Teardrops on My Face: Automatic Weeping Detection from Nonverbal Behavior

Dennis Küster, Lars Steinert, Marc Baker, Nikhil Bhardwaj, and Eva G. Krumhuber

Abstract—Human emotional tears are a powerful socio-emotional signal. Yet, they have received relatively little attention in empirical research compared to facial expressions or body posture. While humans are highly sensitive to others' tears, to date, no automatic means exist for detecting spontaneous weeping. This paper employed facial and postural features extracted using four pre-trained classifiers (FACET, Affdex, OpenFace, OpenPose) to train a Support Vector Machine (SVM) to distinguish spontaneous weepers from non-weepers. Results showed that weeping can be accurately inferred from nonverbal behavior. Importantly, this distinction can be made before the appearance of visible tears on the face. However, features from at least two classifiers need to be combined, with the best models blending three or four classifiers to achieve near-perfect performance (97% accuracy). We discuss how direct and indirect tear detection methods may help to yield important new insights into the antecedents and consequences of emotional tears and how affective computing could benefit from the ability to recognize and respond to this uniquely human signal.

Index Terms—Weeping, tears, facial expression, body posture, support vector machine (SVM).

1 Introduction

Emotional tears are believed to serve uniquely human functions such as appeasement, distress, and helplessness, signaling a need for social support [1], [2]. Their capacity to induce prosocial responses in observers is well-known across many cultures [3]. Tears have been consistently shown to facilitate the perception of sadness [4], [5], with even brief exposures to tearful faces impacting emotion perception [4], [6]. Nevertheless, contemporary research on the functions of emotional tears is still limited by a surprisingly narrow selection of stimuli and methods [7]. Much of the available experimental work has used digitally manipulated (i.e., "photo-shopped") images [5], with only a few studies examining spontaneously elicited tears [8]. Although there have been a few tentative efforts to improve the quality of tear-related stimuli [4], [9], [10], emotional tears have thus far played a less important role in affective computing.

A major reason for such negligence is the lack of annotated training data. Some of the most potent elicitors of emotional tears throughout the lifespan (e.g., funerals, weddings, or divorces) are indeed rare and difficult to assess [11]. Also, a person's proneness to crying is subject to a plethora of factors, including age, situational demands, gender stereotypes, and socio-cultural norms [1], [12]. For example, young adults cry less frequently than other age groups [11], and adult men in Western societies report crying 2-4 times less frequently than women [13]. Crying may thus appear to be less common in certain types of contexts. Despite their enormous variability in occurrence, crying can often be elicited in rather mundane situations, e.g., when watching s sad movie or during interpersonal conflicts with parents, friends, and romantic partners [11]. This makes emotional tears a surprisingly common phenomenon [1], [14]. Furthermore, tears can be successfully evoked in the laboratory by showing participants personally-relevant sadness inducing films [8], [15]. Today, adult crying is increasingly studied across a broad range of settings, including high-stakes social interactions [1], [16]. Tears may also be important cues during therapy [14], [17], [18], in the courtroom [19], or in political advertising [20]. Nevertheless, until the recent publication of a first database on spontaneously elicited dynamic tears [8], none of the works in this field has yielded any openly accessible databases that could have been used for training automatic weeping detectors for videos.

1.1 Tears as Evolved Socio-Emotional Signals

Apart from the lack of suitable training data, there are historical reasons why adult emotional tears used to attract less interest from empirical researchers (cf., [14], [21]). In his seminal work on "The expression of emotions in man and animals" [22], Charles Darwin considered basal tears (i.e., tears serving to nourish and protect the eye) merely to be biologically adaptive responses, concluding that adults' emotional tears may not serve any definite purpose [1]. By contrast, clinicians such as Breuer and Freud [23] thought that shedding tears would facilitate catharsis and aid recovery. Today, we know that Darwin and Freud were most likely both wrong about the role of emotional tears [15], [24].

