Emotions on the Go: Mobile Emotion Assessment in Real-Time using Facial Expressions

Thomas Kosch
LMU Munich
Munich, Germany
thomas.kosch@ifi.lmu.de

Mariam Hassib
Bundeswehr University Munich
Munich, Germany
mariam.hassib@unibw.de

Robin Reutter
University of Stuttgart
Stuttgart, Germany
robin.reutter@gmail.com

Florian Alt
Bundeswehr University Munich
Munich, Germany
florian.alt@unibw.de

ABSTRACT

Exploiting emotions for user interface evaluation became an increasingly important research objective in Human-Computer Interaction. Emotions are usually assessed through surveys that do not allow information to be collected in real-time. In our work, we suggest the use of smartphones for mobile emotion assessment. We use the front-facing smartphone camera as a tool for emotion detection based on facial expressions. Such information can be used to reflect on emotional states or provide emotion-aware user interface adaptation. We collected facial expressions along with app usage data in a two-week field study consisting of a one-week *training phase* and a one-week *testing phase*. We built and evaluated a person-dependent classifier, yielding an average classification improvement of 33% compared to classifying facial expressions only. Furthermore, we correlate the estimated emotions with concurrent app usage to draw insights into changes in mood. Our work is complemented by a discussion of the feasibility of probing emotions on-the-go and potential use cases for future emotion-aware applications.

CCS CONCEPTS

• **Human-centered computing → Smartphones**; *Mobile devices*; *Ubiquitous and mobile computing design and evaluation methods.*

KEYWORDS

Mobile Sensing; Emotion Recognition; Affective Computing; Emotion-Aware Interfaces

ACM Reference Format:
Thomas Kosch, Mariam Hassib, Robin Reutter, and Florian Alt. 2020. Emotions on the Go: Mobile Emotion Assessment in Real-Time using Facial Expressions. In *International Conference on Advanced Visual Interfaces (AVI '20), September 28-October 2, 2020, Salerno, Italy*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3399715.3399928

AVI '20, September 28-October 2, 2020, Salerno, Italy

ACM ISBN 978-1-4503-7535-1/20/09…$15.00
https://doi.org/10.1145/3399715.3399928

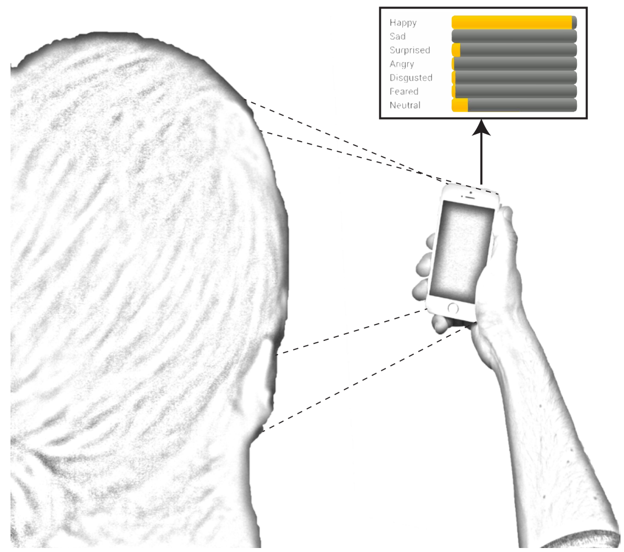
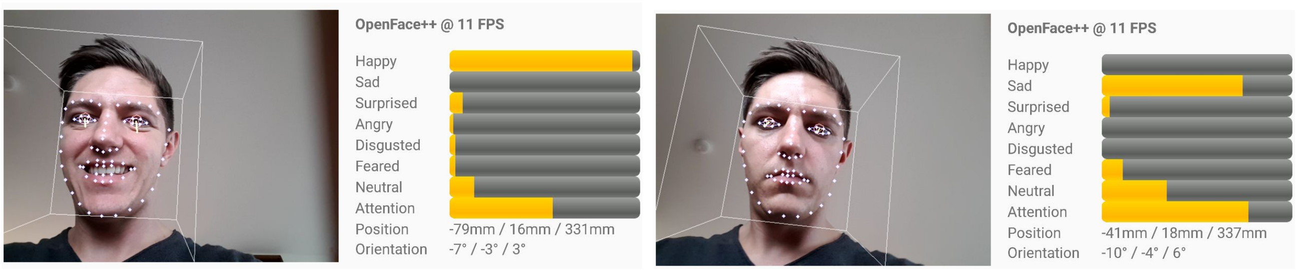


Figure 1: Capturing facial expressions by using the smartphone front camera. Emotions can be accurately predicted through the collection of contextual and facial data.

1 INTRODUCTION

People use their smartphone around 220 times per day in public and private environments [20]. Recently, the role of the smartphone for probing emotions has been investigated from different perspectives, as researchers looked at the effects of usage on mood [34, 35, 55]. Especially the analysis of mood and emotional states from smartphone usage patterns [10, 44] was subject to previous research. Affective computing [42] applications go beyond the desktop and extend smartphones which have recently been explored for tapping into the users' emotional states. We see emotion recognition on the brink of expanding into ubiquitous environments.

At the same time, improved sensing capabilities on smartphones make it possible to obtain detailed insights on micro-interactions humans perform in everyday life. This includes physical activity, behavioral patterns of interaction with the phone, app usage, location, or sound levels. Such parameters can be correlated with perceived



(a) (b)

Figure 2: Estimated emotions through facial expressions by OpenFace. The likelihood of the current emotion is determined by evaluating facial patterns. For example, (a) happy persons are evaluated by detecting the high corners of the mouth, while (b) low corners are a cue for sadness.

emotions and draw accurate predictions about the current mood. Together with the aforementioned ability to assess users' emotions, we envision the smartphone as an emotion-aware companion. Affect recognition sensors provide a holistic view of the users' emotions in an unobtrusive and automated way. For example, emotions can be recorded to enable users to reflect on their mood. Moreover, apps can become emotion-aware, where contextual suggestions can be provided depending on the current mood.

In our work, we utilize the frontal camera and the recording of the current app use to gain insights into emotional states elicited by smartphone apps (see Figure 1). We developed a mobile-based affect recognition application that analyzes facial expressions from the frontal camera stream to detect six basic emotions [13, 15] (see Figure 2). In a field study (N=12), we collected a corpus of subjective mood information using experience sampling as well as facial expressions along with app usage behavior. We used the facial expression data with app usage information to train user-dependent classifiers, achieving an overall improvement of 33% in classification accuracy compared to evaluating facial expressions without app usage data. In the *testing phase* of the study, we tested the classifiers with eight out of the original twelve users in their daily lives and conducted interviews. We analyzed the collected data and together with the qualitative findings, we discuss the feasibility of on-the-go emotion detection and potential use cases.

