to this work.

Authors' Contact Information: Christoph Gebhardt, christoph.gebhardt@inf.ethz.ch; Andreas Brombach; Tiffany Luong; Otmar Hilliges; Christian Holz, Department of Computer Science, ETH Zürich, Zurich, Switzerland.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM 2474-9567/2024/5-ART77
https://doi.org/10.1145/3659606

were sufficient to robustly recognize participants' emotions. It further highlights that within 8 seconds following a change in media content, objective features reveal a participant's new emotional state. We show that grounding labels in a componential emotion model outperforms dimensional models in higher-resolutional state detection. Our findings also demonstrate that using emotional properties of images, predicted by a deep learning model, further improves emotion recognition.

CCS Concepts: • **Human-centered computing → Empirical studies in ubiquitous and mobile computing**; **Social media**.

Additional Key Words and Phrases: Affective computing, emotion detection, classification, social media

ACM Reference Format:
Christoph Gebhardt, Andreas Brombach, Tiffany Luong, Otmar Hilliges, and Christian Holz. 2024. Detecting Users' Emotional States during Passive Social Media Use. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 2, Article 77 (May 2024), 30 pages. https://doi.org/10.1145/3659606

1 INTRODUCTION

Social media, as one of the most prevalently used services on the internet [11], holds a dual-edged influence on users' emotional well-being. On one hand, users share pleasant experiences that can evoke positive affective responses in others seeing this content [5]. These positive emotions can help relax [62], better cope with stress [8], and are even purposely utilized by some to regulate their mood throughout the day [93]. On the other hand, the use of social media is also linked with negative emotions, which, according to user perceptions, often overshadow its positive effects [10]. Users frequently encounter feelings of dissatisfaction, envy [12] and anxiety [2]. In combination with excessive usage patterns, these experiences can manifest in serious mental health problems, such as depression [48], suicidal thoughts [2], and body dysmorphia [87]. They are also recognized as catalysts for "doomscrolling" [3], permanent social comparison [12], and addictions [2].

Previous research has found emotions to be the perpetuating factor of social media use [77]. Elicited emotions increase user engagement [25], which, in turn, leads to more content being posted and consequently to more exposure to emotions [23]. This cycle is well-understood and exploited by big tech companies to maximize user screen time and, consequently, advertising revenue [100].

The centrality of emotions in the dynamics of social media use underscores the necessity for systems capable of recognizing users' emotional states. Effective emotion detection systems could serve as backbone for tools that support users' digital well-being. For instance, these systems could support digital self-control tools [49] in choosing interventions that decrease exposure to emotions during social media use, thus aiding users in reducing their time online. They could also transmit detected emotional states to emotion regulation platforms [82] to support users in becoming self-aware, empowering them to effectively regulate their emotions.

In active social media use, emotion recognition has been a subject of extensive research. This form of engagement encompasses behavioral patterns in which users actively participate in interactions on social network platforms [41]. This may involve activities such as creating and sharing their own content, or commenting on the content of others. Analyzing active social media usage provides a straightforward approach to discerning users' sentiments, as it allows for the visual examination of posted images [86], the linguistic assessment of posts [60], or the analysis of the dense behavioral traces users leave when they are, e.g., writing a comment [45, 72]. However, the prevalent mode of social media engagement is passive [91], where users predominantly consume content without generating noticeable behavioral traces [41]. This input-sparse setting presents a challenge in detecting an individual's emotional state, and has not yet been specifically addressed in emotion recognition research.

In this paper, we are the first to investigate emotional state recognition during passive social media use—the predominant mode of engagement. We conducted an experiment with 29 participants who navigated a controlled social media experience on a smartphone (Figure 1). In an Instagram-like feed, participants saw and scrolled through images of standardized emotional databases [52, 99] that feature common social media content (e.g.,

people, landscapes, faces, animals, and food [44]). Participants reported their emotions via dimensional ratings (Russell's Circumplex Model of Affect [74]) and a set of basic emotions (Mikel's Wheel [57]). Our apparatus passively monitored participants behavior using the smartphone's built-in sensors. Using wearable sensors, we also collected physiological signals of their cardiac, electrodermal, and respiratory activity similar to commercial smartwatches (e.g., Apple Watch, Fitbit, or Garmin).

Emotion detection during passive social media use: Approach and preview of findings (Figure 1)

Our study contributes a deeper understanding of the problem space of supervised emotion detection during passive social media use and beyond. Based on the collected data, we train supervised machine learning models to detect emotional states using objective features extracted from the captured signals and other inputs that are known to influence social media behavior (e.g., gender [92], nationality [50], sexual orientation [19], and personality [37]). For data augmentation, we incorporate a deep learning-based model that predicts elicited emotions exclusively from images as input into our emotion classifier. We analyze the data in four steps, each addressing a key factor of emotion recognition:

(1) A comparison between participants' physiological responses and their behavioral response. We first analyze the predictive power of the features we derive from participants' behavior, such as their movement of the phone, touch interaction, facial expressions, and eye movement. We compare this to the performance using the features we extract from their physiological signals, such as heart rate, respiratory rate, and electrodermal activity.

We find that behavioral features outperform physiological features as input to emotion classifiers. This holds true even when we compare the performance of behavioral features with a combination of behavioral and physiological features. This finding shows the potential practical implications of our study: behavioral features are easy to collect on the devices that are used for social media consumption (e.g., smartphone) and, in contrast to physiological features, do not require additional wearable devices (e.g., smartwatch). The insight can inform and simplify the design of future digital self-control and emotion regulation systems.

(2) The delay between a change in social media content and a discernible response in participants' behavior. We examine the temporal evolution of behavioral and physiological recordings, particularly in response to changing content as participants scrolled through the social media feed. Following a change in the content's affect, we investigate how the timing of feature extraction impacts the performance of the emotion classifier.

Our analysis shows that within 8 seconds following a change in media content, objective features reveal a participant's new emotional state. This novel insight shows the potential for social media apps to link the user's current emotional state to a specific post for observation times of 8 seconds and more, allowing them to implement strategies to better support the user's mental health.

(3) The effect of the emotion model used for subjective rating representation. We investigate two validated emotion models to represent participants' subjective ratings: (a) the established Circumplex Model of Affect [74], founded on the dimensional ratings of valence and arousal, and (b) Mikel's Wheel [57], a recent model that is based on a person's discrete reports of emotions. We were particularly interested in the granularity of detecting participants' emotional states. For this, we compare the performances of our emotion classifiers based on report representations at three abstraction levels: low (2 classes), medium (3–4 classes), and high (up to 8 classes).

We demonstrate that representing self-reports in the polar coordinate space of Mikel's Wheel improves emotion detection in settings of a high state granularity. We attribute this finding to the model's capacity to allow users to select multiple basic emotions, enabling them to express the nuances and ambiguities in their feelings. In a low or medium granularity of emotional states, we find that the two dimensions of the Circumplex Model—valence and arousal—provide more comprehensive insights into the specific constitution of a user's emotional state.

(4) The effect of user-specific and content-specific context information as classifier input. Finally, we show the impact of additional information about the user and the content on the performance of emotion classification. We find that the emotions associated with images as predicted by a deep-learning-based classifier [101] provide a strong additional feature for emotion recognition in our classifier, boosting the classification accuracy significantly.

Contributions

In summary, we make the following contributions in this paper:

(1) A controlled study of passive social media consumption on a mobile phone that monitored 29 participants' physical and physiological behavior in response to emotional content.
(2) An analysis of the impact of behavioral and physiological data on the detection of the emotional state.
(3) A temporal analysis of the delay between a change in the content's affect and its recognizability by a classifier.
(4) A comparison of two emotional models to represent participants' states: the traditional continuous valence-arousal model (Russell's Circumplex [74]) vs. a componential model (Mikel's Wheel [57]).
(5) An investigation into the impact of employing a deep learning-based approach to establish user-independent emotional priors for content, and utilizing them as input for the classifier.

Taken together, our findings have the potential to inform the design and development of future guidance and warning systems for passive social media consumption to help users manage their mental health.

2 RELATED WORK

2.1 Emotion models

Emotion models in psychology are classified into discrete, continuous, and componential models [39]. Discrete models define emotions as a set of basic states, such as Ekman and Friesen's six basic emotions [17] and Izard's ten core emotions [32]. Participants quantify their feelings by selecting combinations of these basic emotions.

Continuous models represent emotions in a coordinate system. The Circumplex Model of Affect by Russell [74] uses a 2D space with valence and arousal axes. To differentiate closely related emotions, the pleasure-arousal-dominance (PAD) model introduces dominance as a third dimension [55]. Emotions in this model are quantified using the self-assessment manikin (SAM) [7] or the Positive and Negative Affect Schedule (PANAS) [98].

Componential models, like Plutchik's Wheel of Emotions [68] and Mikel's Wheel [57], express emotions as combinations of basic states with varying intensities. Quantification is similar to discrete models, with participants selecting combinations of basic emotions. We contrast the robustness of labels represented in Russell's Circumplex Model of Affect [74], a widely used model, and in Mikel's Wheel [57], leveraging a novel parameterization for affect detection from images [101].

2.2 Emotion recognition from behavioral data

Emotion recognition from behavioral data is a central topic in affective computing [66], recent studies on emotion recognition from behavioral data highlight touchscreen and inertial measurement unit (IMU) data as key features for emotion detection during smartphone use [40]. Mottelson and Hornbæk [58] classified positive and negative affect using these sensors. Zualkernan et al. [107] and Wampfler et al. [96] further explored this approach to classify various emotional states. Other studies have utilized phone interaction logs. Sneha et al. [84] predicted seven emotions using typing metrics and context information. Mehrotra et al. [56] and Pielot et al. [67] examined the relationship between phone interactions and emotional states.

Eye-tracking data has also been used for this purpose. Matsuda et al. [53] combined gaze data with IMU and interaction logs to classify emotions. Pupillary response, was used by Heimerl et al. [29] to detect affect. Kosch et al. [35] and Le and Vea [43] used cameras to recognize users' emotions by analyzing their facial expressions.

2.3 Emotion recognition from physiological signals

Emotion's influence on physiological signals has been explored across disciplines. William James posited that emotions emerge from physiological reactions to external stimuli [33]. Building on this, Fairclough identified physiological signals as indicators of a user's psychological state, including emotions [21]. This concept underpins physiological computing, where systems respond to users' physiological reactions. We outline studies recognizing emotions from these signals in both controlled and real-world settings.

2.3.1 Laboratory environments. For studying affect, researchers have examined autonomic nervous system (ANS) activity and key psychophysiological variables, including electrocardiogram (ECG), photoplethysmogram (PPG), electordermal activity (EDA), respiration (RSP), electromyograph (EMG), and electroencephalogram (EEG) [42, 80]. Guo et al. explored ECG-based emotion recognition [26], while Abdullah et al. investigated PPG's valence distinction, correlating it strongly with ECG [1]. Silveira et al. used EDA in movie emotion estimation [81]. Petrantonakis and Hadjileontiadis identified basic emotions with EEG and Adaptive Filtering [64].

Efforts combining modalities show promise, such as Yin's EDA-PPG fusion for valence-arousal clustering [104]. HRV and EDA combined with facial expressions enable fine-grained emotion recognition [79]. Various studies integrate EEG, PPG, and EDA with machine learning [90, 97] and deep learning [46, 47]. Respiration and EMG were also considered, both for emotion identification [30] and assessing emotional stimuli impact on psychophysiological variables [88]. Recent comparisons between medical-grade and consumer-grade sensors show similar recognition rates for valence and arousal in electrodermal and cardiac activity [70]. Panganiban identified stress using PPG data from smartphone cameras and high-quality wearable sensors [61].

2.3.2 Real-world environments. Consumer-grade devices like smartwatches and wristbands have emerged as reliable sources for robust emotion detection in real-world environments [75]. Healy et al. used data from wireless galvanic skin response, heart rate, activity sensors, and a mobile phone to recognize binary valence and affect [28]. Pham et al. fused PPG data from a mobile phone's back-facing camera with facial expressions from the front-facing camera to predict six discrete emotions, noting the respective strengths of each modality [65]. Quiroz et al. combined data from an IMU sensor, a smartwatch, and a heart rate monitor strap to determine affect from a set of three emotions [69]. Subsequent research improved their approach with deep learning on the same dataset [89]. Contrastive Representation Learning was also found to enhance emotional state detection [15].

Schmidt et al. explored CNNs for affect classification using the Empatica E4 wristband, equipped with EDA, PPG, IMU, and skin temperature sensors [76]. Other wristbands with similar sensor modalities have also proven effective for emotional state detection [78]. Expanding the modalities to include EMG and behavioral data, recent research demonstrates the feasibility of reliably classifying binary valence and affect using deep learning-based approaches with signals from consumer-grade mobile devices [103].

2.4 Emotion Recognition in social media use

Research also investigated emotions in the context of social media. Mauri et al. investigated the effect of social media use on physiological signals in a controlled setting, revealing high positive valence and arousal states [54]. Šola et al. analyzed eye movements and facial expressions to study how subconsciously processed stimuli in social media affect emotions, showing that negative emotions decrease when posts contain human faces [108].

Several studies have explored emotion detection during smartphone-based social media use. Lee et al. used Bayesian Networks and phone interaction data to identify seven distinct user emotions during Twitter typing [45]. Ruensuk et al. identified binary states of arousal and valence in both laboratory and real-world experiments using smartphone behavioral data [72]. Their subsequent work focused on negative affective experiences related to appearance comparison and envy, demonstrating the feasibility of identifying such states in various settings [73].

While the studies above span passive and active forms of engagement, our work uniquely focuses on *passive usage*—the predominant mode of social media use. In addition, our research is first to leverage physiological signals exhibited during social media use for the purpose of emotion recognition.