Over the last two decades, Darwin's notion has been challenged by several researchers pointing towards the role of human emotional tears throughout evolution [16], [24], [25]. In this vein, tears may have evolved as a handicap signal towards aggressive or defensive actions. The fact that tears tend to blur vision could make them *reliable signals*

- *D. Küster, L. Steinert, and N. Bhardwaj are with the Department of Computer Science, University of Bremen, 28359 Bremen, Germany. E-Mail: kuester@uni-bremen.de, lars.steinert@uni-bremen.de, nikhil@uni-bremen.de.*
- *M. Baker is with the Department of Psychology, University of Portsmouth, Portsmouth, UK. E-Mail: marc.baker@port.ac.uk.*
- *E. G. Krumhuber is with the Department of Experimental Psychology, University College London, London, UK. E-Mail: e.krumhuber@ucl.ac.uk.*

Manuscript received April 19, 2005; revised August 26, 2015.

for a number of states, such as the need for social support, loss of control, distress, or appeasement. Tears may also carry a certain cost or risk to the encoder (i.e., the weeper), dependent on the type of context [26]. For example, tears that are perceived as inappropriate (e.g., in stressful work situations) may reflect negatively upon the crier, evoking the impression that the person is weak and unprofessional [1], [27]. Thus, advancing knowledge on the antecedents and consequences of crying may help better understand the functions of crying.

1.2 Tears in Concert with Facial Expressions

In terms of ethical concerns, research on the functions of tears in everyday social interaction may pose a challenge for data collection. However, much can be learnt from decoding studies by considering emotional tears as a cue for observers. In a growing number of studies, tears have been shown to specifically enhance sadness ratings [4], [28] and the perceived intensity of emotional states such as sadness, anger, and fear [29]. Enhanced sadness ratings, also called the *tear effect* [21], were also found in the context of facial photographs with neutral expressions, as well as computer-rendered images [4], [25]. Thus, tears appear to be implicitly associated with sadness and negative affect [30]. However, tears could also have more emotion-specific effects, as shown by attenuated perceptions of disgust and surprise in the presence of tears [4]. While most prior decoding studies have been limited to static images and digitally manipulated tears [5], [7], the *tear effect* has recently also been demonstrated for videos [8]. In this work, sadness perceptions interacted with the progression of the videos over time, with the largest differences in perceived sadness between weepers and non-weepers being found when the weepers started to cry. Notably, this effect occurred despite the non-weepers self-reporting very high levels of sadness that were not significantly lower than those of weepers. This raises the question which visual cues or features may have driven the tear effect [8]. Here, machine learning methods may be able to predict the occurrence of tears on the basis of facial actions and other concomitant behaviors (i.e., postural cues). Furthermore, they could help reveal which visual features may be most important for this task.

Automatic weeping detection may perform best when the observable behavioral differences between tearful vs. non-tearful sad encoders are most evident, i.e., when the first teardrops appear. However, it may also be possible to distinguish weepers and non-weepers already during the buildup phase. Compared to other emotional expressions, sadness and tearing tend to require substantially more time to emerge and subside [1]. For example, video-based crying-inductions typically require several minutes to fully take effect [15], followed by a slow recovery period [31]. This slow nature might make weeping predictable well in advance of any visible signs of teardrops on the face. However, it may also be possible to distinguish weepers and non-weepers already during the buildup face. If this is the case, then such a weeping detector might (1) provide a new tool for studying the *psychological* antecedents and consequences of tears (see [1]). Conversely, (2) such an early detection might provide substantial benefits for developing intelligent cognitive systems, e.g., in the context of computer-assisted therapy (e.g., [32]).

1.3 Affective Computing

With the rise of publicly available datasets, computational power, and improvements in (mobile) sensors and algorithms, the field of affective computing has made rapid advances in recent years. Affective computing comprises a machine's ability to recognize, express, respond to and influence its users' emotions [33], making Human-Computer Interaction (HCI) more natural and engaging. Accordingly, numerous studies have employed affect recognition systems to infer emotions from facial expressions [34], body posture and gestures [35], and eye gaze.