CONTRIBUTION STATEMENT

Our contribution is threefold: We (1) present a mobile-based emotion detection system based on the open-source OpenFace platform [3]. We (2) perform a two-week field study consisting of a one-week *training phase* for an emotion predicting classifier that takes facial expressions and current app usage into account. The classifier is evaluated in a one-week *testing phase* to prove its efficiency. Finally, we (3) discuss how emotion-aware mobile applications can be built by temporally probing emotional states.

2 RELATED WORK

Recognizing and inferring affective states was subject to prior research. We provide an overview of (1) facial expression emotion recognition and (2) currently available affective systems.

Figure 3: Six basic emotions defined by Ekman et al. [11, 16]. A modified version of OpenFace [3] uses the front camera to detect landmarks on faces as indicator for emotions.

2.1 Assessing Emotions by Facial Expressions

Deriving the mood of users utilizing facial expression recognition is a compelling topic that has been addressed by previous research. Since the interpretation of facial expressions depends on the gender [14] and is different across cultures [47], emotion recognition is challenging for researchers. Ekman et al. [11, 16] state that a connection between emotions and facial expressions exists. They recorded facial responses while enforcing different emotions (see Figure 3). Picard et al. [43] state, that this kind of emotion detection is crucial regarding emotional intelligence, and thus, describes emotion-sensing as a component that has to be included in machines. However, the underlying trained model uses facial expressions that were expressed intentionally to improve the classification accuracy, resulting in unnatural expressions for perceived emotions. Hess et al. [25] found that aside from underlying emotions, the social setting has a strong impact on facial expressions. The sociality of the situation, the relationship between expresser and audience, and the intensity of stimuli all affect facial expressions. Recently, researchers started investigating mobile facial expression recognition platforms. Filho et al. [19] created and tested their prototypical mobile phone application recognizing smiles via the front camera in a preliminary study. For the face and emotion recognition, they used Orbeus ReKognition API[1]. Other platforms, such as OpenFace [3], Fraunhofer SHORE [46], or the AFFDEX SDK [39] provide ready to use frameworks for recognizing six basic emotions defined by Ekman and Oster [15] using RGB data only. Advances in machine learning have also largely benefited the area of emotion recognition through facial expressions. Neural networks improved the classification of emotions for individuals [9, 21] or were trained to detect single emotions with high accuracy such as pain [45].

2.2 Affect-Aware Systems

Affect-aware systems can be split into awareness and adaptive systems. Awareness systems detect emotional states and either display them back to the user in some form [38] or share the state with other people. The main focus of these systems is to provide feedback reasonably. Adaptive systems, on the other hand, use emotional user states as system input and introduce adaptive changes depending on the state. The rise of ubiquitous sensing technologies enabled the development of affect-aware systems. Researchers explored using emotional states to control games [40], in education [4, 54], or to create seamless communication channels to exchange emotions with others [5, 6, 19]. Ängeslevä et al. [1] used facial expression recognition while writing emails to communicate contextual cues. Recent emotion-aware systems explored the pairing of sensing modalities. For example, McDuff et al. [38] developed an emotion-aware desktop-based system that collects emotion information through sensory data. Besides detecting emotions, a personal dashboard visualized the evaluation in correlation to the user's context which enhanced emotional memory and reflection [38]. El Kaliouby et al. [18] explored the idea of an emotional prosthetic which gives continuous feedback about emotions and helps in reflection and interpretation. Stahl et al. [51] created an affective diary for reflection from data collected from the user's mobile phones, body-worn sensors, and hand-written entries. Researchers also explored the use of smartphones as emotional probes and means of sharing expressed emotions with others. Olsen and Torresen [41] analyze accelerometer data to derive currently experienced emotions. Furthermore, Shapsough et al. [50] analyze emotions through the user's typing behavior in a mobile context. Church et al. [5] designed a proof-of-concept mobile app for in-situ sharing of emotional states between friends to increase awareness and enhance communication [5]. Cui et al. [6] created a mobile system which records frontal camera video reactions to photos being exchanged between users. They found that these emotional video responses support lightweight phatic social interactions. Other researchers investigated the augmentation of communication apps on smartphones or the desktop with emotions sensed explicitly or implicitly using wearable sensors[17, 22, 33]. Furthermore, researchers were concerned with mapping basic emotions into vector space representations [36]. This enables deeper analysis and comparison of emotions among individuals. Recently, methods for collecting experience samples in the wild were presented [24]. Adaptive affective systems use the detected emotions as input for system adaptation. Maat and Pantic [37] created *GazeX*, a multimodal affect sensitive adaptive system which uses facial expressions, eye gaze, and desktop logging to adapt the system and support users in the workplace. Dingler et al. [10] built an application that detects moments of boredom during smartphone usage to help users learn languages. Schrader et al. [49] introduced approaches for integrating emotion in developing game dynamics. In human-robot interaction, Tielman et al. [52] investigated a setup in which a robotic agent gave adaptive emotional feedback to children in different situations based on their emotional state.

Detecting emotions as a basis for adaptive user interfaces or to analyze human behavior has been extensively researched. However, means for detecting emotional states in mobile settings by analyzing facial expressions and app usage only has received little attention. We close this gap by presenting and evaluating our concept of individual emotion recognition using the smartphone front camera and app usage data.

3 FIELD STUDY

We conducted a field study to assess the feasibility of probing emotions in-the-wild through the frontal camera and app usage. We collected emotions as extracted by a version of OpenFace that runs on Android. To collect data for classification and evaluate our approach, we divided the study into two phases. In the *training phase*, facial and app usage data from participants are collected to train a person-dependent classifier. We have chosen a person-dependent approach since facial expressions and experienced emotions are different among individual persons [12]. The *testing phase* comprises the installation of the person-dependent classifier to evaluate its emotion prediction performance. The duration of each phase is one week and the total duration of the study is two weeks (see Figure 4).

3.1 Apparatus

We used the facial expression recognition platform, OpenFace[2] [3] from Baltrusaitis et al. as an emotion detection library for Android. OpenFace is publicly available and enables efficient facial expression evaluation on mobile platforms using the front camera only. OpenFace is able to predict the probability of the six basic emotions, as defined by Ekman and Oster [15], by analyzing facial expressions. These are *happiness*, *sadness*, *surprise*, *fear*, *disgust*, and *anger*.