3 EMOTION RECOGNITION DURING PASSIVE SOCIAL MEDIA USE

Our overview of previous work in this area showed that recognizing users' emotions during social media use is challenging, particularly when monitored passively due to sparse behavioral traces. Our study thus aims to enhance emotion recognition in this context by investigating several key factors:

(F1) *Physiological Signals:* Since physiological signals are known to respond quickly to stimuli, we collect them during our study to improve emotion recognition during passive social media use.
(F2) *Response Delays:* Comprehending the time it takes for participants to exhibit a noticeable emotional response after viewing specific social media content is crucial. We conduct a temporal analysis of the delay between a change in the content's affect and its recognizability by a classifier.
(F3) *Emotion Models:* We study the impact of emotion models on prediction performance, comparing dimensional models, Russell's Circumplex Model of Affect, and Mikel's Wheel.
(F4) *Additional Information:* We analyze the potential benefits of using user-specific factors (e.g., gender, nationality, personality traits) for emotion prediction. Additionally, we investigate if computationally inferred emotions from images enhances emotion prediction.

In summary, our study examines the impact of emotion models, time delays, physiological data, and additional user and content information on emotion recognition during passive social media use. We collect physiological signals that are attainable from consumer-grade devices, including EDA (Fitbit), PPG (smartwatches like Apple Watch), ECG (heart rate monitor straps like Polar's H10), and respiration tracking (available on many smartwatches like Garmin). By systematically investigating these factors, we aim to improve the accuracy of emotion recognition in the unique context of passive social media engagement.

4 METHODS

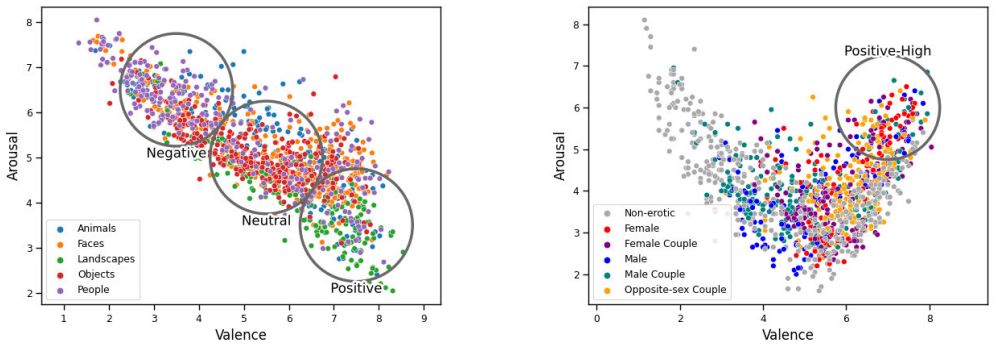
We designed an experiment to investigate the defined factors. The following subsections detail the experimental setting, including stimuli presentation, collected data, procedure, study apparatus and participants. This study has received approval from the Institutional Review Board of the host university ensuring compliance with ethical standards and protection of human subjects participating in the research.

4.1 Curated social media feed

The design of our experiment balances external validity (i.e., resembling a real-world setting of passive social media use) and internal validity (i.e., maintaining control over observed stimuli). We developed a curated social media feed that allows us to regulate the emotionality of stimuli while emulating a typical social network page.

4.1.1 Stimuli. We retrieved the stimuli for our feed from the NAPS, a standardized emotional database containing high-quality images along with their corresponding emotional ratings [52]. Additionally, we sourced images from NAPS-ERO, a complementary database that incorporates erotic images to induce positive valence and high arousal [99]. We manually ensured that all selected images adhere to the community guidelines of Instagram. The image categories present in the NAPS (animals, faces, landscapes, objects, and people) align closely with the most prevalent categories of images typically found on social media platforms. These categories encompass self-portraits (featuring only the face or the entire person), group portraits, scenes (including human-made and urban settings, both indoor and outdoor activities), animals, and food[1] [31, 44]. With the stimuli form NAPS-ERO,

[1]While "food" is not explicitly categorized in the NAPS, images related to food are encompassed within the broader "object" class.



(a) Average ratings from NAPS dataset.

(b) Average ratings from NAPS-ERO dataset.

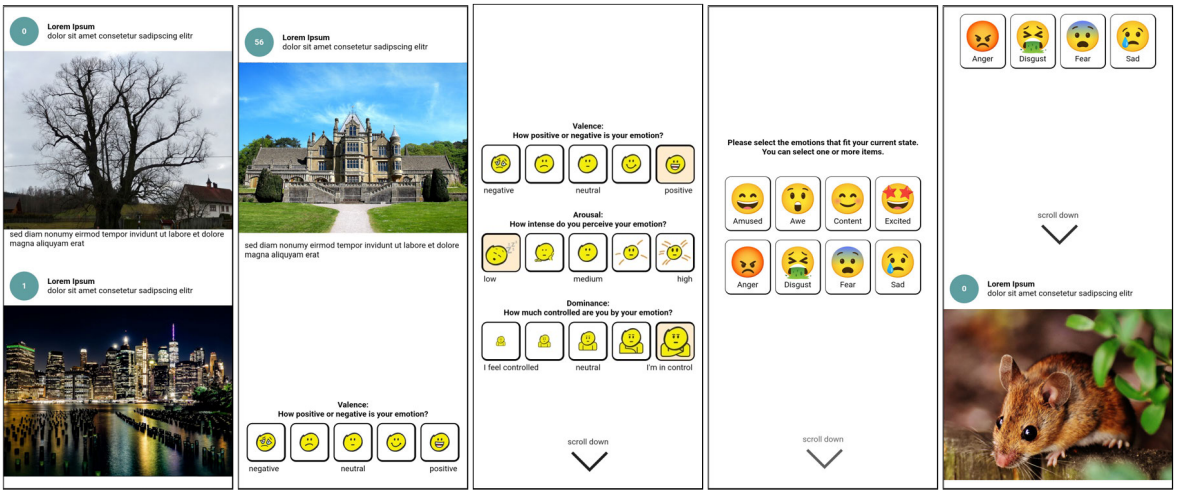
Fig. 2. Image ratings of NAPS [52] and NAPS-ERO [99] displayed in the valence-arousal space, annotated with clusters

we could reflect the high proportion of erotic and semi-erotic content on social media in our feed [16]. As no significant difference in the elicitation of emotion between videos and images within a social network context was shown in prior work [72], we limited our stimuli set to images.

To regulate the emotional content of stimuli, we specified clusters based on the emotional ratings which we then used to sample images during the study. We defined these clusters as non-overlapping circles around a fixed point in the valence-arousal space [74], each with a radius of 1.25 (scales reach from 1 to 9). Figure 2a shows the distribution of ratings of NAPS in this space and the position of the defined clusters. Specifically, for the NAPS set, we positioned the negative cluster at <val=3.5, aro=6.5>, the neutral cluster at <val=5.5, aro=5.0>, and the positive cluster at <val=7.5, aro=3.5>. For the NAPS-ERO set, the positive-high cluster was placed at <val=7.0, aro=6.0> (Figure 2b). For each NAPS image category (animals, faces, landscapes, objects, and people), we then sampled 60 to 80 images for the negative, neutral, and positive cluster, resulting in 15 sets of images. To attain images for the positive-high cluster, we sampled four more image sets from the social media-compliant NAPS-ERO subset, each representing the highest-rated images by heterosexual male, homosexual male, heterosexual female, and homosexual female raters respectively. This results in a total of 19 distinct category-cluster image sets to which participants were exposed during the study.

4.1.2 User interface. To present stimuli to participants, we developed an Android application that displays the sets of images in a vertical, Instagram-like feed layout (Figure 3). To ensure that elicited emotions arise solely from the controlled stimuli while maintaining a resemblance to real-world social media feeds, we augmented images with *Lorem Ipsum* text (Figure 3a). In a trial, the app operates as follows: when the first image from a set becomes visible in the feed, a 20-second timer is activated. Throughout this period, participants can scroll through the feed at their preferred pace, while the application randomly samples images from the corresponding cluster-category image set. After the timer expired, the app automatically inserts the self-assessment questionnaires into the feed (Figure 3b), so that no further images can be displayed. Thus, emoticon representations of SAM and the distinct emotions of Mikel's Wheel are used (Figure 3c, 3d). After completing the self-reports, the next feed becomes available and an arrow at the bottom of the screen prompted the participant to continue (Figure 3e).

We purposely designed the app such that no neutral stimuli are presented between two feeds stemming from different cluster-category image sets. Thus, stimuli experienced in the previous feed continue to influence a participant's current emotional state. With this departure from highly controlled experimental settings where



(a) View of curated feed to scroll through. (b) Self-reports appear in feed after timeout. (c) SAM items adopted from [94]. (d) Emoji items based on Mikel's Wheel [57]. (e) Next feed is shown after self-reports.

Fig. 3. User interface of the curated social media feed (images are not from NAPS and shown for illustrative purposes).

neutral stimuli are inserted between different stimulus groups, we better reflect the dynamic nature of social media. In social media, various other factors than images influence user emotions (e.g., captions, relationship to the poster, likes) which are constantly affected by brief exposures to new content [108]. By introducing emotional contagion across stimuli groups, our experimental design moves closer to realism, allowing us to investigate the time delay at which a change in content can robustly be detected in users' emotional state.

4.2 Collected data

4.2.1 Behavioral & physiological signals. Table 1 provides an overview of the behavioral and physiological data collected during our study, categorizing it into feature groups, and specific features. It furthermore references examples of related work where these feature groups have been used for emotion recognition. For a more detailed description of each feature and its computation, we refer to Section 4.5.

4.2.2 Subjective ratings. We asked participants to report their affective state on two different type of questionnaires. The first questionnaire is the self-assessment manikin (SAM) [7], which returns ratings in the valence-arousal space of Russell's Circumplex Model of Affect [74]. Specifically, we used emoti-SAM [27], a validated 5-point-scale version of the SAM that is better suited for smartphone screens [96] (see Figure 3c). The second questionnaire allowed participants to select one or more of the eight emotional categories from Mikel's Wheel: fear, sadness, disgust, anger, contentment, amusement, awe, and excitement [57]. To increase its ease-of-use, we added icons of the Noto Emoji Library[2] to the textual descriptions of the distinct emotions (see Figure 3d). From the multiple emotional categories selected in such a self-report, a composite emotional state needs to be computed to use as a label in supervised learning for emotion recognition. Thus, we leveraged recent research in Computer Vision, which has introduced a parameterization of Mikel's Wheel in a 2D polar coordinate space [101]. With this parameterization, we computed a compound emotional vector for each subjective rating of

[2]https://github.com/adobe-fonts/noto-emoji-svg

Table 1. Data type, feature groups and specific features of the collected data. If not stated differently, the references point to examples of related work where respective feature groups have been used for emotion recognition.

Type	Group	Features
Behavioral	Phone interactions (e.g., [56, 67])	timestamps of visible elements, duration of each set, number of images displayed until timeout, total display time of each image
	Motion (e.g.,[58, 96])	acceleration (x,y,z) angular speed (x,y,z)
	Touch (e.g.,[59, 95])	number of gestures, duration, length, speed, acceleration, time interval between gestures, pressure, change in pressure, direction and directness of stroke
	Facial expressions (e.g., [36, 65])	distribution over distinct facial expressions
	Eye-tracking (e.g., [53, 72])	eye blinks, eye aspect ratio of left and right eye
Physiological	ECG (e.g., [1, 106])	heart rate, heart rate variability (mean and st. dev of RR intervals)
	PPG (e.g.,[70, 104]	heart rate, heart rate variability
	EDA (e.g., [28, 70])	phasic/tonic amplitude, for the phasic signal: number of peaks, time to first onset, rise time, peak amplitude, half-way recovery time
	RSP (e.g., [30, 103])	respiration rate, amplitude
User	Demographics (e.g., [84])	gender, nationality, sexual orientation
	Traits (e.g., [38])	personality traits (according to [85]) self-esteem (according to [71])
Content	Image emotions	predicted image emotions

participants where the vector's angle represents the valence and its radial distance from origin the emotional intensity (see [101] for details of the algorithm).

As we are interested in the effect of the emotion model on performance across different number of emotional states to be detected, we summarized the subjective ratings of the different scales into affective states of varying levels of resolution. This is straightforward for SAM, where divisions of the ordinal scale can function as distinct levels of resolution. Specifically, we divided the scales of valence and arousal into low resolution (two negative scale points vs. rest), medium resolution (two negative scale points vs. neutral scale point vs. two positive scale points), and high resolution (each individual scale point, Figure 4a). To attain different levels of detail for the emotional vectors of Mikels' Wheel, we specified reference angles in the polar coordinate space and mapped each self-report vector to the nearest one. The resolutions of Mikel's wheel that we utilized in our analysis include low resolution (positive vs. negative emotions), medium resolution (fear, sadness vs. disgust, anger vs. contentment, amusement vs. awe, excitement), and high resolution (the eight distinct emotions, Figure 4b).

(a) State resolutions of the Circumplex Model.

(b) State resolutions of Mikel's Wheel.

Fig. 4. Visualization of how the different state resolutions are attained in the respective emotion model. For the Circumplex Model, the ordinal ratings are split between different scale points. For Mikel's Wheel, reference angles were specified according to the respective resolution. Emotional vectors that represent self-reports were then mapped to the nearest one.

4.2.3 Demographic & personality-related data. Prior to the study, we asked participants to fill out a demographic questionnaire asking for basic demographic information, such as age and education level, as well as gender, nationality and sexual orientation. We also inquired about participants usage patterns of the most popular social network platforms. Following related work [72, 73], we queried frequency of use, daily time spent, content preferences, and if their usage behavior is primarily active or passive. Finally, we assessed participants' self-esteem using the Rosenberg Self-Esteem Scale [71] and their personality traits with the Big Five Inventory 2 [85].