As human behavior is multimodal by nature, there has been a growing consensus that "ideal" systems for automatic affect analysis should be multimodal [36]. Consequently, the combination of different modalities has been the focus of multiple studies [37], [38], [39], public datasets [40], [41], [42], and affect recognition challenges [43], [44]. However, the face arguably remains the most important nonverbal source of affective information [45]. Facial expressions can be recorded non-intrusively and at low cost using ordinary video cameras, and analyzed through various user-friendly commercial classifiers (e.g., FACET, Affdex, [46]). Many of these systems additionally provide basic information about head pose and eye gaze [47], which open-source software tools (e.g., OpenFace, OpenPose) can further enrich to extract features related to body posture and gestures [48]. The latter two may be particularly informative about the individual's emotional state [35], [49]. For instance, Gunes and Picardi [50], [51] demonstrated that the combination of facial expressions and upper-body gestures outperforms unimodal approaches. Not surprisingly, additional modalities such as speech are increasingly leveraged for multimodal affective computing. For example, Kessous et al. [52] combined facial expressions, body gestures, and speech features to automatically classify eight discrete emotional states. Classifiers trained on body gestures were even found to outperform (67.1 %) those that relied on facial expressions (48.3 %) and speech (57.1 %) only. The best recognition results were obtained when fusing speech and gesture features (75 %) at a feature level. Hence, non-verbal behavior conveyed by body posture and gestures can be an important channel for emotion communication. Supportive evidence comes from clinical research suggesting that tears may be associated with other nonverbal behaviors, such as sudden movements, wiping, touching, or hiding the face [1], [53].

1.4 Towards Automatic Tear Detection

Concordance between different components of the human emotion system is typically limited [54], and emotional tears are unlikely to be an exception to this rule. While the lack of training data remains a challenge, there are now a few publicly available data-sets featuring tears in still images [5] and videos [8]. To this end, a machine learning approach could be taken to detect when someone is or will be crying. If successful, this offers a powerful new research tool for the

study of sadness[1] and depression, helping to enhance affect sensing across a broad range of applications.

Given that open-access data on tears are still rather limited [5], [8], we decided against more data-hungry deep-learning approaches [55]. Hence, the present work examines facial and postural features extracted by several well-established machine classifiers for emotion recognition, which were then submitted to more traditional Support Vector Machines (SVMs). The major questions we aimed to address are the following: (1) Is it possible to infer weeping based on non-verbal facial and postural behaviors? If yes, which features are most important for automatic weeping detection? (2) Which classifiers, or combinations thereof, are most suitable for indirectly detecting (non-)weepers? (3) Can weeping be predicted before the moment the first tear becomes visible?

2 METHODOLOGY

2.1 Data Collection

The video data used in the present study comprised 24 participants from the Portsmouth Dynamic Spontaneous Tears Database (PDSTD) [8] and 10 participants from the same subject pool [56] but whose data were not included in the PDSTD[2] [56]. As detailed in [8] and [56], female students (hereafter referred to as encoders) were invited to the laboratory to watch a self-selected sad movie[3] (10-15 min) and a neutral film clip about owls (approximately 5 min). They identified the scene of the sad film they found most emotionally arousing (i.e., saddest). Dynamic facial behavior was recorded with a frame rate of 30 fps using a Logitech C920 Pro HD webcam and a video resolution of 1920 x 1080 pixels [8]. Weeping was detected manually via infrared thermal imaging (FLIR A655sc). The resulting dataset consisted of 30 s episodes extracted from the end of the neutral films, the 30 s immediately prior to the saddest moment, and the 30 s from 10 s before to 20 s after the saddest moment (non-weepers) or the first tear (weepers) (see also [8]).

2.2 Data set

We analyzed the videos of thirty-four encoders (M_{Age}: 22.18, SD= 4.67), who were either weeping (n=16, M_{Age}= 23.94, SD= 5.90) or not weeping (n= 18, M_{Age}= 20.61, SD= 6.13) in response to sad movies. We expected the time around the moment of the first tear to be the most informative for distinguishing between weepers and non-weepers. We therefore selected the 30 s of highest emotional intensity as indicated by the onset of the first tear[4] (weepers) or the self-identified saddest moment (non-weepers).