We first adapted the platform to work on specific Android devices with version 5.0[3] and up. The application streams data from the frontal camera at a rate of ten frames per second to save computing power and reduce battery consumption. Data is only collected when the display of the smartphone was turned on and the face of the user is detected. Thereby, the frontal camera was turned off when the screen was turned off. Facial expressions were only collected when the user's face was turned towards the smartphone screen. OpenFace calculates an attention score, a metric that correlates with landmarks and eye contact between user and smartphone display, which is used to detect if a user is looking towards the

[1]Has been replaced by Amazon Rekognition - www.aws.amazon.com/rekognition - last access 2020-05-31

[2]www.github.com/TadasBaltrusaitis/OpenFace - last access 2020-05-31

[3]Android Lollipop

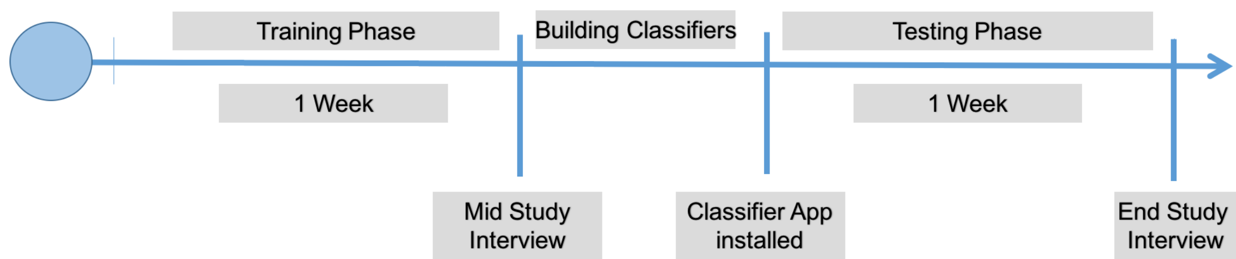


Figure 4: Timeline of the whole experiment. The *training phase* comprised the collection of training data and experience samples. A person-dependent classifier is built afterwards, which is evaluated in a *testing phase*.

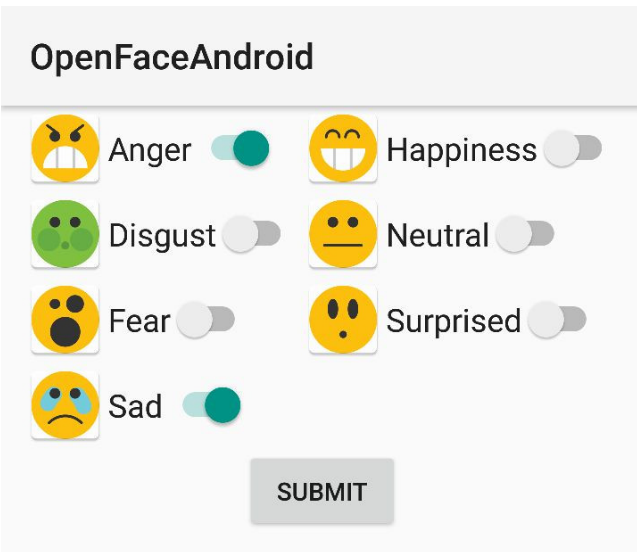


Figure 5: Experience sampling probe about the users' current mood. Participant made a binary selection of multiple emotions which were close to their current mood.

smartphone. With this, we avoid recordings were the users' face is partially detected but does not pay attention to the displayed content. Figures 2a and 2b show facial expressions detected by the mobile application such as happiness or sadness with an attention score over 50%. In addition to the collected emotion predictions from OpenFace, the current app usage was collected. All data is logged locally and no images are stored or sent to external servers. The app runs as a background service to prevent user interventions, such as manual shutdowns or reboots. To collect ground truth about the currently experienced emotions, participants receive an experience sampling probe notification every 15 minutes. The notification asks them to rate their current emotion for each of the six basic emotions in addition to a default emotion *neutral* (see Figure 5). All facial data which was gathered within the *current* uptime of the smartphone screen was labeled with the rated emotion. However, as smartphones may be used for a long duration at a time, we limited the labeling of facial data to the *first* 15 minutes after answering the experience sample questionnaire. With this, we intended to avoid incorrect labeling due to changes in emotions. Samples that were collected during the experience sampling probe were discarded as we did not want to include facial expressions *while* answering the experience sampling probe. Furthermore, the experience sampling probe instructed participants to rate their experienced emotions for apps they have used recently excluding the experience sampling probe.

3.2 Participants and Procedure

We recruited twelve participants to take part in the field study (six female, 21–45 years, M = 26.92, SD = 5.94). All participants had smartphones running with at least Android 5.0. Participants were recruited through acquaintances, colleagues, and university mailing lists. They received 20 Euro for participating. The overall duration of the study was two weeks. In the following, we outline the procedure of the *training phase* and *testing phase*.

3.2.1 Training Phase. We explained the purpose of our study and the data being logged. Furthermore, we explained that all data is recorded locally and is not stored remotely. We also explained that participation is voluntary and that they can abort the study at any time. Participants then signed an informed consent form and provided demographic data. Afterward, the application was downloaded to their phones. We explained how the Android app is used and we showed the participants how to fill in the experience sampling questionnaire. Afterward, we showed participants where the logs are saved on their phones and asked them to return to our lab again after one week to collect the gathered log files and conduct short semi-structured interviews to conclude phase 1 of the study (*training phase*, see Figure 4). This included questions about their experiences with the app and how much they think it affected the overall functionality of their smartphone. Two participants aborted the study during the *training phase*.

3.2.2 Testing Phase. Out of the initial twelve participants from the *training phase*, eight participants took part in the second week of the study phase which is the *testing phase* (four female, 21–45 years, M = 26, SD = 3.08). We asked participants again to come to our lab and to reinstall the app on their smartphones. However, this time the app included their person-dependent classifier model. Four participants dropped out for the testing phase due to a low number of experience samples (N = 3) or personal absence (N = 1).