4.2.4 Predicted image emotions. To attain emotional priors for the content participants observed during the study, we employed a deep-learning-based approach [101]. This method predicts the emotions evoked in observers based on the self-reports it was initially trained on. We re-implemented the approach, using the same training procedure and losses. Subsequently, we trained it on two social media datasets, namely Flickr_ldl and Twitter_ldl, which were labeled with emotions by a population of viewers [102]. Employing the trained network, we predicted a distribution over discrete emotions for each image in NAPS and NAPS-ERO. To use these distributions as input for a classifier, we aggregated the images shown in the feed over each time window for which features were computed and then normalized the sum by its number of images. Figure 5 explains this procedure visually.

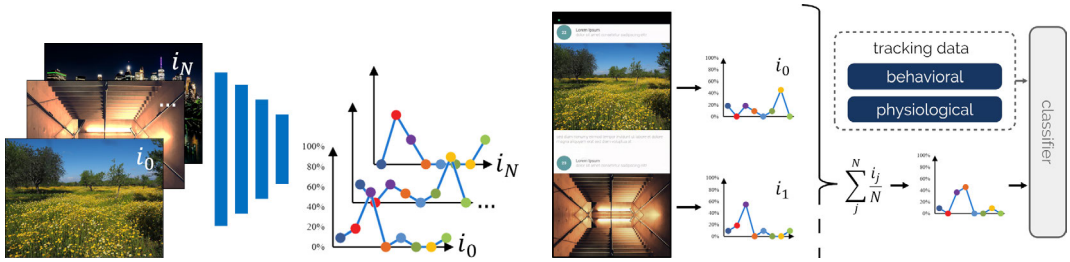


Fig. 5. To attain emotional priors for the content participants observed during the study, we used a neural network to predict a distributions over discrete emotions for each image in NAPS and NAPS-ERO (left). To train a classifier, the distributions of images shown in the feed were aggregated over each time window for which features were computed and then normalized by its number of images. The resulting emotional distribution was used as input to the classifier (right).

4.3 Procedure

Before the start of the study, each participant was led to a separate room where they signed the consent form and completed the pre-questionnaires. It was emphasized that participants did not have to answer any subsequent questions that they felt were too private or personal. Participants were then asked to mount the physiological sensors in an isolated space with blinds. They were told that they are free to leave at any time if wearing sensors directly on their skin makes them feel uncomfortable. After that, the experimenter gave a short introduction to the study and showed participants the Android app with the curated social media feed. They could then familiarize themselves with the app, which displayed images in the feed that were not utilized in the actual experiment. Participants were told to scroll through the images at their own pace, as if it were a feed from a social media platform they frequently use.

After finishing the briefing, participants had to complete a relaxation session, in which a calming video of a beach was shown for two minutes. Participants were then instructed on-screen to begin with the experiment. In the experiment, participants were exposed sequentially to the 19 cluster-category image sets. The order of image sets was randomized to ensure that no two consecutive sets were sampled from the same cluster. Additionally, for each set and trial, the images were arranged in a random order. After each block of browsing the feed they went through the self-assessments. Upon conclusion of the study, participants were instructed to unmount the sensors, and the experimenter conducted a short debriefing session. Overall, a trial took between 60 and 90 minutes.

4.4 Apparatus

4.4.1 Platform. The experimental platform consisted of two smartphones and a data collection backend. On the first phone (Huawei P9 Plus), participants interacted with the curated social media feed (see Section 4.1). The feed was implemented using the JavaScript framework Vue.js and was accessed within an Android app using a web-view widget. The Android application also collected camera, IMU and touch screen data in the background. We chose the Huawei P9 Plus for its pressure-sensitive display. Pressure features have been identified as robust predictors of arousal in related studies [95]. The second phone (Xiaomi Mi A3) was solely used to log the data collected by the various physiological sensors (participants did not engage with it). Thus, we developed another Android application that accessed the sensors' Android SDKs. Both applications transmit the logged data to a dedicated Python backend, which stores the logs and also serves as the host for the web service delivering stimuli for the curated social media feed. Participants used a separate tablet to answer the pre-questionnaires.

4.4.2 Sensors. EDA and PPG were recorded with a Shimmer3 GSR+ unit at a sampling rate of 128 Hz. A Polar H10 belt was used to record ECG, sampled at 130 Hz. Breathing rate was monitored with a Vernier Go Direct Respiration Belt at a sampling rate of 20 Hz. In addition, images were recorded from the camera of the smartphone at a frequency of 10 Hz. Similarly, the smartphones' IMU was recorded at a frequency of 100 Hz. We used experimental sensors to collect physiological data to establish an upper bound in terms of signal quality when compared to the sensors on consumer-grade devices. Thus, we only considered sensor that are integrated in consumer-grade wearable devices.

4.5 Data preprocessing

We split the 20 second time windows that participants are exposed to a cluster-category image set into four sub-windows, ranging from 0-8 seconds, 4-12 seconds, 8-16 seconds, and 12-20 seconds. The sub-windows were selected to maintain a 50% overlap, ensuring both significant signal variation and comprehensive temporal coverage. For comparison purposes, we included the full 20 seconds in our analysis. We chose the length of 8 seconds based on findings of prior work related to the response time of physiological variables to a stimulus (3-5 seconds for EDA [6], 6 seconds for HR data [4]). The 20-second time duration of exposure to specific groups of

emotional stimuli ensures that participants are influenced by them (related work uses 5-10 seconds [94]) and allows for the detection of effects in behavioral data (related work uses 15 seconds [72, 73]).

4.5.1 Physiological signals. Signals from the individual physiological sensors (ECG, EDA, PPG, RSP) were pre-processed using the Neurokit2 library [51]. First, to eliminate small fluctuations caused by randomly occurring transmission or processing delays, the recorded data was imputed and re-sampled to match the respective target sampling frequency of the sensor. Then, we removed noise from the signals by using a Butterworth filter with a sensor-specific order as well as lower and higher cutoff frequencies (ECG: 5th-order, cutoff > 0.5 Hz; PPG: 3rd order; cut off < 0.5 Hz and > 8 Hz; EDA: 4th-order, cutoff > 3 Hz; RSP: 2nd-order, cut off < 0.05 Hz and > 3 Hz). For the ECG signal, an additional filtering step is included to remove powerline noise by applying a moving average kernel with the width of 1/50 Hz. For the ECG signal, we detected the location of the R-peaks (maximum amplitude in the R wave) as local maxima in the signal to calculate the NN intervals (time between two detected heartbeats). From the NN intervals, we then computed the heart rate (HR) and heart rate variability (HRV) using the Neurokit2 library. PPG data is processed by performing systolic peak detection [18] to mark their positions in the signal. HR and HRV are then computed analogously to the ECG signal. Using a Butterworth filter, the EDA signal is decomposed into its phasic and tonic components. To find the skin conductance response to a stimulus, peaks are then located as local maxima in the phasic component. From the location of peaks, the rise time (time it takes to reach peak amplitude from onset), the peak amplitude, and the half-recovery time (time it takes from peak to decrease to half amplitude) are computed using NeuroKit2. For respiration, we detect peaks [34] and then calculate breathing rate and amplitude from their locations.

Finally, we split all signals into the 20-second cluster-category time window and subsequently into the four corresponding sub-windows. For each sub-window and signal, we then compute summary statistics, i.e., number of peaks, mean, standard deviation, minimum, maximum. These features were then normalized by subtracting the respective statistics of the recorded signals of the last 30 seconds of the relaxation phase prior to the experiment.

4.5.2 Behavioral data. In a first step, all collected behavioral signals underwent a cleaning process and were imputed, wherever feasible. Within the browser, user interactions such as button presses, or page scrolling were recorded. The corresponding timestamps of elements entering or leaving the viewport were then used to calculate the number of images displayed in a set, average image display duration, as well as overall rating duration. Individual touch data points were grouped into coherent strokes for which distance, duration and pressure was computed. The speed and acceleration of a stroke was then calculate by dividing its distance and duration. Additional stroke features were determined by measuring the disparity between the straight-line distance from the start to the endpoint of a stroke and its total distance. The direction of a stroke is computed as the angle of the line between the endpoints. We also recorded the pressure value of each touch point and used pressure differences of the first and last point of a stroke as additional feature. Similarly, differences in speed and acceleration between the first and last segment in a stroke were computed. For the signals attained from the inertial measurement unit (IMU) of the smartphone, i.e., acceleration, rotation, and magnetic field strength, the magnitudes and the differences in magnitude were computed. Facial expressions were predicted based on the captured images from the front-facing camera of the phone using an implementation of the method described in [105], which provides a distribution over seven distinct facial expressions. For eye blink detection, we predicted facial landmarks on the same images and then followed the approach in [9] to compute eye aspect ratios and blinks.

Like for physiological signals, we aggregated the behavioral features to their mean, standard deviation, minimum, maximum and absolute number for any given time window. As participants did not engage with the phone during the relaxation phase, we standardized the behavioral data on a per-participant basis. Thus, for each feature, we subtracted the mean of collected data over the whole experiment and divided the result by their standard deviation. The predicted distributions of facial expressions were aggregated over a time window and then normalized by the number of recorded frames.

4.6 Participants

Participants were recruited using snowball sampling in social media channels related to the host university (e.g., channels of student associations). They needed to confirm that they were not taking tranquillisers, psychotropic drugs (e.g., anti-depressants), or narcotics, and were not diagnosed with cardiovascular diseases. A total of 29 participants (12 female, 17 male) between ages 18–32 (M=25.48, SD=2.84) were recruited. 22 participants identified as heterosexual, 4 participants as bisexual and 3 as homosexual. They possessed 8 different nationalities. All of them were frequently using at least one social media application, spending on average 2.59 hours (SD=1.37 h) per day on social media. 23 participants stated to use social media predominantly passive, 5 participants to use it equally active and passive, and one person indicated primarily active use. We had to exclude the data of three participants due to connection problems with the physiological sensors. The data of the other 26 participants were used in our analysis.

5 RESULTS

In the following, we present the results of our experiment. For significance testing, we performed an ANOVA if data was normally distributed (Shapiro-Wilk $p > .05$) and exhibited equal variances between groups (Levene's $p > .05$). Pairwise comparisons were performed using t-tests with Bonferroni-adjusted p-values. If either assumption was violated, we assessed differences with the non-parametric Aligned Rank Transform (ART) ANOVA. Post-hoc comparisons were then performed using the ART-C algorithm and Bonferroni corrections. If not stated differently, independent variables were considered as within-subject factors and participants as a random factor.

5.1 Statistical analysis of self-reports

In a first step, we assessed whether the distinct clusters indeed provoked varied emotional responses among participants. Thus, we regarded them as independent variables and treated the different categories as repetitions. We then examined their impact on valence and arousal, as measured through the SAM self-reports. Figure 6 shows the distributions of valence and arousal ratings averaged per participant and grouped by cluster. Due to the ordinal nature of these dependent variables, we assessed differences with the repeated-measures ART ANOVA and the respective post-hoc test.

We found a significant effect of clusters on valence ratings [$F_{3,446} = 239.77$, $p < .0001$]. Post-hoc tests have shown that clusters caused the intended increase of valence according to the experimental design ($p < .0001$ for all). No significant differences were observed solely between the positive-high and positive clusters. Similarly, the effect of clusters on arousal ratings was significant [$F_{3,446} = 50.13$, $p < .0001$]. The pairwise comparison has shown that differences in arousal between all clusters were significant ($p < .0001$ for all). The only non-significant difference

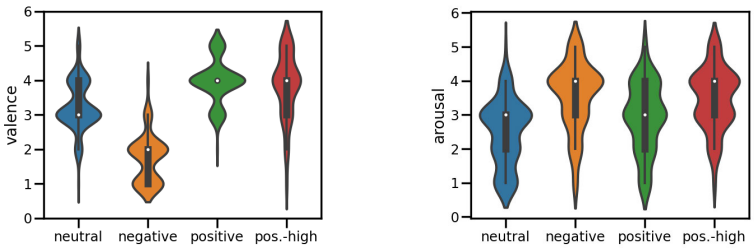


Fig. 6. Subjective ratings of participants for valence (left) and arousal (right) per cluster (negative, neutral, positive, pos.-high), averaged over category according to emoti-SAM [27].

Table 2. Fraction of physiological and behavioral features for each time period where a significant difference between clusters of a particular magnitude ($^{*}\ p < .05$, $^{**}\ p < .01$) was identified.

Type	Group	0 - 8 sec		4 - 12 sec		8 - 16 sec		12 - 20 sec		0 - 20 sec	
		*	**	*	**	*	**	*	**	*	**
physiological	ECG	0.43	0.14	0.62	0.52	0.62	0.52	0.43	0.43	0.62	0.43
	PPG	0.48	0.33	0.43	0.24	0.38	0.33	0.33	0.33	0.52	0.33
	EDA	0.04	-	0.12	0.04	0.23	-	0.54	0.23	0.08	-
	RSP	0.33	-	0.33	-	0.44	-	0.28	0.11	0.39	0.11
	All	**0.26**	**0.09**	**0.30**	**0.16**	**0.37**	**0.17**	**0.36**	**0.22**	**0.36**	**0.17**
behavioral	Phone interactions	0.50	0.50	0.80	0.80	0.90	0.60	0.60	0.50	1.00	0.90
	Eye tracking	-	-	-	-	-	-	0.67	0.33	-	-
	Facial expressions	-	-	0.57	0.14	0.71	0.57	0.57	0.43	0.57	0.14
	Motion	0.12	-	0.44	0.19	0.31	0.25	0.31	-	0.50	0.06
	Touch	0.35	0.16	0.54	0.32	0.43	0.30	0.53	0.31	0.63	0.40
	All	**0.19**	**0.13**	**0.47**	**0.29**	**0.47**	**0.34**	**0.54**	**0.31**	**0.54**	**0.30**

in the case of arousal was observed between the positive-high and negative clusters, as was intended by the experimental design (see Figure 2). In summary, the analysis confirmed that our experimental design effectively evoked distinct emotions within participants.