2.3 Feature Sets

We focused on individual facial muscles (so-called Action Units, AUs) as defined by the Facial Action Coding System (FACS; [57]). We processed the videos using four (commercial Affdex [58], FACET [59]) and non-commercial (OpenFace 2.0 [60], and OpenPose [61]) classifiers to extract features based on facial activity and body posture. While several established off-the-shelf classifiers provide estimates of AUs, cross-system evaluation studies to date have focused mainly on basic emotions [62]. Thus, little is known about the comparative reliability of single AUs. However, prior work has demonstrated significant differences in classifier performance between posed and spontaneous expressions [47], [63], as well as between classifiers [62], suggesting that recognition accuracy may differ substantially between systems, emotions, and individual AUs. We therefore extracted AU-features from three different facial expression recognition systems (Affdex, FACET, OpenFace 2.0). Although all of them output similar numbers of AUs, it is likely that one system performs vastly better for some AUs than another and vice-versa. Hence, each classifier might contribute substantial unique information to our model. Besides facial information, we also assessed postural features using OpenPose [61], as suggested by earlier works on objective coding of crying behavior [53]. Table 1 provides an overview of the different machine classifiers, the included channels, and the features considered in this study.

TABLE 1
Overview of the machine classifiers, channels, and feature sets.

Classifier	*Channel*[1]	*Features*
FACET (FA)	FE	Intensity of AU1,2,4-7,9,10,12, 14,15,17,18,20,23-26,28,43
Affdex (AF)	FE	Intensity of AU1,2,4-7,9,10,12(L/R), 14,15,17,18,20,24-26,28,43
OpenPose (OP)	BP	D:E2E, D:E2S, D:Ea2Ea, D:S2S, D:Ha2N, N Mov[2]
OpenFace (OF)	FE	Intensity and presence of AU1,2, 4-7,9,10,12,14,15,17,20,23-26,45 and presence of AU28
	EG	EG X, EG Y[3]
	HP	H X-Lo, H Y-Lo H Z-Lo, H X-Ro, H Y-Ro, H Z-Ro[4]

[1] FE=Facial Expressions, BP=Body Pose, EG=Eye Gaze, HP=Head Pose. [2] Euclidean distance (D:) between the X, Y coordinates for eyes (E), ears (Ea), hands (Ha), shoulders (S), hands to the nose (N), and movement (Mov) of the nose compared to the previous frame. [3] Eye gaze direction in radians in world coordinates averaged for both eyes for the x-axis and y-axis. [4] Location (Lo) and Rotation (R) of the head in radians around X, Y, Z.

2.3.1 FACET (FA)

FACET (SDK v6.3; iMotions, 2016) is a commercial software for automatic facial expression recognition that was originally developed based on the Computer Expression Recognition Toolbox algorithm CERT [59]. FACET classifies frame-based facial expressions both in terms of FACS AUs as well as the six basic emotions [64].

1. Tears are also known to occasionally occur in the context of other emotions, such as very intense experiences of happiness [1]. However, they have most consistently been shown to be relevant in contexts involving sadness and a need for social support [3].

2. These additional participants were not included in the PDSTD because they did not provide the extended informed consent required to publish their non-anonymous raw video data. For the purpose of the present study, the original data could be processed locally, without revealing the participants' identity.

3. The original study involved only female participants due to the greater success of video-based weeping-elicitation in female encoders [15], [31].

4. The presence of tears was determined via infrared thermal imaging using a FLIR A655sc [56].

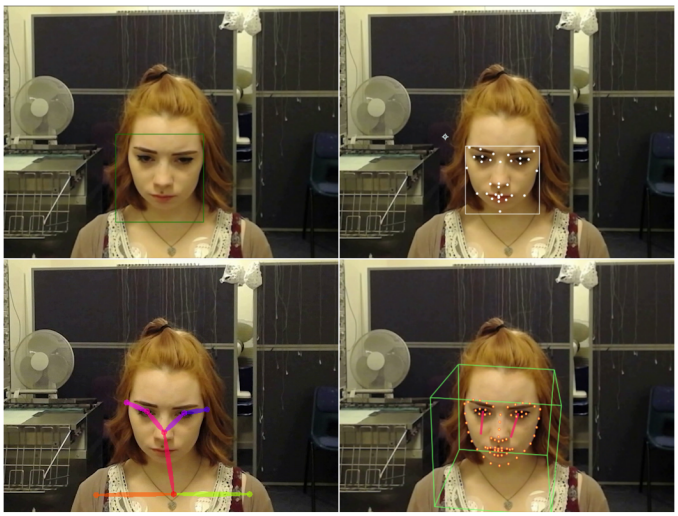


Fig. 1. Example of an annotated video frame from our data-set using FACET (top left), Affdex (top right), OpenPose (bottom left), and OpenFace (bottom right).