Participant	Emotion Samples	Experience Samples	Number of Instances	F1 Score	F1 Score with App Usage
1	10,774	89	10,735	.94	.97
2	21,154	280	20,909	.68	.77
3	6,936	37	6,905	.70	.79
4	14,249	46	14,049	.56	.72
5	2,777	12	2,777	.66	.82
6	25,778	107	25,777	.57	.69
7	2,838	23	2,838	.81	.88
8	3,430	42	3,425	.75	.83
9	6,249	94	6,079	.61	.70
10	4,717	88	4,714	.58	.71
11	3,368	19	3,368	.71	.82
12	5,459	27	4,775	.83	.87

Table 1: Emotion samples, answered experience samples, and F1 scores of a cross-validation (k = 10) per participant collected during the *training phase*. We removed participants P5, P7, and P11 from the *testing phase* due to a low number of samples.

Before starting the *testing phase*, we explained the new app to the participants and showed them that they will be prompted with the same experience sampling method as in the previous study. However, this time we do not use a fixed time window of 15 minutes to ask the participants about their current mood. Instead, we trigger the experience questionnaire when 25 frames of the same classified emotion were counted regardless of the time frame. The overall duration of the *testing phase* was one week. Afterward, we met with the participants in our lab to conduct semi-structured interviews about events they experienced in the last week which could have affected their emotional perception and overall experience with the app. Overall, we collect the user experience sampling probes, emotion samples as assessed by OpenFace, and emotion prediction outcome from our trained classifier which used the OpenFace emotion labels and currently used application information as features.

3.3 Training Phase: Emotion Classification

Participants who aborted the training phase were removed from the *training phase* analysis. Hence, data of twelve participants were used for further analysis. In total, we collected 864 experience sampling questionnaires and 107,405 facial expression samples through OpenFace. We used the collected data to train person-dependent classifiers. using facial expressions and app usage as features.

3.3.1 Data Preprocessing. We performed several steps to clean the data before training a classifier. To label the data points with ground truth emotion labels, we used the same defined time frame of when the screen was turned on until the experience sampling took place or full 15 minutes of facial data if the smartphone was used for a longer duration. For the currently used application, we labeled the data with the current open Android activity. No data was collected when the smartphone screen was locked. With this, we avoided the pollution of data during short smartphone interaction time frames that are not related to the currently perceived emotions of our participants.

3.3.2 Evaluating Emotion Prediction. As emotions are highly person-dependent [8, 12], and the fact that our system is intended to run on users' phones, we chose to train person-dependent classifiers for emotion recognition from facial expressions and logged data. The features used to train the classifier were the measured facial expressions for a maximum of 15 minutes after every experience-sampling probe and the currently used app on the smartphone. Data were only collected when the smartphone screen was unlocked and an experience sample questionnaire was filled. Interrupting the data collection, by for example locking the smartphone screen, quit the current data collection session until answering the next experience sampling questionnaire. We use the six basic rated emotions and *neutral* as class labels. The experience sampled emotions served as ground truth. For each participant, we evaluated the performance of two classifiers: one that included only the facial expressions retrieved from OpenFace as a feature, and one that combined the two features: facial expressions and used application. We use random forests to train the person-dependent classifiers since random forests are computationally more efficient compared to other training methods such as deep neural networks. Furthermore, random forests can be easily applied on mobile platforms without the need for high computational demand or a network connection to exchange personal data with a server that calculates a model. Our model results in instances with seven features, which are the probability of all six basic emotions and neutral as determined by OpenFace, and instances with eight dimensions using the current app usage in addition to the seven emotional metrics raised by OpenFace. Unlabeled data was not used for training. Therefore, a different number of instances was defined per participant. We conducted a person-dependent cross-validation (k = 10). Thus, the data was randomly partitioned into ten folds. Nine folds were used for training, while the remaining fold was used for evaluation. The number of iterations was set to 100 and the minimum variance for splits was adjusted to 0.001. This procedure was repeated ten times per participant and accuracies were averaged. Additionally, we calculated the averaged F1 score from each fold. The average F1 score was .7 (SD = .1) when considering facial expressions only and .79 (SD = .08) when training facial expressions in combination with app usage. Thus, higher F1 scores are found when considering app usage as an additional feature. Table 1 provides an overview of F1 scores with and without app usage per participant.

Emotion	User Rating	Facial Expression	Facial Expression Classification
Anger	10,267	41,731	11,801
Disgust	3,713	4,670	5,772
Fear	937	9,692	932
Happiness	98,720	73,607	104,966
Neutral	327,118	150,323	254,478
Sad	31,433	94,913	29,500
Surprised	4,706	34,585	2,072

Table 2: Number of classified emotions per sample. We compare the collected samples of user ratings, facial expressions, and facial expressions using classification with app usage. The number of classified emotions taking app usage into account approximates the number of samples by user ratings.

3.3.3 Limitations. The *training phase* only lasted for one week. While that is suitable to detect a variety of emotions, data collected over a longer period may increase accuracy as well as allow a broader set of emotions to be considered – in particular, such that occur rather rarely and for short amounts of time. To show the feasibility of mobile emotion classification, we started with a small sample size to control the effects of causalities. Furthermore, we have only considered emotional probes at specific time intervals instead of a continuous data stream. Also, participants might have been aware of the study intention and provide supportive experience sampling probes to support the results.

3.4 Testing Phase: Evaluation in the Wild

We used the same app from the *training phase* but encapsulated the machine learning model for the individual eight participants who continued in the *testing phase*. The model continuously evaluates facial expressions in combination with currently used application data. For the *testing phase*, we tested (1) if our classifier yielded comparable results to those of the *training phase*, and hence test the feasibility of continuous emotion detection over a period of time with the same trained classifier and (2) to again compare the outcomes of using only facial expressions versus adding contextual information (e.g., app usage) to classify emotions in the wild. We cleaned the data using the same processing pipeline as in the *training phase*. We then analyzed the number of recorded facial expressions data samples to compare the user ratings with the outcomes of facial expressions only from OpenFace and the overall classification results with the application used as an additional feature. We compared the user ratings with facial expression only and the trained classifier. We find that facial expressions including app usage show approximately the user rating compared to using facial expression only (see Figure 6 and Table 2). We analyzed the user ratings and estimated the facial expressions per participant to investigate the overall performance between classifier and using facial expressions only. Comparing the user ratings of the *testing phase* with the facial expressions only and classification of facial expression together with app usage results in an improved prediction of emotions. Averaging the classification results in a prediction accuracy of 33% (SD = 20.91) for using facial expression only and 66% (SD = 20.91) when combining facial expressions with current app usage, hence resulting in an improvement of 33%.