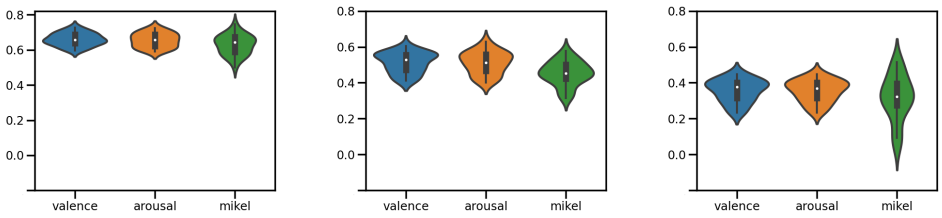
5.2 Statistical analysis of features

After verifying that the clusters did evoke varied emotions, we statistically analyzed their effect on the distributions of the summary statistics of features for each time window. Thus, the subject ID was considered a repeated measure. Table 2 provides a summary of the fraction of physiological and behavioral features for each time period where a significant difference between clusters was identified. Results show that both physiological and behavioral features immediately react to the change in the emotional content and already in the initial time window an effect of clusters on features of each group (other than eye tracking) can be observed. Interestingly, there is a difference in the fraction of features that show a main effect of clusters in the different time windows between PPG and ECG, even though the sets are identical. In general, physiological and behavioral features appear to exhibit an increasing reactivity to clusters over time with the 8 - 16 window (physiological) and the 12 - 20 window (behavioral) showing the highest fraction of significant features. For both feature types, the entire 20-second period tends to exhibit a similar ratio of significant features than the late 8-second windows.

Similarly, we conducted a correlation analysis between physical and behavioral features and participant's valence and arousal ratings. No distinct temporal correlation pattern is observable (see Appendix A).

5.3 Emotion model (**F3**)

We applied various classifiers to the defined model-dependent resolutions of collected self-reports and the extracted features from all time windows. Thus, we selected six classifiers that have demonstrated superior performance in prior studies [24, 72, 73, 75]: AdaBoost (AB), Logistic Regression (LR), Random Forest (RF), RBF-kernel Support Vector Machine (SVM), Multi-layer Perceptron (MLP), and XGBoost (XGB). We also included a baseline classification strategy (ZeroR) that predicts the class that is most frequent in the distribution of the training data. For all classifiers, we used their scikit-learn implementation with default parameters. To provide a robust and unbiased evaluation of models, we assessed the performance of each classifier using leave-one-subject-out cross-validation. As class balancing is crucial for an adequate performance in emotion recognition [75], we over-sampled minority classes in the training data using the Synthetic Minority Oversampling Technique. Finally, we standardized the data by employing scikit-learn's MinMaxScaler, fitting it to the training data, and



(a) low resolution (b) medium resolution (c) high resolution

Fig. 7. Classifier accuracy for the different emotion models (valence, arousal, and mikel) per resolution.

subsequently applying its transformation to both the training and test sets. The accuracies of the different classifiers is summarized in Appendix B.

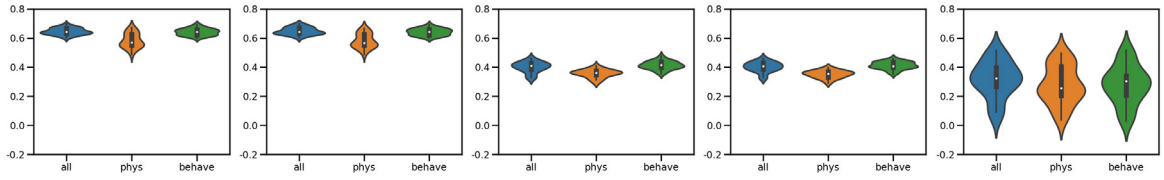
We compared the effect of the different emotion models (*Mikel*, *valence*, and *arousal*) across all time windows on classification accuracy for each resolution. The ANOVA found a significant effect of the emotion model on classification performance for the medium resolution [$F_{2,87} = 7.74$, $p < .0001$]. In terms of medium resolutions (Figure 7b), *valence* and *arousal* caused a better classification performance than *Mikel* (both $p < .0.004$). This can be attributed to the difference in number of states in the resolutional setting (three for *valence* and *arousal*, four for *Mikel*). For low- (Figure 7a) and high resolutions (Figure 7c), we found no main effect of the different emotion models on accuracy.

In high resolutions, classifiers had to predict eight distinct emotional states with *Mikel* as opposed to five with *valence* and *arousal*. Nevertheless, no difference in accuracy between the emotion models was found. Thus, we constrain the remainder of this analysis to use *Mikel* for the high range of predicted states. For low and medium resolutions, we use *valence* and *arousal* due to their lower variance of accuracies and as their dimensional nature provides more insights into the specific constitution of a user's emotional state.

5.4 Physiological signals (**F1**) and response delays (**F2**)

We analyzed the effect of the different feature types (*phys*: physiological; *behav*: behavioral; *all*: combination of both) across all time windows on classification accuracy for the previously specified resolution-model pairs (*valence-low*, *arousal-low*, *valence-medium*, *arousal-medium*, and *Mikel-high*). Time windows were considered as within-subject factor and feature types as between-subject factors.

The ANOVA revealed a main effect of the feature type on accuracy for *valence-low* [$F_{2,75} = 18.04$, $p < .0001$], *arousal-low* [$F_{2,75} = 19.44$, $p < .0001$], *valence-medium* [$F_{2,75} = 48.80$, $p < .0001$], and *arousal-medium* [$F_{2,75} = 62.91$,



(a) valence-low (b) arousal-low (c) valence-medium (d) arousal-medium (e) Mikel-high

Fig. 8. Classifier accuracy for the different feature types (all, phys., behave.) per model-resolution pair.

$p < .0001$]. Pairwise comparisons have shown that, in all these settings, *all* and *behav* caused a higher accuracy than *phys* (all $p < .0001$; Figure 8a-8d). No effect of the feature type was found in *Mikel-high* (Figure 8e). The ANOVA also showed that there was a significant effect of the time window on accuracy in the settings of *valence-medium* [$F_{4,75} = 15.15$, $p < .0001$] and *arousal-medium* [$F_{4,75} = 11.99$, $p < .0001$]. Post-hoc tests have shown that in both cases, the *0-8 sec* time window caused a lower classification performance than all other windows for *valence-medium* ($p < .02$) and than all other windows but *0-20 sec* for *arousal-medium* ($p < .03$). Our analysis did not find an effect of the time windows on classification accuracy in other settings. Similarly, the ANOVA found interaction effects between feature type and time window for *valence-medium* [$F_{8,75} = 2.57$, $p < .02$] and *arousal-medium* [$F_{8,75} = 3.0$, $p < .006$]. Results of post-hoc tests are in line with previously reported pairwise comparisons of other main effects, i.e., behavioral features outperformed physiological features for the same time windows.

To confirm how changes in cluster-category image sets impact emotional state detectability, we assessed the prediction performance of self-reports within the current set by comparing features computed over the *0-20 sec* time window of the previous and current sets. A Wilcoxon signed-rank test revealed that there was a significant difference between the accuracy of the features attained from the current and previous valence-arousal cluster for *valence-low* [$Z = 0.0$, $p < .001$], *arousal-low* [$Z = 1.0$, $p < .001$], *valence-medium* [$Z = 11.0$, $p < .001$] and *arousal-medium* [$Z = 15.0$, $p < .002$]. No significant difference was found for *Mikel-high*.

To further unpack the indicated differences in terms of feature types, we inspected the impurity-based feature importances computed for the Random Forest classifier trained on all feature types in all model-resolution settings and time windows. We utilized the Random Forest for this analysis, as it exhibited the highest accuracy among classifiers across the various settings (see Appendix C). Table 3 shows the 30 most representative features according to their impurity-based feature importance averaged over all runs. Results affirmed that behavioral

Table 3. Summary of the 30 most representative features according to their impurity-based feature importances of the Random Forest, trained on and averaged over all model-resolution settings and time windows. The table shows their data type, feature group, the feature and its summary statistics. It reports the median rank and average importance score per feature group and data type.

Type	Group	Features (summary statistics)	Ranks	Median rank	Average score
behavioral	**Phone int.**	screen times (mean, st. dev., max, min), number images, time stamps (max)	1,2,4,8, 18,20	6	0.037
	Motion	acceleration (max) angular speed (max)	13,19	16	0.030
	Touch	speed (mean, min, max), duration(mean, st. dev., min, max), pressure decline (min), pressure (st. dev., min), number, acceleration (mean, min)	5,6,7,9,10, 11,12,15,16, 17,23,25,26, 29,30	15	0.031
	Facial exp.	angry expression	3	3	0.041
	Eye tracking	right eye's aspect ratio (mean)	14	14	0.030
	All			13	0.034
physiological	**ECG**	HR (max)	21	21	0.028
	PPG	HRV (st. dev.)	27	27	0.027
	EDA	phasic component	28	28	0.027
	RSP	respiration rate (min, max)	23	23	0.028
	All			24	0.028

features exhibit a greater predictive capability compared to physiological features. They possessed a higher count (despite being balanced in total number), a higher median rank and a superior average feature importance score.

We observed that only two features computed from images captured by the front-facing camera are on the list. To assess the impact of not utilizing the camera, we trained the Random Forest using the same configuration as previously, but this time excluding all features derived from the camera's images. Results show a slight decrease in the mean accuracy across all time windows and resolution-model pairs (with cam: 0.48; with no cam: 0.46)

5.5 Additional information (**F4**)

In the final stage of our analysis, we contrasted classification performance across different settings with varying degrees of available information in different dimensions. Thus, in terms of emotion model and resolution, we

Table 4. Results attained by systematically varying available information when training the Random Forest classifier. The table shows results for 5 model-resolution pairs (Valence-low, Arousal-low, Valence-medium, Arousal-medium, Mikel-high), 2 feature sets (all, behavioral), 2 time windows (0-8 sec, 8-16 sec), and 5 different settings of additional information (no additional information [no], demographic data, personality-related data, image emotions, their combination). The best performance per combination of a model-resolution pair, feature set and time window is highlighted in bold. If in one of these combinations two settings perform similarly, the one using the lower amount of information is highlighted.

Model	**Res.**	**ZeroR**	**Info.**	**0 - 8 sec**				**8 - 16 sec**			
				all		**behav.**		**all**		**behav.**	
				acc.	**F1**	**acc.**	**F1**	**acc.**	**F1**	**acc.**	**F1**
Valence	**low**	**0.28**	**no**	0.72	0.7	0.71	0.7	0.73	0.7	0.7	0.7
			demographic	0.73	0.7	0.75	0.8	0.72	0.7	0.73	0.7
			personality	0.75	0.7	0.72	0.7	0.72	0.7	0.7	0.7
			image emotions	**0.76**	**0.8**	**0.76**	**0.8**	0.74	0.7	**0.83**	**0.8**
			combination	0.75	0.7	0.76	0.8	**0.77**	**0.7**	0.77	0.8
Arousal	**low**	**0.28**	**no**	0.72	0.7	0.7	0.7	0.69	0.7	0.72	0.7
			demographic	0.68	0.7	0.67	0.7	0.69	0.6	0.73	0.7
			personality	**0.73**	**0.7**	0.7	0.7	0.7	0.7	0.72	0.7
			image emotions	0.73	0.7	0.76	0.8	0.77	0.7	0.78	0.8
			combination	0.72	0.7	**0.78**	**0.8**	**0.79**	**0.8**	**0.8**	**0.8**
Valence	**med.**	**0.28**	**no**	0.42	0.4	0.47	0.5	0.46	0.4	0.47	0.5
			demographic	0.46	0.5	0.34	0.3	0.53	0.5	0.52	0.5
			personality	0.39	0.4	0.44	0.4	0.51	0.5	0.51	0.5
			image emotions	0.49	0.5	0.48	0.5	0.58	0.5	**0.6**	**0.6**
			combination	**0.51**	**0.5**	**0.53**	**0.5**	**0.59**	**0.6**	0.59	0.6
Arousal	**med.**	**0.28**	**no**	0.47	0.4	0.46	0.4	0.49	0.5	0.53	0.5
			demographic	0.35	0.3	0.42	0.4	0.51	0.5	0.52	0.5
			personality	0.35	0.4	0.39	0.4	0.52	0.5	0.47	0.5
			image emotions	0.49	0.5	**0.54**	**0.5**	0.59	0.6	**0.59**	**0.6**
			combination	**0.54**	**0.5**	0.51	0.5	**0.62**	**0.6**	0.57	0.5
Mikel	**high**	**0.08**	**no**	0.45	0.4	0.48	0.4	0.53	0.4	**0.53**	**0.4**
			demographic	**0.52**	**0.4**	**0.55**	**0.4**	0.53	0.4	0.53	0.4
			personality	0.52	0.4	0.52	0.4	**0.56**	**0.5**	0.53	0.4
			image emotions	0.52	0.4	0.55	0.4	0.56	0.4	0.53	0.4
			combination	0.52	0.4	0.52	0.4	0.5	0.4	0.5	0.4

again used our five previously selected model-resolution pairs (*valence-low*, *arousal-low*, *valence-medium*, *arousal-medium*, and *Mikel-high*). To vary information regarding type of features, we decided to compare the full feature set with behavioral features, as the latter outperformed physiological features. All time windows were considered for this analysis as comparison of accuracies was inconclusive in this case (0-8 sec, 4-12 sec, 8-16 sec, 12-20 sec, 0-20 sec). Finally, we considered the additional information that we have identified as promising for enhancing emotion recognition (see Section 3). In particular, we decided to analyze the influence of demographic data (gender, age, sexual orientation), personality-related information [71, 85], predicted image emotions [101] and their combination on classification performance. Subsequently, we trained the Random Forest on the resulting 250 distinct settings (5 model-resolution pairs x 2 feature sets x 5 time windows x 5 different type of additional information). To ensure optimal performance, we conducted a grid search on its parameters including the number of trees, tree depth, and whether boosting is employed. Table 4 summarizes the results of this analysis. For the sake of clarity, the table only shows results from the 0-8 sec (shortest time period after stimulus change) and the 8-16 sec time windows (best performance).