2.3.2 Affdex (AF)

Affdex (v7.0, iMotions) was developed by Affectiva, a spin-off company of the MIT Media Lab [62]. It uses SVM classifiers and Histogram of Oriented Gradient (HOG) features [58] to recognize basic emotions as well as 19 different AUs[5].

2.3.3 OpenPose (OP)

OpenPose is an open-source system for the detection of human body, hand, facial and foot key-points in real-time [61]. For all sessions, the position of 25 body landmarks[6] and 20 hand landmarks per hand[7] were extracted for all frames.

2.3.4 OpenFace (OF)

OpenFace 2.0 is an open-source facial behavior analysis toolkit [60] which allows for facial landmark detection, head pose and eye-gaze and estimation, and AU recognition. For all videos, we extracted facial features, the location and rotation of the head (head pose), and the direction of eye-gaze based in individual frames.

2.4 Pre-Processing and Classification

We sliced all (facial and postural) feature streams into 5 s segments with 50 % overlap and assigned the label of the corresponding encoder (crier, non-crier) to it. Next, we aggregated these segments by calculating statistical functionals for each feature, namely the mean, median, max, skewness and kurtosis. We followed Kessous et al. [52] by combining the feature sets in an early fusion approach. We applied L2 normalization on each feature vector to have a unit norm. We used a Support Vector Machine (SVM) with an RBF kernel for classification and optimized the γ ($\gamma \in \{.0001, .001, .01, .1, 1.0\}$) and C ($C \in \{.001, .1, 10, 25, 50, 100, 1000\}$) parameters in a 3-fold Cross Validation (CV) using the training data. We evaluated this approach by using a user-independent[8] 5-fold CV with the final prediction for each encoder (crier, non-crier) being obtained through majority voting across all samples of that individual. To test for statistical significance, we conducted non-parametric McNemar-Tests on the global prediction level for each feature set against the baseline (chance level). The baseline accuracy is .558 for all feature sets which results from the slightly imbalanced class distribution (55,8 % of the participants belong to the class crier). We used Accuracy (Acc.), Precision, Recall and F1-Score (F1) as evaluation metrics.

5. Due to copyright reasons, Affdex AUs are not officially labeled as such.

6. Nose, shoulders, elbows, wrists, middle Hip, left hip, right hip, ankles, knees, eyes, ears, heels, big toes, small toes

7. Wrist, and each of the three joints and the beginning of the finger for thumb, index, middle, ring and pinky finger

3 Results

3.1 Performance per Classifier

Tab. 2 and Fig. 2 show the user-independent classification results based on individual feature sets (per system) and their combination.

TABLE 2
Classification results based on a user-independent 5-fold CV. The baseline accuracy is .558 for all machine classifiers which results from the class distribution (55,8 % of the participants belong to the class crier). The level of significance is indicated by: * (p<0.05) and *** (p<0.001).

Classifier	*Acc.*	*Precision*	*Recall*	*F1*
OF	.588	.609	.737	.667
FA	.618	.636	.737	.683
OP	.618	.667	.632	.649
AF	.765	.824	.737	.778
AF_FA	.647	.667	.737	.700
OP_FA	.676	.700	.737	.718
OF_FA	.676	.700	.737	.718
OF_OP	.706	.765	.684	.722
OF_AF	.765	.762	.842	.800
AF_OP*	.853	.889	.842	.865
OF_OP_FA	.735	.812	.684	.743
AF_OP_FA*	.853	.889	.842	.865
OF_AF_FA*	.853	.889	.842	.865
OF_AF_OP***	.971	.950	1.00	.974
OF_AF_OP_FA***	.971	.950	1.00	.974