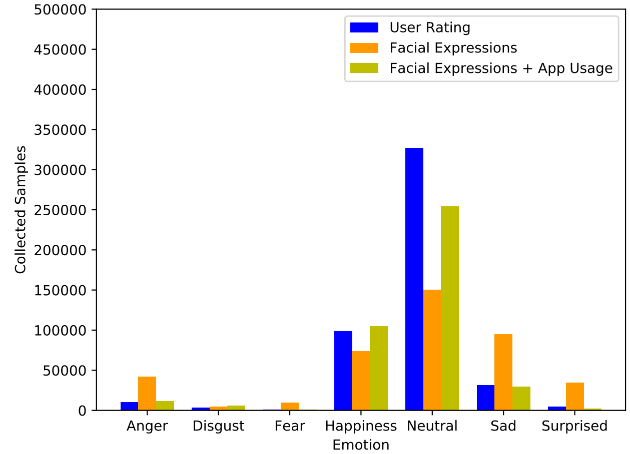


Figure 6: The number of classified emotions taking app usage into account approximates the number of samples by user ratings. These results are representative for the states *Happiness*, *Sad*, and *Neutral* due to a sufficient sample size.

3.5 Emotions and App Usage

We conclude that including contextual information such as app usage can lead to better emotion predictions. Recent research showed that emotions and specific apps, such as social media apps, can alter emotions into sustainable changes [7, 48]. This can be caused by a reward feedback loop, where the human brain recognizes social media rewards (e.g., likes or shares), as a rewarding event. This can train and alter the typical brain behavior [48].

To find which apps cause specific emotions, we grouped the logged apps into categories, comprising *Games*, *Messaging*, *Social Media*, and *Browser* applications. *Games* describe all applications which required participants to be interactively involved in an objective. *Messaging* contains all apps which enable direct or group communications with specific known persons. *Social Media* comprises all apps which enable personal content to be shared with a larger group of people to which they can respond (e.g., Facebook, Instagram, LinkedIn). Finally, *Browser* apps describe apps running in a browser, where generic content is displayed. Table 3 shows an overview of user ratings when using a specific app from one of the previously mentioned app categories. *Neutral* was rated very frequently. *Happiness* and *sadness* were rated more often than *anger*, *disgust*, *fear*, or *surprise*. Generally, people perceived *happiness* or *sadness* frequently when using *messaging* or *social media* apps.

3.6 Qualitative Feedback

We present the qualitative feedback collected from the one-hour semi-structured interviews conducted after the *training phase* and *testing phase*. We asked participants several questions related to their general behavior with the phone, privacy perceptions, and suggested use cases for emotion-awareness on smartphones. All participants stated they did not have major privacy concerns since they knew that the actual images were not saved either locally on the smartphone nor on an external server. Two participants stated that they did cover the camera intentionally or did not take the phone with them in sensitive contexts such as going to the

Category	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Unique Ratings	26	6	5	259	787	55	6
Browser	2	0	2	55	170	9	4
Games	0	0	2	28	150	6	7
Messaging	21	5	5	183	508	39	6
Social Media	12	8	1	95	350	45	1

Table 3: Emotions rated by participants per app category. The *unique ratings* represent cases in which participants provided a vote for a single emotion in the questionnaire within any of the app categories.

bathroom (P2, P6). Only three participants mentioned that they told the people around them, such as partners, siblings, and colleagues at work about the app. When we asked participants about their perceived emotions in relation to the apps they used every day, we got varying feedback which we compared to the outcomes from our classifiers. Two participants stated that they felt happy when scrolling through different social media platforms such as Facebook, Instagram, and YouTube (P1, P2, P4, P5), where they experienced a whole lot of emotions while using communication apps such as WhatsApp. Participants stated that the experience sampling probes often recorded their "reaction" to a certain application of activity rather than the current overall mood they are in (P3, P5). They also stated that the experience sampling made them really think of their current emotions and try to correlate it with various contexts (P3).

4 DISCUSSION

We discuss our findings from the study in terms of feasibility when it comes to detecting emotions in mobile settings. While considering contextual parameters, such as app usage, we train a classifier and show that the classified sample points approximate the user rated sample points compared to evaluating pure facial analysis.

4.1 Contextual Emotion Classification

Much of the research in emotion classification using facial expressions and partly in combination with other sensors was done in lab setups [28]. Few research projects attempted to further push the limits of emotion recognition by assessing its feasibility in natural and uncontrolled setups [39, 53]. The effects of other surrounding conditions that are present in the natural environment of ubiquitous computing usage need to be assessed as a step towards emotion-aware systems, providing users with an added value in their lives. Through our work, we found that it is feasible to collect meaningful facial expression information and contextual data from the mobile phone in an in-the-wild setting which enabled emotion detection on mobile phones. Compared to using facial expressions only, including app usage yielded a higher accuracy for most participants. Facial recognition in combination with current app usage shows that more sample points are assigned to emotion instead of *neutral*. We imply that the collected samples approximate the user rated emotions. We believe that the classification towards an emotion rather than *neutral* can be improved with more training data. We do not claim to predict the current emotion a user perceives *while* using a certain app. However, the questions that arise from these findings is whether users open up a certain application *when* they are in a certain state, or the users feel a certain emotion while they are using a particular application. Testing this hypothesis in future work would bring us a step closer to ubiquitous emotion recognition in the wild.

4.2 Feasibility of Mobile Emotion Recognition

Based on our results, real-time emotion recognition through facial expressions is feasible and provides interesting prospects for creating new affect-aware systems. Whereas camera-based facial detection still suffers from drawbacks due to occlusions and light conditions, through our studies we collected a large amount of data in a short timespan that still enabled us to obtain acceptable accuracies for detecting emotions. Also, recent studies showed that although users' faces are not always fully visible in mobile scenarios, a full face is visible around 30% of the time [26, 27]. Furthermore, we expect that these issues will be mitigated through the integration of depth sensors in smartphones.

Since emotions do not change frequently within short periods, obtaining accurate classifications at points where the user is in a context suitable for detection is feasible. However, we found that certain emotions can be scarcely identified through our application. These include *anger*, *disgust*, *fear*, and *surprise*. Our participants stated that they hardly felt these emotions during the study, hence we did not have enough data points for a thoughtful comparison. Determining the emotions mentioned above require often unusual facial movements to be recognized. Our results show that some emotions were perceived, but not detected by the app using facial evaluation and current app usage alone, such as *anger* and *surprise*. Although we were able to increase the emotion detection accuracy, a *training phase* is required. This can be achieved by evaluating emotions in the background during regular app usage while probing emotional states. If the confidence of the classifier reaches a threshold, apps can become emotion-aware and increase their accuracy by facilitating reinforcement learning.