Results revealed that for low resolutions of valence and arousal behavioral features outperformed or performed on par with the full feature set for both time windows. This observation holds for medium resolutions except for the case of arousal classification in the 8 - 16 sec time window. In terms of high resolution, no clear trend in terms of feature set can be observed. The analysis also showed that the later time window resulted in higher classification accuracies across all resolutions and models. However, these differences are minor, with an average difference in accuracy of 0.04 across all setting and 0.09 for medium state resolutions, for which a significant difference in the pairwise comparison of the these time windows was found (Section 5.4). Furthermore, it was observed that including predicted image emotions in the feature set boosted classification performance across all emotion models, time windows and resolutions. In contrast, for the other forms of additional information (demographic, personality-related), we did not observe a clear trend.

6 DISCUSSION

Our experiment assessed the effect of four key factors on the performance of emotion recognition during passive social media use: (**F1**) data of physiological activity, (**F2**) the time window for data collection, (**F3**) the impact of emotion models, and (**F4**) user-specific factors as well as inferred emotion distributions from images. We discuss our findings in the light of these factors and their implication for future work on emotion recognition during passive social media use.

(**F1**): Emotion recognition using behavioral versus physiological data

Results revealed that across emotion models, time windows, and the predicted range of emotional states, behavioral features were sufficient to recognize participants' emotions during passive social media use, even outperforming the combination of behavioral and physiological features in some settings. Combined with the fact that both feature types displayed a similar response to cluster configurations and a comparable correlation with valence and arousal ratings, our findings suggest that behavioral features hold greater predictive power for this task.

We hypothesize that the observed finding could be due to the delayed manifestation of responses to a stimulus in physiological signals compared to behavioral signals. In terms of behavioral signals, facial expressions adjust to stimuli within only 0.3 to 0.4 seconds [14] and fixation behavior changes within 0.5 seconds after exposure to an image [83]. In our study, fixation behavior is captured by the various screen time metrics, e.g., the time a participant looked at a post. In contrast, physiological features demonstrate a longer reaction time, with heart rate requiring approximately 6 seconds to respond [4] and EDA varying between 3 to 5 seconds upon exposure to a stimulus [6]. These response times were derived from conventional stimulus-response studies where emotionally distinct stimuli are presented after neutral ones. Our study design excludes a neutralization phase between stimuli

of differing polarity to better reflect real-world social media use. We believe that this adjustment further amplified the disparities in stimulus responses of physiological and behavioral features, offering a potential explanation for why behavioral data proved to be more indicative of participants' emotions across settings. Future research should explore how omitting a neutralization phase between two consecutive emotional stimuli effects responses in physiological and behavioral signals, in a more controlled setting.

As behavioral data can be gathered directly from the devices used for social media consumption (e.g., smartphones), our findings demonstrate that users' emotional states during passive social media use can be reliably identified solely based on information that is immediately available on these devices. This eliminates the necessity of equipping the user with additional wearables. Note that, in this study, experimental sensors were used to collect physiological data, for which a superior signal quality in comparison to the sensors found on consumer-grade devices can be assumed. Thus, we expect found differences in terms of feature type to even be more pronounced in real-world settings where, furthermore, motion artifacts will affect signal quality.

(**F2**): Response delays to changing content

Our results show that stimuli from different valence-arousal clusters caused differences in self-reported affect and in distributions of collected physiological and behavioral features. They also show that physiological and behavioral features did not serve as reliable predictors across distinct valence-arousal clusters. This indicates the dependency of participants' emotional state on observed content and, further, suggests emotional contagion across different stimuli groups. Nevertheless, the analysis of peak performances revealed that the classifier achieved similar levels of performance for emotion recognition with features from just the initial 8 seconds as input as it did with features from the best-performing time window (8–16 seconds). The consistent but slight improvement in emotion detection performance in the later time window, along with the difference in performance between the initial 8 seconds and the subsequent time frames in medium-resolution settings, characterize an interesting trade-off between the promptness at which emotional states can be detected and the accuracy of their recognition. While these findings stand in contrast to the peak-end rule [22], they are in line with Di Lascio et al.'s results, who observed that different time windows can be most indicative for self-reported experiences [13].

Our results indicate that during passive social media use, users' responses to changes in content can be detected with sufficient robustness as early as within 8 seconds. This discovery holds significant implications for social media applications supporting users' mental health. It enables the association of a user's current emotional state with a specific post, provided the observation time exceeds 8 seconds. Subsequently, these applications could dynamically respond to the content and tailor interventions accordingly. On social media, users typically spend less than 3 seconds on a piece of content [20]. Despite these brief durations of exposure, they can still influence their emotional state [108]. Future research should, thus, extend our research by investigating the association of content and users' emotional state for even shorter durations of exposure.

(**F3**): Effect of emotion model on emotion recognition

For a high state granularity, our analysis showed no discernible differences between emotion models despite the higher number of predicted states of Mikel's Wheel compared to the Circumplex Model. This indicates the superiority of Mikel's Wheel in high-resolutional settings. We explain this finding by the compound nature of the emotion model. By enabling users to select multiple basic emotions, the model facilitates the expression of the nuances and ambiguities in their feelings.

Representing these subjective ratings in polar coordinate space (using a recent parameterization [101]), allowed us then to attain robust labels from these compound emotional states.

In settings of a low- or medium range of predicted emotional states, our results indicated no definite answer on which emotion model to use. However, we argue that the two dimensions of the Circumplex Model, valence

and arousal, provide more comprehensive insights into the specific constitution of a user's emotional state. The model further detects states of neutral affect—a capability that Mikel's Wheel does not possess. In addition, it caused lower variances of accuracy in the prediction of the emotional state.

Mikel's Wheel also allows attaining multi-class- as well as distributional labels from self-reports. Its positive impact on emotion recognition accuracy in our task and this representational flexibility, suggests testing Mikel's Wheel in other domains of affective computing as an interesting direction for future work.

(**F4**): Effect of additional inputs on emotion detection

Our findings reveal that using the emotional properties of images participants observed in the feed as input to the classifier increases emotion recognition peak performances across emotion models and resolutions. These properties were attained through an existing deep-learning-based approach [101], which predicts the distribution over discrete emotions experienced by a population of viewers in response to an image. Their positive impact on detection accuracy indicates that these distributions provide meaningful priors for the task of emotion recognition during passive social media use. This suggests their potential as a valuable feature for emotion detection in other application areas where users' emotions towards images are of interest.

In contrast, for the other forms of additional information (demographic, personality-related), we could not observe a clear trend. While participants varied in gender, nationality, and sexual orientation (see Section 4.6), minority groups were underrepresented. Future research needs to consider larger participant pools with balanced groups across different factors to establish conclusive findings.

Comparison with active social media use

To understand how the form of engagement, active or passive, impacts the recognition of emotions, we compare our results with the accuracies of *binary* emotional state estimation reported in closely related work [72, 73]. Both studies present accuracies in unconstrained settings of engagement (active and passive) in the range of 92–96%. The best result in terms of binary emotional state prediction in our experiment was 83% (binary valence, behavioral data, predicted image emotions, 8–16 sec). This minor decrease in performance holds despite the fact that passive social media use is a more challenging setting as users' behavioral traces are more sparse than in active social media use. In particular, passive use does not comprise typing behavior, which has been demonstrated to be a strong indicator of users' emotional state [95]. In the two studies of active social media use [72, 73], participants exhibited this behavior through activities such as commenting on posts. In our study of passive social media use, models cannot capitalize on this behavioral cue, potentially explaining the decline in performance compared to the settings of active use. Future work should further investigate differences in emotion recognition between active and passive use, by comparing them with identical feature sets and settings.

Privacy- & ethical considerations

In our study, we leveraged features computed from images captured by the front-facing camera as inputs to the classifier, as they have demonstrated utility in emotion detection during active social media use [72]. Similarly, we analyzed factors that are known to influence social media behavior, i.e., gender [92], nationality [50], sexual orientation [19], and personality [37]. We recognize the sensitivity of this information and the severe consequences if it were to be misused. Fortunately, our results indicated that excluding camera-based features led to only a marginal 2% decrease in emotion detection accuracy. Excluding demographic- and personality-related information resulted in an even smaller performance decrease of 0.3% (averaged across settings in Table 4). This suggests that emotions during passive social media use can robustly be detected even without these features.

While our aim to use emotion detection systems for enhancing digital well-being, we underscore the potential of this research to be misused in manipulating users' emotions in accordance with malicious objectives. By

increasing the understanding of the factors that contribute to a robust emotion detection, our research can inform legislators in creating laws to protect user data that might otherwise be exploited for emotional manipulation. This could become relevant, e.g., in the development of legislation under the European Union AI Act, which also plans to regulate emotion detection systems [63]. Moreover, our research constitutes to raising awareness among users about exposing their emotional state when browsing online. Both aspects are crucial, as already in today's internet, black-box algorithms utilize content that triggers highly arousing emotions to increase people's time online for the financial benefit of platform providers [100].

7 LIMITATIONS & FUTURE RESEARCH

Our study was conducted in a controlled social media environment with fewer features and influences than found on actual platforms. In social media, various factors in addition to the shown content influence users' emotions, such as captions, relationships to the author of a post, likes, reposts, recency, prior exposure, and others. To select from this high-dimensional space a manageable subset of factors for controlled evaluation, we used affective images and replicated key interaction features like infinite scroll to establish an simulated real-world social media scenario. While this design choice impacted the external validity of our study, it increased its internal validity by allowing us to control the stimuli presented to participants. Thus, we could ensure that all participants saw the same posts and reduce external influences, e.g, familiarity with a stimulus. Recognizing the significance of unexplored dimensions and features, we recommend future research to investigate their influence on emotion detectability during passive social media use to complement our current assessment.

One limitation of our study is the small sample size of 26 participants that may limit the reliability of observed effects. Although we gained valuable preliminary insights into the emotional state detection during passive social media use, future studies with larger participant pools are warranted to validate and extend our findings.

While our study sheds light on the emotional state detection of users aged 18 to 35 years, the limited age range restricts the generalizability of our findings to other age groups. Given that problematic internet use predominantly affects adolescents [2, 48], they stand to gain the most from applications that facilitate digital health based on the detected emotional state. Therefore, expanding the age range to encompass this group would be particularly insightful. Future research should include a broader spectrum of ages to capture nuances in the emotional state detection across different age periods.

Our model attains accuracies comparable to those reported in other recent studies on wearable emotion recognition, addressing both binary [15] and three-class emotion detection problems [15, 96]. However, more advanced models for emotional state recognition exist that further improve prediction performance [46, 47]. Future studies should explore whether our findings on supervised emotion detection also hold for these methods.

8 CONCLUSION

In this study, we addressed the challenge of robust emotional state recognition during passive social media engagement, a mode of interaction that has not been extensively explored in previous research. Our experiment involved 29 participants browsing a controlled social media feed, where they were exposed to typical content of social media stemming from a standardized emotional database.

Our findings show that behavioral features, derived from participants' interaction with the phone, outperform physiological signals in informing emotion classifiers. This underscores the practicality of utilizing behavioral data, attainable on every smartphone, for informing, e.g., digital self-control tools. Additionally, we observed that within 8 seconds following a change in media content, objective features can discern a participant's new emotional state. This would allow social media applications supporting mental health to link a user's current emotional state to a specific post for observation durations of 8 seconds or longer.

Our investigation of two validated emotion models, the Circumplex Model of Affect and Mikel's Wheel, shed light on their respective strengths. Mikel's Wheel proves particularly effective in high state granularity settings, allowing users to express nuances and ambiguities in their feelings by selecting multiple basic emotions. In contrast, the Circumplex Model's dimensions of valence and arousal offer more comprehensive insights into a user's emotional state in low- or medium-granularity settings. Lastly, we examined the impact of supplemental user- and content-specific information on emotion detection performance. We found that leveraging image emotions predicted by a deep learning model significantly boosts classification accuracy.

In summary, our study provides valuable insights for robust emotional state recognition during passive social media use. These findings have implications for the development of systems that support users' digital health, emphasizing the importance of considering behavioral features for accurate emotion detection.

ACKNOWLEDGEMENTS

We are grateful for the valuable feedback of Max Möbus and Shkurta Gashi on analysis design and paper drafts.