When investigating the performance of FACET (.618), OpenPose (.618), and OpenFace (.588) separately, tear classification was close to the baseline, thereby failing to reach significance (all ps>.05). Classification performance was somewhat higher for Affdex (.765), but did not reach significance either (p=.210). Interestingly, the combination of Affdex and OpenPose significantly ($\chi^2 = 3$, p=.002) exceeded the baseline accuracy (.853). The combined classifications from both systems were thus similarly accurate as three of the four three-system combinations. Furthermore, the combination of Affdex and OpenPose outperformed other two-system combinations, such as OpenFace and OpenPose. This suggests that Affdex may have performed better than other classifiers at detecting some of the most relevant AU-features features in this context. Best classification results were achieved when

8. A user-independent Cross Validation considers each subject separately to estimate performance for users based on the data of the other users.

combining OpenFace, Affdex and OpenPose (.971) which significantly exceeded the baseline ($\chi^2 = 0, p<.001$) and were similarly accurate as the combination of all four systems (.971, $\chi^2 = 0$, $p<.001$). These results suggest that the respective classifiers capture complementary sources of information, which can be exploited by the combined models.

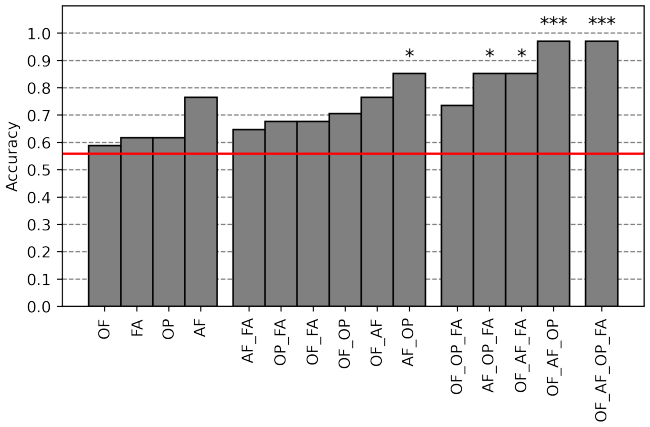


Fig. 2. Classification accuracy of the four classifiers and their combinations. The red line indicates the baseline (.558). The level of significance is indicated by: * ($p<0.05$) and *** ($p<0.001$).

3.2 Performance per Encoder

While the combined models generally outperformed models based on individual systems, a more fine-grained analysis may help reveal (1) which systems contributed good (vs. poor) or unique (vs. redundant) information, (2) which individual videos might show easily recognizable "typical" (non-) weeping, and (3) which videos could be regarded as borderline cases. We thus explored the classification accuracies of individual and combined systems at the level of individual encoders, separated by weepers and non-weepers.

As illustrated by Fig. 3, four of the nineteen non-weepers (21%) were always classified correctly, and an additional five non-weepers (26%) were misclassified only once or twice. In comparison, none of the weepers were always correctly classified by the individual- and combined systems. Overall, this exploratory analysis suggests no substantial difference in classification accuracies for weepers and non-weepers[9]. Nevertheless, a small number of encoders appeared to be particularly challenging to classify as either weepers or non-weepers. In particular, encoder numbers 18, 20, 23, and 24 can be regarded as borderline cases.

As further shown by Fig. 3, the three-system combinations generally matched and improved the performance patterns of the best-performing two-system combinations. I.e., gains were largely achieved without introducing any new mistakes. For example, the best-performing three-system combination of OpenFace, Affdex, and OpenPose (OF_AF_OP) eliminated nearly all of the misclassified cases of the best two-system combination of Affdex and OpenPose (AF_OP). Likewise, the

9. We refrained from any statistical comparisons between weepers and non-weepers due to the large amount of feature-overlap between these models. Furthermore, the two best-performing system-combinations already achieved a near-perfect performance.