4.3 Future Work

We plan to provide an enhanced version of our app on a software distribution platform to gather more data for analysis purposes. We want to provide a general classifier with apps that adapt their content based on the user's mood. For example, a game or newspaper app can change displayed content based on the currently measured mood. This can have a direct influence on the collected emotions and hence the classifier accuracy and can be avoided by building classifiers which are being continuously enhanced. In a future study, the outcome from the classifier can be compared directly to the subjective feedback from the user and, if it was a false classification, the new instance would be added and used to retrain the model. This ensures a continuous enhancement of the overall classifier accuracy. Since the 15-minute time window between the collection of ground truth labels poses a limitation of our work, we will leverage novel unobtrusive strategies to collect experience samples during smartphone usage [56]. We will further utilize other emerging physiological measures to extend the contextual space of emotion classification in-the-wild, such as eye tracking [29, 30] and brain-computer interfaces [32] for workload assessments as well as electrodermal activity for arousal measurements [2, 31]. Finally, we consider the use of deep neural networks to train facial expressions.

Cost-intensive training through deep learning can be calculated and downloaded remotely from a server. However, this has to meet privacy regulations as personal data regarding facial expressions and app usage would be uploaded to a foreign cloud service.

4.4 Use Cases for Mobile Emotion Detection

We present several use cases for inferring emotional states on mobile devices using the insights we collected from the user study.

4.4.1 Personal Ubiquitous Affective Diaries. The idea of having affective diaries for reflection has been explored and captured the interests of several researchers [18, 38, 51]. Through our studies, we found that using the frontal camera of the phone holds a lot of potential for creating an emotional prosthetic without augmenting the users with further sensors. This has also been mentioned during our interviews where users stated that a tool for self-reflection would be an interesting way for emotion regulation.

4.4.2 Content Adaptation. In the vision of a fully context-aware system, adaptivity is always a discussed topic. In the case of affective systems, as we charted in the background section, very few systems provided users with adaptive content based on their current status. Through our explorations, we find that such a system indeed would have an added value for users. For example, our users stated that filtering our content that is known to make them feel sad, or angry is a desired feature. An emotion-aware mobile phone can choose which apps a user could preferably use to elicit a certain mood.

4.4.3 Seamless Emotion Sharing Channel. Sharing emotions on the go has taken many forms in recent years – from text to emojis, videos, GIFs, and stickers, to the concept of "liking" on social media which has recently even advanced to a full list of possible emotional reactions to a post. This, coupled with findings from recent research [23], shows the interest in sharing emotional information with other users implicitly and explicitly. Several users stated that this can serve as a warning system to share emotions with others.

5 CONCLUSION

This paper presents a field study that uses a smartphone app to analyze facial expressions and current app use to build a classifier that predicts emotional states in mobile settings. In a *testing phase* of our study, we show the feasibility of our approach for certain emotions using a person-dependent classifier. Our results show a 33% improvement regarding emotion classification when considering the analysis of facial expressions combined with app use. We discuss potential use cases that arise from our work and the ability to use the phone as an emotion-aware companion. We believe, that the classification of facial expressions in combination with contextual data will lead to emotional prosthetics which enable users to reflect, adapt, and communicate emotions with mobile devices as well as with their relatives. We publish the source code of our application to foster research in this area[4].

ACKNOWLEDGMENTS

This research was supported by the Deutsche Forschungsgemeinschaft (DFG) under grant agreement no. 316457582 and 425869382.

[4]www.github.com/hcum/emotions-on-the-go - last access 2020-05-31

REFERENCES

[1] Jussi Ängeslevä, Carson Reynolds, and Sile O'Modhrain. 2004. EmoteMail. In *ACM SIGGRAPH 2004 Posters* (Los Angeles, California) *(SIGGRAPH '04)*. ACM, New York, NY, USA. https://doi.org/10.1145/1186415.1186426

[2] Yadid Ayzenberg, Javier Hernandez Rivera, and Rosalind Picard. 2012. FEEL: Frequent EDA and Event Logging – a Mobile Social Interaction Stress Monitoring System. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI EA '12)*. Association for Computing Machinery, New York, NY, USA, 2357–2362. https://doi.org/10.1145/2212776.2223802

[3] Tadas Baltrusaitis, Peter Robinson, and Louis-Philippe Morency. 2016. OpenFace: An open source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 1–10. https://doi.org/10.1109/WACV.2016.7477553

[4] Marian Stewart Bartlett, Gwen Littlewort, Ian Fasel, and Javier R. Movellan. 2003. Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction.. In *2003 Conference on Computer Vision and Pattern Recognition Workshop*, Vol. 5. 53–53. https://doi.org/10.1109/CVPRW.2003.10057

[5] Karen Church, Eve Hoggan, and Nuria Oliver. 2010. A Study of Mobile Mood Awareness and Communication Through MobiMood. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries* (Reykjavik, Iceland) *(NordiCHI '10)*. ACM, New York, NY, USA, 128–137. https://doi.org/10.1145/1868914.1868933

[6] Yanqing Cui, Jari Kangas, Jukka Holm, and Guido Grassel. 2013. Front-camera Video Recordings As Emotion Responses to Mobile Photos Shared Within Close-knit Groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. ACM, New York, NY, USA, 981–990. https://doi.org/10.1145/2470654.2466125

[7] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting Depression via Social Media. *ICWSM* 13 (2013), 1–10.

[8] Jozefien De Leersnyder, Michael Boiger, and Batja Mesquita. 2013. Cultural regulation of emotion: Individual, relational, and structural sources. *Frontiers in psychology* 4 (2013).

[9] Hui Ding, Shaohua K. Zhou, and Rama Chellappa. 2017. FaceNet2ExpNet: Regularizing a Deep Face Recognition Net for Expression Recognition. In *2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)*. 118–126. https://doi.org/10.1109/FG.2017.23

[10] Tilman Dingler. 2016. Cognition-aware Systems As Mobile Personal Assistants. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct* (Heidelberg, Germany) *(UbiComp '16)*. ACM, New York, NY, USA, 1035–1040. https://doi.org/10.1145/2968219.2968565

[11] Paul Ekman. 1984. Expression and the nature of emotion. *Approaches to emotion* 3 (1984), 19–344.