REFERENCES

[1] Kemal Jabir Abdullah, Izzat Aulia Akbar, Bambang Setiawan, Febriliyan Samopa, and Nisfu Asrul Sani. 2022. Analysis of the Effect of Comedic Film on Changes of Heart Rate Using Photoplethysmogram and Electrocardiogram. *Procedia Comput. Sci.* 197, C (jan 2022), 208–214. https://doi.org/10.1016/j.procs.2021.12.133

[2] Hosam Al-Samarraie, Kirfi-Aliyu Bello, Ahmed Ibrahim Alzahrani, Andrew Paul Smith, and Chikezie Emele. 2021. Young users' social media addiction: causes, consequences and preventions. *Information Technology & People* (2021). https://doi.org/10.1108/itp-11-2020-0753

[3] Nitin Anand, Manoj Kumar Sharma, Pranjali Chakraborty Thakur, Ishita Mondal, Maya Sahu, Priya Singh, Ajith S. J., Jayesh Suresh Kande, Neeraj MS, and Ripudaman Singh. 2022. Doomsurfing and doomscrolling mediate psychological distress in COVID-19 lockdown: Implications for awareness of cognitive biases. *Perspectives in Psychiatric Care* 58, 1 (Jan. 2022), 170–172. https://doi.org/10.1111/ppc.12803

[4] Jenni Anttonen and Veikko Surakka. 2005. Emotions and Heart Rate While Sitting on a Chair. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Portland, Oregon, USA) *(CHI '05)*. Association for Computing Machinery, New York, NY, USA, 491–499. https://doi.org/10.1145/1054972.1055040

[5] Natalya N. Bazarova, Yoon Hyung Choi, Victoria Schwanda Sosik, Dan Cosley, and Janis Whitlock. 2015. Social Sharing of Emotions on Facebook: Channel Differences, Satisfaction, and Replies. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 154–164. https://doi.org/10.1145/2675133.2675297

[6] Wolfram Boucsein. 2012. *Electrodermal activity*. Springer Science & Business Media.

[7] Margaret M. Bradley and Peter J. Lang. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25, 1 (March 1994), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

[8] Moira Burke and Robert Kraut. 2013. Using Facebook after Losing a Job: Differential Benefits of Strong and Weak Ties. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (San Antonio, Texas, USA) *(CSCW '13)*. Association for Computing Machinery, New York, NY, USA, 1419–1430. https://doi.org/10.1145/2441776.2441936

[9] Jan Cech and Tereza Soukupova. 2016. Real-time eye blink detection using facial landmarks. *Cent. Mach. Perception, Dep. Cybern. Fac. Electr. Eng. Czech Tech. Univ. Prague* (2016), 1–8.

[10] Shirley Cramer and Becky Inkster. 2017. *Status of Mind: Social media and young people's mental health.* Online Report. Royal Society For Public Health. https://www.rsph.org.uk/static/uploaded/d125b27c-0b62-41c5-a2c0155a8887cd01.pdf

[11] DataReportal, & We Are Social, & Meltwater. 2023. Most popular websites worldwide as of November 2022, by total visits. https://www.statista.com/statistics/1201880/most-visited-websites-worldwide/

[12] Dian A. de Vries, A. Marthe Möller, Marieke S. Wieringa, Anniek W. Eigenraam, and Kirsten Hamelink. 2018. Social Comparison as the Thief of Joy: Emotional Consequences of Viewing Strangers' Instagram Posts. *Media Psychology* 21, 2 (April 2018), 222–245. https://doi.org/10.1080/15213269.2016.1267647

[13] Elena Di Lascio, Shkurta Gashi, and Silvia Santini. 2018. Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–21.

[14] Ulf Dimberg and Monika Thunberg. 1998. Rapid facial reactions to emotional facial expressions. *Scandinavian journal of psychology* 39, 1 (1998), 39–45.

[15] Vipula Dissanayake, Sachith Seneviratne, Rajib Rana, Elliott Wen, Tharindu Kaluarachchi, and Suranga Nanayakkara. 2022. SigRep: Toward Robust Wearable Emotion Recognition With Contrastive Representation Learning. *IEEE Access* 10 (2022), 18105–18120. https://doi.org/10.1109/ACCESS.2022.3149509

[16] Jenna Drenten, Lauren Gurrieri, and Meagan Tyler. 2020. Sexualized labour in digital culture: Instagram influencers, porn chic and the monetization of attention. *Gender, Work & Organization* 27, 1 (2020), 41–66.

[17] Paul Ekman and Wallace V. Friesen. 1971. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology* 17 (1971), 124–129. https://doi.org/10.1037/h0030377

[18] Mohamed Elgendi, Ian Norton, Matt Brearley, Derek Abbott, and Dale Schuurmans. 2013. Systolic Peak Detection in Acceleration Photoplethysmograms Measured from Emergency Responders in Tropical Conditions. *PLOS ONE* 8, 10 (Oct. 2013), e76585. https://doi.org/10.1371/journal.pone.0076585 Publisher: Public Library of Science.

[19] César G. Escobar-Viera, Ariel Shensa, Jaime Sidani, Brian Primack, and Michael P. Marshal. 2020. Association Between LGB Sexual Orientation and Depression Mediated by Negative Social Media Experiences: National Survey Study of US Young Adults. *JMIR Mental Health* 7, 12 (Dec. 2020), e23520. https://doi.org/10.2196/23520

[20] Facebook. 2016. Capturing Attention in Feed: The Science Behind Effective Video Creative. https://www.facebook.com/business/news/insights/capturing-attention-feed-video-creative. Accessed on 11/08/2023.

[21] Stephen H. Fairclough. 2009. Fundamentals of physiological computing. *Interacting with Computers* 21, 1 (Jan. 2009), 133–145. https://doi.org/10.1016/j.intcom.2008.10.011

[22] Barbara L Fredrickson and Daniel Kahneman. 1993. Duration neglect in retrospective evaluations of affective episodes. *Journal of personality and social psychology* 65, 1 (1993), 45.

[23] David Garcia, Arvid Kappas, Dennis Küster, and Frank Schweitzer. 2016. The dynamics of emotions in online interaction. *Royal Society open science* 3, 8 (2016).

[24] Shkurta Gashi, Elena Di Lascio, Bianca Stancu, Vedant Das Swain, Varun Mishra, Martin Gjoreski, and Silvia Santini. 2020. Detection of artifacts in ambulatory electrodermal activity data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (2020), 1–31.

[25] Amit Goldenberg and James J Gross. 2020. Digital emotion contagion. *Trends in cognitive sciences* 24, 4 (2020), 316–328.

[26] Han-Wen Guo, Yu-Shun Huang, Chien-Hung Lin, Jen-Chien Chien, Koichi Haraikawa, and Jiann-Shing Shieh. 2016. Heart Rate Variability Signal Features for Emotion Recognition by Using Principal Component Analysis and Support Vectors Machine. In *2016 IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE)*. 274–277. https://doi.org/10.1109/BIBE.2016.40

[27] Elaine C. S. Hayashi, Julián E. Gutiérrez Posada, Vanessa R. M. L. Maike, and M. Cecília C. Baranauskas. 2016. 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)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3033701.3033728

[28] Jennifer Healey, Lama Nachman, Sushmita Subramanian, Junaith Shahabdeen, and Margaret Morris. 2010. Out of the lab and into the fray: Towards modeling emotion in everyday life. In *Pervasive Computing: 8th International Conference, Pervasive 2010, Helsinki, Finland, May 17-20, 2010. Proceedings 8*. Springer, 156–173.

[29] Alexander Heimerl, Linda Becker, Dominik Schiller, Tobias Baur, Fabian Wildgrube, Nicolas Rohleder, and Elisabeth Andre. 2022. We've Never Been Eye to Eye: A Pupillometry Pipeline for the Detection of Stress and Negative Affect in Remote Working Scenarios. In *Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments* (Corfu, Greece) *(PETRA '22)*. Association for Computing Machinery, New York, NY, USA, 486–493. https://doi.org/10.1145/3529190.3534729

[30] Murtadha D. Hssayeni and Behnaz Ghoraani. 2021. Multi-Modal Physiological Data Fusion for Affect Estimation Using Deep Learning. *IEEE Access* 9 (2021), 21642–21652. https://doi.org/10.1109/ACCESS.2021.3055933

[31] Yuheng Hu, Lydia Manikonda, and Subbarao Kambhampati. 2014. What We Instagram: A First Analysis of Instagram Photo Content and User Types. In *Eighth International AAAI Conference on Weblogs and Social Media*. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8118

[32] Carroll E. Izard. 1977. *Human Emotions*. Springer US, Boston, MA. https://doi.org/10.1007/978-1-4899-2209-0

[33] William James. 1948. What is emotion? 1884. (1948).

[34] D. Khodadad, S. Nordebo, B. Müller, A. Waldmann, R. Yerworth, T. Becher, I. Frerichs, L. Sophocleous, A. van Kaam, M. Miedema, N. Seifnaraghi, and R. Bayford. 2018. Optimized breath detection algorithm in electrical impedance tomography. *Physiological Measurement* 39, 9 (Sept. 2018), 094001. https://doi.org/10.1088/1361-6579/aad7e6 Publisher: IOP Publishing.

[35] Thomas Kosch, Mariam Hassib, Robin Reutter, and Florian Alt. 2020. Emotions on the Go: Mobile Emotion Assessment in Real-Time Using Facial Expressions. In *Proceedings of the International Conference on Advanced Visual Interfaces* (Salerno, Italy) *(AVI '20)*. Association for Computing Machinery, New York, NY, USA, Article 18, 9 pages. https://doi.org/10.1145/3399715.3399928

[36] Thomas Kosch, Mariam Hassib, Robin Reutter, and Florian Alt. 2020. Emotions on the Go: Mobile Emotion Assessment in Real-Time using Facial Expressions. In *Proceedings of the International Conference on Advanced Visual Interfaces (AVI '20)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3399715.3399928

[37] Michal Kosinski, David Stillwell, and Thore Graepel. 2013. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences* 110, 15 (2013), 5802–5805. https://doi.org/10.1073/pnas.1218772110 arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.1218772110

[38] Nikola Kovačević, Christian Holz, Tobias Günther, Markus Gross, and Rafael Wampfler. 2023. Personality Trait Recognition Based on Smartphone Typing Characteristics in the Wild. *IEEE Transactions on Affective Computing* (2023), 1–11. https://doi.org/10.1109/TAFFC.2023.3253202

[39] Agata Kołakowska, Agnieszka Landowska, Mariusz Szwoch, Wioleta Szwoch, and Michał Wróbel. 2015. Modeling emotions for affect-aware applications. In *Information Systems Development and Applications*. University of Gdańsk, Poland, 55–69. https://wzr.ug.edu.pl/nauka/upload/files/Informationsystemsdevelopmentandapplications.pdf

[40] Agata Kołakowska, Wioleta Szwoch, and Mariusz Szwoch. 2020. A Review of Emotion Recognition Methods Based on Data Acquired via Smartphone Sensors. *Sensors (Basel, Switzerland)* 20, 21 (Nov. 2020), 6367. https://doi.org/10.3390/s20216367

[41] Hanna Krasnova, Helena Wenninger, Thomas Widjaja, and Peter Buxmann. 2013. Envy on Facebook: a hidden threat to users' life satisfaction?. In *Proceedings of the 11th International Conference on Wirtschaftsinformatik (WI2013)*, Vol. 2. Universität Leipzig, Leipzig, Germany, 1477–1491. https://boris.unibe.ch/47080/

[42] Sylvia D. Kreibig. 2010. Autonomic nervous system activity in emotion: A review. *Biological Psychology* 84, 3 (July 2010), 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010

[43] Hoa T. Le and Larry A. Vea. 2016. A Customer Emotion Recognition through Facial Expression Using Kinect Sensors v1 and v2: A Comparative Analysis. In *Proceedings of the 10th International Conference on Ubiquitous Information Management and Communication* (Danang, Viet Nam) *(IMCOM '16)*. Association for Computing Machinery, New York, NY, USA, Article 80, 7 pages. https://doi.org/10.1145/2857546.2857628

[44] Helmut Leder, Jussi Hakala, Veli-Tapani Peltoketo, Christian Valuch, and Matthew Pelowski. 2022. Swipes and Saves: A Taxonomy of Factors Influencing Aesthetic Assessments and Perceived Beauty of Mobile Phone Photographs. *Frontiers in Psychology* 13 (2022). https://www.frontiersin.org/articles/10.3389/fpsyg.2022.786977

[45] Hosub Lee, Young Sang Choi, Sunjae Lee, and I. P. Park. 2012. Towards unobtrusive emotion recognition for affective social communication. In *2012 IEEE Consumer Communications and Networking Conference (CCNC)*. 260–264. https://doi.org/10.1109/CCNC.2012.6181098

[46] Chao Li, Zhongtian Bao, Linhao Li, and Ziping Zhao. 2020. Exploring temporal representations by leveraging attention-based bidirectional LSTM-RNNs for multi-modal emotion recognition. *Information Processing & Management* 57, 3 (2020), 102185. https://doi.org/10.1016/j.ipm.2019.102185

[47] Chao Li, Boyang Chen, Ziping Zhao, Nicholas Cummins, and Björn W. Schuller. 2021. Hierarchical Attention-Based Temporal Convolutional Networks for Eeg-Based Emotion Recognition. In *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 1240–1244. https://doi.org/10.1109/ICASSP39728.2021.9413635

[48] Liu Yi Lin, Jaime E Sidani, Ariel Shensa, Ana Radovic, Elizabeth Miller, Jason B Colditz, Beth L Hoffman, Leila M Giles, and Brian A Primack. 2016. Association between social media use and depression among US young adults. *Depression and anxiety* 33, 4 (2016), 323–331.

[49] Ulrik Lyngs, Kai Lukoff, Laura Csuka, Petr Slovák, Max Van Kleek, and Nigel Shadbolt. 2022. The Goldilocks level of support: Using user reviews, ratings, and installation numbers to investigate digital self-control tools. *International journal of human-computer studies* 166 (2022), 102869.

[50] Shilpa Madan, Shankha Basu, Sharon Ng, and Elison Ai Ching Lim. 2018. Impact of Culture on the Pursuit of Beauty: Evidence from Five Countries. *Journal of International Marketing* 26, 4 (2018), 54–68. https://www.jstor.org/stable/26979336 Publisher: [Sage Publications, Inc., American Marketing Association].