[12] Paul Ekman. 1993. Facial expression and emotion. *American psychologist* 48, 4 (1993), 384. https://doi.org/10.1037/0003-066X.48.4.384

[13] Paul Ekman. 1999. Facial expressions. *Handbook of cognition and emotion* 16 (1999), 301–320.

[14] Paul Ekman, Wallace V Friesen, Maureen O'sullivan, Anthony Chan, Irene Diacoyanni-Tarlatzis, Karl Heider, Rainer Krause, William Ayhan LeCompte, Tom Pitcairn, Pio E Ricci-Bitti, et al. 1987. Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of personality and social psychology* 53, 4 (1987), 712. https://doi.org/10.1037/0022-3514.53.4.712

[15] Paul Ekman and Harriet Oster. 1979. Facial expressions of emotion. *Annual review of psychology* 30, 1 (1979), 527–554.

[16] Paul Ekman and Erika L Rosenberg. 1997. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA.

[17] Rana El Kaliouby and Peter Robinson. 2004. FAIM: Integrating Automated Facial Affect Analysis in Instant Messaging. In *Proceedings of the 9th International Conference on Intelligent User Interfaces* (Funchal, Madeira, Portugal) *(IUI '04)*. ACM, New York, NY, USA, 244–246. https://doi.org/10.1145/964442.964493

[18] Rana el Kaliouby, Alea Teeters, and Rosalind W. Picard. 2006. An exploratory social-emotional prosthetic for autism spectrum disorders. In *International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06)*. 2 pp.–4. https://doi.org/10.1109/BSN.2006.34

[19] Jackson Feijó Filho, Thiago Valle, and Wilson Prata. 2014. Non-verbal communications in mobile text chat: emotion-enhanced mobile chat. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*. ACM, 443–446.

[20] Marian Harbach, Emanuel Von Zezschwitz, Andreas Fichtner, Alexander De Luca, and Matthew Smith. 2014. It's a hard lock life: A field study of smartphone (un) locking behavior and risk perception. In *Symposium on usable privacy and security (SOUPS)*. 9–11.

[21] Behzad Hasani and Mohammad H Mahoor. 2017. Spatio-Temporal Facial Expression Recognition Using Convolutional Neural Networks and Conditional Random Fields. *arXiv preprint arXiv:1703.06995* (2017).

[22] Mariam Hassib, Daniel Buschek, Paweł W. Wozniak, and Florian Alt. 2017. HeartChat: Heart Rate Augmented Mobile Chat to Support Empathy and Awareness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. ACM, New York, NY, USA, 2239–2251. https://doi.org/10.1145/3025453.3025758

[23] Mariam Hassib, Mohamed Khamis, Stefan Schneegass, Ali Sahami Shirazi, and Florian Alt. 2016. Investigating User Needs for Bio-sensing and Affective Wearables. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. ACM, New York, NY, USA, 1415–1422. https://doi.org/10.1145/2851581.2892480

[24] Javier Hernandez, Daniel McDuff, Christian Infante, Pattie Maes, Karen Quigley, and Rosalind Picard. 2016. Wearable ESM: Differences in the experience sampling method across wearable devices. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 195–205. https://doi.org/10.1145/2935334.2935340

[25] Ursula Hess, Rainer Banse, and Arvid Kappas. 1995. The intensity of facial expression is determined by underlying affective state and social situation. *Journal of personality and social psychology* 69, 2 (1995), 280. https://doi.org/10.1037/0022-3514.69.2.280

[26] Qiong Huang, Ashok Veeraraghavan, and Ashutosh Sabharwal. 2017. TabletGaze: dataset and analysis for unconstrained appearance-based gaze estimation in mobile tablets. *Machine Vision and Applications* 28, 5 (01 Aug 2017), 445–461. https://doi.org/10.1007/s00138-017-0852-4

[27] Mohamed Khamis, Anita Baier, Niels Henze, Florian Alt, and Andreas Bulling. 2018. Understanding Face and Eye Visibility in Front-Facing Cameras of Smartphones used in the Wild. *Proceedings of the 36th Annual ACM Conference on Human Factors in Computing Systems* 36 (2018), 5. https://doi.org/10.1145/3152832.3173854

[28] Kyung Hwan Kim, Seok Won Bang, and Sang Ryong Kim. 2004. Emotion recognition system using short-term monitoring of physiological signals. *Medical and biological engineering and computing* 42, 3 (2004), 419–427.

[29] Thomas Kosch, Mariam Hassib, Daniel Buschek, and Albrecht Schmidt. 2018. Look into My Eyes: Using Pupil Dilation to Estimate Mental Workload for Task Complexity Adaptation. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI EA '18)*. ACM, New York, NY, USA, Article LBW617, 6 pages. https://doi.org/10.1145/3170427.3188643

[30] Thomas Kosch, Mariam Hassib, Paweł W. Woźniak, Daniel Buschek, and Florian Alt. 2018. Your Eyes Tell: Leveraging Smooth Pursuit for Assessing Cognitive Workload. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, Article 436, 13 pages. https://doi.org/10.1145/3173574.3174010

[31] Thomas Kosch, Jakob Karolus, Havy Ha, and Albrecht Schmidt. 2019. Your Skin Resists: Exploring Electrodermal Activity As Workload Indicator During Manual Assembly. In *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems* (Valencia, Spain) *(EICS '19)*. ACM, New York, NY, USA, Article 8, 5 pages. https://doi.org/10.1145/3319499.3328230

[32] Thomas Kosch, Albrecht Schmidt, Simon Thanheiser, and Lewis L. Chuang. 2020. One does not Simply RSVP: Mental Workload to Select Speed Reading Parameters using Electroencephalography. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. ACM, New York, NY, USA. https://doi.org/10.1145/3313831.3376766

[33] Myungho Lee, Kangsoo Kim, Hyunghwan Rho, and Si Jung Kim. 2014. Empa Talk: A Physiological Data Incorporated Human-computer Interactions. In *Proceedings of the Extended Abstracts of the 32Nd Annual ACM Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI EA '14)*. ACM, New York, NY, USA, 1897–1902. https://doi.org/10.1145/2559206.2581370

[34] Yu-Kang Lee, Chun-Tuan Chang, You Lin, and Zhao-Hong Cheng. 2014. The dark side of smartphone usage: Psychological traits, compulsive behavior and technostress. *Computers in Human Behavior* 31, Supplement C (2014), 373 – 383. https://doi.org/10.1016/j.chb.2013.10.047

[35] Robert LiKamWa, Yunxin Liu, Nicholas D. Lane, and Lin Zhong. 2013. MoodScope: Building a Mood Sensor from Smartphone Usage Patterns. In *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services* (Taipei, Taiwan) *(MobiSys '13)*. ACM, New York, NY, USA, 389–402. https://doi.org/10.1145/2462456.2464449

[36] Zhe Liu, Anbang Xu, Yufan Guo, Jalal U. Mahmud, Haibin Liu, and Rama Akkiraju. 2018. Seemo: A Computational Approach to See Emotions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, Article 364, 12 pages. https://doi.org/10.1145/3173574.3173938

[37] Ludo Maat and Maja Pantic. 2007. Gaze-X: Adaptive, affective, multimodal interface for single-user office scenarios. In *Artifical Intelligence for Human Computing*. Springer, 251–271.

[38] Daniel McDuff, Amy Karlson, Ashish Kapoor, Asta Roseway, and Mary Czerwinski. 2012. AffectAura: An Intelligent System for Emotional Memory. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. ACM, New York, NY, USA, 849–858. https://doi.org/10.1145/2207676.2208525

[39] Daniel McDuff, Abdelrahman Mahmoud, Mohammad Mavadati, May Amr, Jay Turcot, and Rana el Kaliouby. 2016. AFFDEX SDK: A Cross-Platform Real-Time Multi-Face Expression Recognition Toolkit. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Santa Clara, California, USA) *(CHI EA '16)*. ACM, New York, NY, USA, 3723–3726. https://doi.org/10.1145/2851581.2890247

[40] Mohammad Obaid, Charles Han, and Mark Billinghurst. 2008. "Feed the Fish": An Affect-aware Game. In *Proceedings of the 5th Australasian Conference on Interactive Entertainment* (Brisbane, Queensland, Australia) *(IE '08)*. ACM, New York, NY, USA, Article 6, 6 pages. https://doi.org/10.1145/1514402.1514408

[41] Andreas Fsrøvig Olsen and Jim Torresen. 2016. Smartphone accelerometer data used for detecting human emotions. In *Systems and Informatics (ICSAI), 2016 3rd International Conference on*. IEEE, IEEE, New York, NY, USA, 410–415. https://doi.org/10.1109/ICSAI.2016.7810990

[42] Rosalind W. Picard. 2003. Affective computing: challenges. *International Journal of Human-Computer Studies* 59, 1 (2003), 55 – 64. https://doi.org/10.1016/S1071-5819(03)00052-1 Applications of Affective Computing in Human-Computer Interaction.

[43] Rosalind W. Picard, Elias Vyzas, and Jennifer Healey. 2001. Toward machine emotional intelligence: analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 10 (Oct 2001), 1175–1191. https://doi.org/10.1109/34.954607

[44] 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)*. ACM, New York, NY, USA, 825–836. https://doi.org/10.1145/2750858.2804252

[45] Paul Rodriguez, Guillem Cucurull, Jordi Gonzàlez, Josep M. Gonfaus, Kamal Nasrollahi, Thomas B. Moeslund, and F. Xavier Roca. 2017. Deep Pain: Exploiting Long Short-Term Memory Networks for Facial Expression Classification. *IEEE Transactions on Cybernetics* PP, 99 (2017), 1–11. https://doi.org/10.1109/TCYB.2017.2662199

[46] Tobias Ruf, Andreas Ernst, and Christian Küblbeck. 2011. *Face Detection with the Sophisticated High-speed Object Recognition Engine (SHORE)*. Springer Berlin Heidelberg, Berlin, Heidelberg, 243–252. https://doi.org/10.1007/978-3-642-23071-4_23

[47] James A Russell. 1994. Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological bulletin* 115, 1 (1994), 102. https://doi.org/10.1037/0033-2909.115.1.102

[48] Jan Scholz, Miriam C Klein, Timothy EJ Behrens, and Heidi Johansen-Berg. 2009. Training induces changes in white-matter architecture. *Nature neuroscience* 12, 11 (2009), 1370–1371. https://doi.org/10.1038/nn.2412

[49] Claudia Schrader, Julia Brich, Julian Frommel, Valentin Riemer, and Katja Rogers. 2017. Rising to the challenge: An emotion-driven approach toward adaptive serious games. In *Serious Games and Edutainment Applications*. Springer, New York, NY, USA, 3–28.

[50] Shams Shapsough, Ahmed Hesham, Youssef Elkhorazaty, Imran A. Zualkernan, and Fadi Aloul. 2016. Emotion recognition using mobile phones. In *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*. IEEE, New York, NY, USA, 1–6. https://doi.org/10.1109/HealthCom.2016.7749470

[51] Anna Ståhl, Kristina Höök, Martin Svensson, Alex S. Taylor, and Marco Combetto. 2009. Experiencing the Affective Diary. *Personal and Ubiquitous Computing* 13, 5 (01 Jun 2009), 365–378. https://doi.org/10.1007/s00779-008-0202-7

[52] Myrthe Tielman, Mark Neerincx, John-Jules Meyer, and Rosemarijn Looije. 2014. Adaptive Emotional Expression in Robot-child Interaction. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction* (Bielefeld, Germany) *(HRI '14)*. ACM, New York, NY, USA, 407–414. https://doi.org/10.1145/2559636.2559663

[53] Hitomi Tsujita and Jun Rekimoto. 2011. HappinessCounter: Smile-encouraging Appliance to Increase Positive Mood. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI EA '11)*. ACM, New York, NY, USA, 117–126. https://doi.org/10.1145/1979742.1979608

[54] Beverly Woolf, Winslow Burleson, Ivon Arroyo, Toby Dragon, David Cooper, and Rosalind Picard. 2009. Affect-aware tutors: recognising and responding to student affect. *International Journal of Learning Technology* 4, 3-4 (2009), 129–164.

[55] SungHyuk Yoon, Sang-su Lee, Jae-myung Lee, and KunPyo Lee. 2014. Understanding Notification Stress of Smartphone Messenger App. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI EA '14)*. ACM, New York, NY, USA, 1735–1740. https://doi.org/10.1145/2559206.2581167

[56] Tianyi Zhang, Abdallah El Ali, Chen Wang, Alan Hanjalic, and Pablo Cesar. 2020. RCEA: Real-Time, Continuous Emotion Annotation for Collecting Precise Mobile Video Ground Truth Labels. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376808