[51] Dominique Makowski, Tam Pham, Zen J. Lau, Jan C. Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and S. H. Annabel Chen. 2021. NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior Research Methods* 53, 4 (Aug. 2021), 1689–1696. https://doi.org/10.3758/s13428-020-01516-y

[52] Artur Marchewka, Łukasz Żurawski, Katarzyna Jednoróg, and Anna Grabowska. 2014. The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods* 46, 2 (June 2014), 596–610. https://doi.org/10.3758/s13428-013-0379-1

[53] Yuki Matsuda, Dmitrii Fedotov, Yuta Takahashi, Yutaka Arakawa, Keiichi Yasumoto, and Wolfgang Minker. 2018. EmoTour: Estimating emotion and satisfaction of users based on behavioral cues and audiovisual data. *Sensors* 18, 11 (2018), 3978.

[54] Maurizio Mauri, Pietro Cipresso, Anna Balgera, Marco Villamira, and Giuseppe Riva. 2011. Why Is Facebook So Successful? Psychophysiological Measures Describe a Core Flow State While Using Facebook. *Cyberpsychology, Behavior, and Social Networking* 14, 12 (Dec. 2011), 723–731. https://doi.org/10.1089/cyber.2010.0377

[55] A. Mehrabian and J. A. Russell. 1974. The basic emotional impact of environments. *Perceptual and Motor Skills* 38, 1 (Feb. 1974), 283–301. https://doi.org/10.2466/pms.1974.38.1.283

[56] Abhinav Mehrotra, Fani Tsapeli, Robert Hendley, and Mirco Musolesi. 2017. MyTraces: Investigating Correlation and Causation between Users' Emotional States and Mobile Phone Interaction. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 83

(sep 2017), 21 pages. https://doi.org/10.1145/3130948

[57] JOSEPH A. MIKELS, BARBARA L. FREDRICKSON, GREGORY R. LARKIN, CASEY M. LINDBERG, SAM J. MAGLIO, and PATRICIA A. REUTER-LORENZ. 2005. Emotional category data on images from the International Affective Picture System. *Behavior Research Methods* 37, 4 (Nov. 2005), 626–630. https://doi.org/10.3758/BF03192732

[58] Aske Mottelson and Kasper Hornbæk. 2016. An affect detection technique using mobile commodity sensors in the wild. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 781–792. https://doi.org/10.1145/2971648.2971654

[59] Aske Mottelson, Jarrod Knibbe, and Kasper Hornbæk. 2018. Veritaps: Truth Estimation from Mobile Interaction. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3174135

[60] Mathieu Pagé Fortin and Brahim Chaib-draa. 2019. Multimodal multitask emotion recognition using images, texts and tags. In *Proceedings of the ACM Workshop on Crossmodal Learning and Application*. 3–10.

[61] Francis C. Panganiban and Franz A. de Leon. 2021. Stress Detection Using Smartphone Extracted Photoplethysmography. In *2021 IEEE Region 10 Symposium (TENSYMP)*. 1–7. https://doi.org/10.1109/TENSYMP52854.2021.9550905 ISSN: 2642-6102.

[62] Galen Panger. 2018. People Tend to Wind Down, Not Up, When They Browse Social Media. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 133 (nov 2018), 29 pages. https://doi.org/10.1145/3274402

[63] European Parlament. 2023. Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI. https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai. Accessed on 01/31/2024.

[64] Panagiotis C Petrantonakis and Leontios J Hadjileontiadis. 2010. Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis. *IEEE Transactions on affective computing* 1, 2 (2010), 81–97.

[65] Phuong Pham and Jingtao Wang. 2017. Understanding Emotional Responses to Mobile Video Advertisements via Physiological Signal Sensing and Facial Expression Analysis. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces (IUI '17)*. Association for Computing Machinery, New York, NY, USA, 67–78. https://doi.org/10.1145/3025171.3025186

[66] Rosalind W Picard. 2000. *Affective computing*. MIT press.

[67] Martin Pielot, Tilman Dingler, Jose San Pedro, and Nuria Oliver. 2015. When Attention is Not Scarce - Detecting Boredom from Mobile Phone Usage. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 825–836. https://doi.org/10.1145/2750858.2804252

[68] Robert Plutchik. 2001. The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist* 89, 4 (2001), 344–350. https://www.jstor.org/stable/27857503

[69] Juan Carlos Quiroz, Elena Geangu, and Min Hooi Yong. 2018. Emotion Recognition Using Smart Watch Sensor Data: Mixed-Design Study. *JMIR Mental Health* 5, 3 (Aug. 2018), e10153. https://doi.org/10.2196/10153 Company: JMIR Mental Health Distributor: JMIR Mental Health Institution: JMIR Mental Health Label: JMIR Mental Health Publisher: JMIR Publications Inc., Toronto, Canada.

[70] Martin Ragot, Nicolas Martin, Sonia Em, Nico Pallamin, and Jean-Marc Diverrez. 2018. Emotion Recognition Using Physiological Signals: Laboratory vs. Wearable Sensors. In *Advances in Human Factors in Wearable Technologies and Game Design (Advances in Intelligent Systems and Computing)*, Tareq Ahram and Christianne Falcão (Eds.). Springer International Publishing, Cham, 15–22. https://doi.org/10.1007/978-3-319-60639-2_2

[71] Morris Rosenberg. 1965. *Society and the Adolescent Self-Image*. Princeton University Press, Princeton, USA. https://www.jstor.org/stable/j.ctt183pjjh

[72] Mintra Ruensuk, Eunyong Cheon, Hwajung Hong, and Ian Oakley. 2020. How Do You Feel Online: Exploiting Smartphone Sensors to Detect Transitory Emotions during Social Media Use. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (Dec. 2020), 150:1–150:32. https://doi.org/10.1145/3432223

[73] Mintra Ruensuk, Taewan Kim, Hwajung Hong, and Ian Oakley. 2022. Sad or just jealous? Using Experience Sampling to Understand and Detect Negative Affective Experiences on Instagram. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3491102.3517561

[74] James A. Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39 (1980), 1161–1178. https://doi.org/10.1037/h0077714

[75] Stanislaw Saganowski, Bartosz Perz, Adam Polak, and Przemyslaw Kazienko. 2022. Emotion Recognition for Everyday Life Using Physiological Signals from Wearables: A Systematic Literature Review. *IEEE Transactions on Affective Computing* 1, 1 (2022), 1–1. https://doi.org/10.1109/TAFFC.2022.3176135

[76] Philip Schmidt, Robert Dürichen, Attila Reiss, Kristof Van Laerhoven, and Thomas Plötz. 2019. Multi-Target Affect Detection in the Wild: An Exploratory Study. In *Proceedings of the 2019 ACM International Symposium on Wearable Computers* (London, United Kingdom) *(ISWC '19)*. Association for Computing Machinery, New York, NY, USA, 211–219. https://doi.org/10.1145/3341163.3347741

[77] Melanie Schreiner, Thomas Fischer, and Rene Riedl. 2021. Impact of content characteristics and emotion on behavioral engagement in social media: literature review and research agenda. *Electronic Commerce Research* 21 (2021), 329–345.

[78] Feri Setiawan, Sunder Ali Khowaja, Aria Ghora Prabono, Bernardo Nugroho Yahya, and Seok-Lyong Lee. 2018. A Framework for Real Time Emotion Recognition Based on Human ANS Using Pervasive Device. In *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, Vol. 01. 805–806. https://doi.org/10.1109/COMPSAC.2018.00129

[79] Feri Setiawan, Aria Ghora Prabono, Sunder Ali Khowaja, Wangsoo Kim, Kyoungsoo Park, Bernardo Nugroho Yahya, Seok-Lyong Lee, and Jin Pyo Hong. 2020. Fine-grained emotion recognition: fusion of physiological signals and facial expressions on spontaneous emotion corpus. *International Journal of Ad Hoc and Ubiquitous Computing* 35, 3 (Jan. 2020), 162–178. https://doi.org/10.1504/ijahuc.2020.110824

[80] Lin Shu, Jinyan Xie, Mingyue Yang, Ziyi Li, Zhenqi Li, Dan Liao, Xiangmin Xu, and Xinyi Yang. 2018. A Review of Emotion Recognition Using Physiological Signals. *Sensors* 18, 7 (July 2018), 2074. https://doi.org/10.3390/s18072074

[81] Fernando Silveira, Brian Eriksson, Anmol Sheth, and Adam Sheppard. 2013. Predicting audience responses to movie content from electro-dermal activity signals. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. 707–716.

[82] Petr Slovak, Alissa Antle, Nikki Theofanopoulou, Claudia Daudén Roquet, James Gross, and Katherine Isbister. 2023. Designing for Emotion Regulation Interventions: An Agenda for HCI Theory and Research. *ACM Trans. Comput.-Hum. Interact.* 30, 1, Article 13 (mar 2023), 51 pages. https://doi.org/10.1145/3569898

[83] Ashley R Smith, Simone P Haller, Sara A Haas, David Pagliaccio, Brigid Behrens, Caroline Swetlitz, Jessica L Bezek, Melissa A Brotman, Ellen Leibenluft, Nathan A Fox, et al. 2021. Emotional distractors and attentional control in anxious youth: eye tracking and fMRI data. *Cognition and Emotion* 35, 1 (2021), 110–128.

[84] H.R. Sneha, Mohammed Rafi, M.V. Manoj Kumar, Likewin Thomas, and B. Annappa. 2017. Smartphone based emotion recognition and classification. In *2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)*. 1–7. https://doi.org/10.1109/ICECCT.2017.8117872

[85] Christopher J. Soto and Oliver P. John. 2017. 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 (July 2017), 117–143. https://doi.org/10.1037/pspp0000096 Place: US Publisher: American Psychological Association.

[86] Luca Surace, Massimiliano Patacchiola, Elena Battini Sönmez, William Spataro, and Angelo Cangelosi. 2017. Emotion recognition in the wild using deep neural networks and Bayesian classifiers. In *Proceedings of the 19th ACM international conference on multimodal interaction*. 593–597.

[87] Shabbir Syed-Abdul, Luis Fernandez-Luque, Wen-Shan Jian, Yu-Chuan Li, Steven Crain, Min-Huei Hsu, Yao-Chin Wang, Dorjsuren Khandregzen, Enkhzaya Chuluunbaatar, Phung Anh Nguyen, and Der-Ming Liou. 2013. Misleading Health-Related Information Promoted Through Video-Based Social Media: Anorexia on YouTube. *Journal of Medical Internet Research* 15, 2 (Feb. 2013), e2237. https://doi.org/10.2196/jmir.2237

[88] Wioleta Szwoch. 2015. Emotion Recognition Using Physiological Signals. In *Proceedings of the Mulitimedia, Interaction, Design and Innnovation (MIDI '15)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/2814464.2814479

[89] Giuseppe Romano Tizzano, Matteo Spezialetti, and Silvia Rossi. 2020. A Deep Learning Approach for Mood Recognition from Wearable Data. In *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. 1–5. https://doi.org/10.1109/MeMeA49120.2020.9137218

[90] Goran Udovičić, Jurica Đerek, Mladen Russo, and Marjan Sikora. 2017. Wearable Emotion Recognition System Based on GSR and PPG Signals. In *Proceedings of the 2nd International Workshop on Multimedia for Personal Health and Health Care* (Mountain View, California, USA) *(MMHealth '17)*. Association for Computing Machinery, New York, NY, USA, 53–59. https://doi.org/10.1145/3132635.3132641

[91] Patti M Valkenburg, Irene I van Driel, and Ine Beyens. 2022. The associations of active and passive social media use with well-being: A critical scoping review. *New media & society* 24, 2 (2022), 530–549.

[92] Scott R. Vrana and David Rollock. 2002. The role of ethnicity, gender, emotional content, and contextual differences in physiological, expressive, and self-reported emotional responses to imagery. *Cognition and Emotion* 16, 1 (Jan. 2002), 165–192. https://doi.org/10.1080/02699930143000185 Publisher: Routledge _eprint: https://doi.org/10.1080/02699930143000185.

[93] Greg Wadley, Wally Smith, Peter Koval, and James J. Gross. 2020. Digital Emotion Regulation. *Current Directions in Psychological Science* 29, 4 (2020), 412–418. https://doi.org/10.1177/0963721420920592 arXiv:https://doi.org/10.1177/0963721420920592

[94] Rafael Wampfler, Severin Klingler, Barbara Solenthaler, Victor Schinazi, and Markus Gross. 2019. Affective State Prediction in a Mobile Setting using Wearable Biometric Sensors and Stylus. In *Proceedings of the 12th International Conference on Educational Data Mining, EDM 2019, Montréal, Canada, July 2-5, 2019. International Educational Data Mining Society (IEDMS) 2019*. Université du Québec; Polytechnique Montréal, Montréal, Canada, 198–207. https://doi.org/10.3929/ethz-b-000393912

[95] Rafael Wampfler, Severin Klingler, Barbara Solenthaler, Victor R. Schinazi, and Markus Gross. 2020. Affective State Prediction Based on Semi-Supervised Learning from Smartphone Touch Data. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376504

[96] Rafael Wampfler, Severin Klingler, Barbara Solenthaler, Victor R. Schinazi, Markus Gross, and Christian Holz. 2022. Affective State Prediction from Smartphone Touch and Sensor Data in the Wild. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3491102.3501835

[97] Chunting Wan, Dongyi Chen, and Honghang Lv. 2022. Emotion Recognition Using the Fusion of Frontal 2-Channel EEG Signals and Peripheral Physiological Signals. In *Proceedings of the 12th International Conference on Biomedical Engineering and Technology* (Tokyo, Japan) *(ICBET '22)*. Association for Computing Machinery, New York, NY, USA, 69–74. https://doi.org/10.1145/3535694.3535707
[98] D. Watson, L. A. Clark, and A. Tellegen. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology* 54, 6 (June 1988), 1063–1070. https://doi.org/10.1037//0022-3514.54.6.1063
[99] Małgorzata Wierzba, Monika Riegel, Anna Pucz, Zuzanna Leśniewska, Wojciech Dragan, Mateusz Gola, Katarzyna Jednoróg, and Artur Marchewka. 2015. Erotic subset for the Nencki Affective Picture System (NAPS ERO): cross-sexual comparison study. *Frontiers in Psychology* 6 (Sept. 2015), 13. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.01336
[100] Tim Wu. 2017. *The attention merchants: The epic scramble to get inside our heads.* Vintage.
[101] Jingyuan Yang, Jie Li, Leida Li, Xiumei Wang, and Xinbo Gao. 2021. A circular-structured representation for visual emotion distribution learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4237–4246.
[102] Jufeng Yang, Ming Sun, and Xiaoxiao Sun. 2017. Learning visual sentiment distributions via augmented conditional probability neural network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31.
[103] Kangning Yang, Chaofan Wang, Yue Gu, Zhanna Sarsenbayeva, Benjamin Tag, Tilman Dingler, Greg Wadley, and Jorge Goncalves. 2023. Behavioral and Physiological Signals-Based Deep Multimodal Approach for Mobile Emotion Recognition. *IEEE Transactions on Affective Computing* 14, 2 (2023), 1082–1097. https://doi.org/10.1109/TAFFC.2021.3100868
[104] Zhong Yin, Mengyuan Zhao, Yongxiong Wang, Jingdong Yang, and Jianhua Zhang. 2017. Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. *Computer methods and programs in biomedicine* 140 (2017), 93–110.
[105] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. 2016. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE signal processing letters* 23, 10 (2016), 1499–1503.
[106] Lili Zhu, Petros Spachos, and Stefano Gregori. 2022. Multimodal Physiological Signals and Machine Learning for Stress Detection by Wearable Devices. In *2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. 1–6. https://doi.org/10.1109/MeMeA54994.2022.9856558
[107] I. Zualkernan, F. Aloul, S. Shapsough, A. Hesham, and Y. El-Khorzaty. 2017. Emotion recognition using mobile phones. *Computers & Electrical Engineering* 60 (May 2017), 1–13. https://doi.org/10.1016/j.compeleceng.2017.05.004
[108] Hedda Martina Šola, Mirta Mikac, and Ivana Rončević. 2022. Tracking unconscious response to visual stimuli to better understand a pattern of human behavior on a Facebook page. *Journal of Innovation & Knowledge* 7, 1 (Jan. 2022), 100166. https://doi.org/10.1016/j.jik.2022.100166

A CORRELATION OF FEATURES WITH VALENCE AND AROUSAL

We conducted a correlation analysis by computing Spearman's rank correlation coefficient between physical and behavioral features and participant's valence and arousal ratings. Table 5 show the fraction of physiological and behavioral features for each time period that correlate significantly with valence respectively arousal. In terms of correlations, no distinct temporal pattern is observable. Interestingly, the fraction of physiological features that correlate with Arousal ratings is higher than for valence ratings. The opposite is true for behavioral features where a higher fraction correlates with valence than with Arousal self-reports.

B EMOTION MODEL AND RESOLUTION

In Table 6, we summarize the results of our classifier analysis.

C TEMPORAL AND FEATURE TYPE DEPENDENCY ON EMOTIONAL CONTENT CHANGES

Table 7 summarizes our comparison of the classification accuracy for each resolution for different feature types, time windows, and resolutions. Results confirm that the highest state detection accuracy per resolution is always either achieved by a classifier trained on behavioral- or all features. In terms of temporal dependency, results show no clear pattern. However, none one the best performing classifier per resolution was trained on data from the first sub-time-window (0-8 sec).

Table 5. Fraction of physiological and behavioral features for each time period where a significant Spearman correlation between features and Valence- and Arousal ratings of a particular magnitude ($^{*}\ p < .05$, $^{**}\ p < .01$) was identified.

Dim.	Type	Group	0 - 8 sec		4 - 12 sec		8 - 16 sec		12 - 20 sec		0 - 20 sec	
			*	**	*	**	*	**	*	**	*	**
Valence	physiol.	ECG	-	-	-	-	-	-	0.05	-	-	-
		PPG	0.10	0.10	0.24	0.05	0.19	0.14	0.24	0.14	0.33	0.24
		EDA	-	-	-	-	0.38	0.08	0.19	0.08	-	-
		RSP	0.06	-	-	-	-	-	-	-	-	-
		All	**0.03**	**0.02**	**0.05**	**0.01**	**0.11**	**0.04**	**0.10**	**0.04**	**0.07**	**0.05**
	behavioral	Phone interactions	0.40	0.40	0.90	0.50	0.50	0.20	0.30	0.20	0.90	0.90
		Eye tracking	0.67	0.33	0.67	0.67	0.67	0.33	0.67	0.67	0.67	0.67
		Facial expressions	0.14	-	0.57	0.29	0.71	0.57	0.57	0.57	0.43	-
		Motion	0.06	-	0.06	-	0.06	-	0.06	-	0.19	0.06
		Touch	0.43	0.22	0.46	0.26	0.51	0.23	0.43	0.17	0.56	0.32
		All	**0.34**	**0.19**	**0.53**	**0.34**	**0.49**	**0.27**	**0.41**	**0.32**	**0.55**	**0.39**
Arousal	physiol.	ECG	0.29	0.05	0.57	0.38	0.52	0.29	0.43	0.19	0.48	0.24
		PPG	0.14	0.05	0.43	0.33	0.33	0.29	0.48	0.29	0.33	0.24
		EDA	-	-	0.35	0.12	0.31	0.12	0.08	0.04	0.04	-
		RSP	-	-	0.06	-	0.11	-	0.06	-	-	-
		All	**0.21**	**0.10**	**0.44**	**0.21**	**0.37**	**0.22**	**0.33**	**0.18**	**0.33**	**0.14**
	behavioral	Phone interactions	-	-	0.20	0.20	0.70	0.60	0.30	0.20	0.60	0.20
		Eye tracking	1.00	0.67	1.00	0.67	0.67	0.33	0.67	0.33	1.00	0.67
		Facial expressions	0.43	0.14	0.14	0.14	0.14	-	0.29	0.29	0.29	0.14
		Motion	-	-	0.06	-	0.06	-	-	-	-	-
		Touch	0.01	-	0.17	0.06	0.10	0.04	0.01	-	0.10	0.06
		All	**0.29**	**0.16**	**0.31**	**0.21**	**0.33**	**0.19**	**0.25**	**0.16**	**0.40**	**0.21**

Table 6. Accuracy of emotional state detection over resolutions of emotion models. The table shows the best (Max) and mean (M) accuracy and its standard deviation (SD) of each classifier over all time windows. The best performing classifier for each emotion model and resolution is highlighted in bold.

Model	**Res.**	**ZeroR**	**SVM**	**RF**	**MLP**	**AB**	**LR**	**XGB**
		Acc.	**Max** **M±SD**	**Max** **M±SD**	**Max** **M±SD**	**Max** **M±SD**	**Max** **M±SD**	**Max** **M±SD**
Valence	**low**	0.28	0.68 0.65±0.02	0.67 0.67±0.01	0.65 0.64±0.01	0.65 0.63±0.02	0.64 0.63±0.01	**0.70** **0.66±0.03**
	med.	0.28	**0.46** **0.40±0.05**	0.43 0.40±0.03	0.43 0.40±0.04	0.44 0.38±0.04	0.45 0.40±0.05	0.45 0.41±0.02
	high	0.1	**0.29** **0.26±0.02**	**0.29** **0.26±0.02**	0.26 0.24±0.02	**0.29** **0.26±0.02**	0.27 0.25±0.02	**0.29** **0.27±0.01**
Arousal	**low**	0.28	0.68 0.65±0.02	0.69 0.67±0.02	0.67 0.65±0.02	0.65 0.63±0.02	0.64 0.63±0.01	**0.70** **0.66±0.03**
	med.	0.28	**0.46** **0.40±0.05**	0.43 0.41±0.02	0.43 0.40±0.02	0.44 0.38±0.04	0.45 0.40±0.05	0.45 0.41±0.02
	high	0.1	0.29 0.26±0.02	0.27 0.26±0.01	**0.29** **0.25±0.03**	**0.29** **0.26±0.02**	0.27 0.25±0.02	**0.29** **0.27±0.01**
Mikel	**low**	0.26	0.69 0.65±0.06	**0.72** **0.67±0.03**	0.64 0.59±0.04	0.65 0.60±0.04	0.64 0.58±0.05	0.71 0.65±0.05
	med.	0.14	**0.53** **0.49±0.03**	**0.53** **0.51±0.02**	0.46 0.41±0.05	0.58 0.48±0.09	0.45 0.39±0.06	0.47 0.44±0.02
	high	0.08	0.37 0.35±0.01	**0.51** **0.49±0.02**	0.30 0.28±0.02	0.15 0.10±0.04	0.31 0.27±0.03	0.43 0.40±0.02

Table 7. Accuracy of feature types over different time windows and model-resolution pairs (*Valence-low*, *Arousal-low*, *Valence-medium*, *Arousal-medium*, and *Mikel-high*). The table shows the accuracy of the ZeroR baseline (ZeroR) and the best performing classifier (Max; low=XGB, medium&high=RF). The time window in which the best-performing classifier was trained is indicated in bold. If it is also the best-performing classifier for a given resolution, it is additionally marked in italics.

Res.	Model	Typ.	ZR	0-8 sec		4-12 sec		8-16 sec		12-20 sec		0-20 sec	
				M SD	Max Clf	M SD	Max Clf	M SD	Max Clf	M SD	Max Clf	M SD	Max Clf
low	**Valence**	**all**	0.28	0.63 ±0.0	0.66 RF	0.65 ±0.0	0.67 RF	0.64 ±0.0	0.67 SVM	0.65 ±0.0	***0.70*** ***XGB***	0.65 ±0.0	0.68 SVM
		phy.	0.28	0.56 ±0.1	0.65 RF	0.58 ±0.0	0.63 RF	0.59 ±0.1	**0.67** **RF**	0.60 ±0.0	0.66 XGB	0.59 ±0.0	0.66 RF
		beh.	0.28	0.64 ±0.0	0.66 XGB	0.64 ±0.0	**0.69** **RF**	0.64 ±0.0	0.66 RF	0.64 ±0.0	0.67 XGB	0.66 ±0.0	**0.69** **SVM**
	Arousal	**all**	0.28	0.62 ±0.0	0.64 RF	0.65 ±0.0	0.67 XGB	0.65 ±0.0	0.69 RF	0.66 ±0.0	***0.70*** ***XGB***	0.66 ±0.0	0.69 RF
		phy.	0.28	0.55 ±0.1	0.63 RF	0.57 ±0.0	0.63 RF	0.58 ±0.1	0.65 RF	0.60 ±0.0	**0.66** **RF**	0.59 ±0.0	**0.66** **RF**
		beh.	0.28	0.64 ±0.0	0.66 XGB	0.64 ±0.0	0.67 RF	0.63 ±0.0	0.66 RF	0.64 ±0.0	0.67 XGB	0.66 ±0.0	**0.69** **SVM**
med.	**Valence**	**all**	0.28	0.34 ±0.0	0.39 XGB	0.40 ±0.0	0.42 RF	0.43 ±0.0	**0.46** **SVM**	0.42 ±0.0	0.45 LR	0.41 ±0.0	0.44 SVM
		phy.	0.28	0.33 ±0.0	0.36 XGB	0.36 ±0.0	0.38 SVM	0.36 ±0.0	0.38 RF	0.38 ±0.0	**0.41** **MLP**	0.35 ±0.0	0.37 RF
		beh.	0.28	0.39 ±0.0	0.42 XGB	0.41 ±0.0	0.44 XGB	0.41 ±0.0	0.44 XGB	0.42 ±0.0	***0.48*** ***RF***	0.44 ±0.0	0.46 SVM
	Arousal	**all**	0.28	0.35 ±0.0	0.39 XGB	0.40 ±0.0	0.42 RF	0.42 ±0.0	***0.46*** ***SVM***	0.42 ±0.0	0.45 LR	0.41 ±0.0	0.44 SVM
		phy.	0.28	0.33 ±0.0	0.36 XGB	0.36 ±0.0	0.38 SVM	0.36 ±0.0	0.39 RF	0.37 ±0.0	**0.40** **LR**	0.34 ±0.0	0.37 XGB
		beh.	0.28	0.39 ±0.0	0.42 XGB	0.41 ±0.0	0.44 XGB	0.41 ±0.0	0.44 XGB	0.41 ±0.0	0.44 XGB	0.44 ±0.0	***0.46*** ***SVM***
high	**Mikel**	**all**	0.08	0.30 ±0.2	***0.51*** ***RF***	0.32 ±0.1	0.47 RF	0.32 ±0.1	0.48 RF	0.33 ±0.1	0.49 RF	0.32 ±0.1	***0.51*** ***RF***
		phy.	0.08	0.27 ±0.1	0.45 RF	0.25 ±0.1	0.45 RF	0.29 ±0.1	**0.50** **RF**	0.29 ±0.1	0.45 RF	0.29 ±0.1	0.48 RF
		beh.	0.08	0.28 ±0.1	0.50 RF	0.29 ±0.1	0.47 RF	0.28 ±0.1	0.47 RF	0.26 ±0.1	0.47 RF	0.28 ±0.1	***0.51*** ***RF***