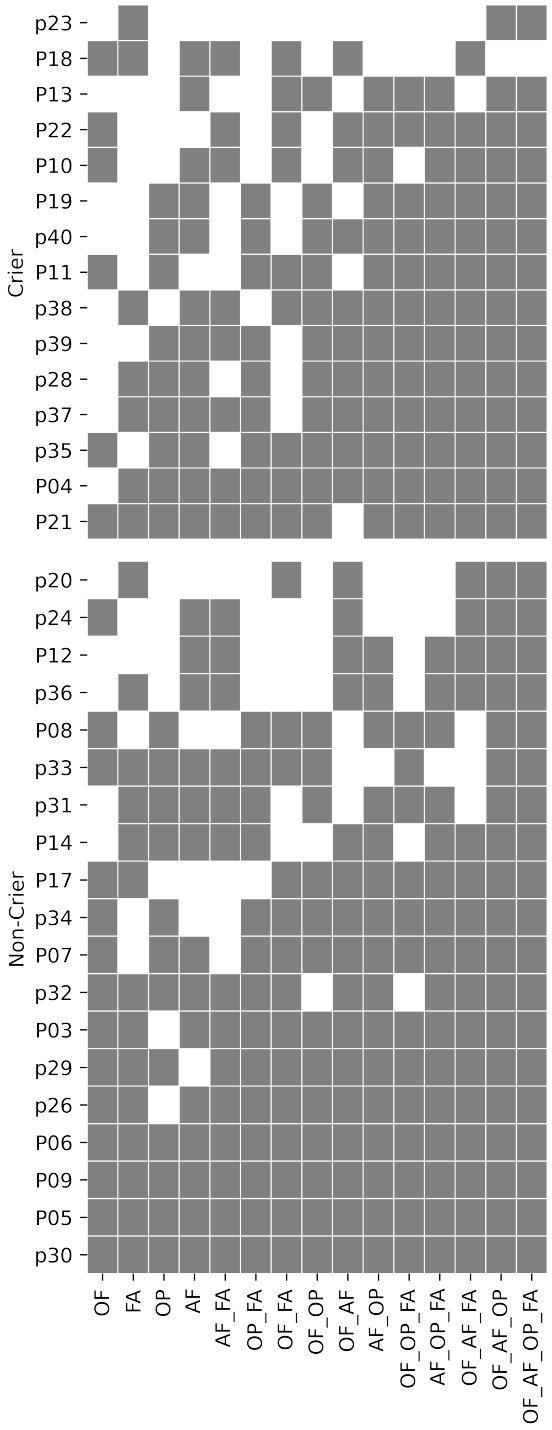


Fig. 3. Classification performance for individual encoders based on the four classifiers and their possible combinations. The grey cells indicate correct classifications. White cells show incorrect classifications.

combination of OpenFace, Affdex, and FACET (OF_AF_FA) contributed a few additional correct classifications on top of the two-system combination of OpenFace and Affdex (OF_AF). The combined system of Affdex, OpenPose, and FACET (AF_OP_FA), perfectly duplicated the per-subject level performance of Affdex and OpenPose (AF_OP). Finally, OpenFace, OpenPose and FACET (OF_OP_FA) appeared to slightly improve performance of the two-system combination

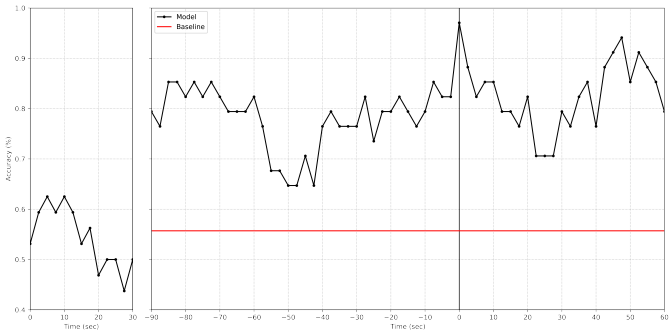


Fig. 4. Classifier performance across time for the first 30 s of the neutral owl video (left panel), and 90 s before and 60 s after the *sad/tear* moment (right panel), as indicated by the vertical line at t=0. The red line represents the baseline performance.

of OpenFace and OpenPose (OF_OP). In comparison, performance patterns based on individual systems appeared to be more variable. Individually, OpenFace (FA) and OpenPose (OP) showed markedly different patterns of mistakes across individual weepers and non-weepers. Together, these results suggest that most of the systems contributed relevant features to the combined models, despite their relatively poor individual performance.

3.3 Weeping Prediction and Feature Importance

Our classification results suggest that the presence of tears may be accurately predicted by combining features from extant classifiers for posture- and expression recognition. For example, the combined features from Affdex and OpenPose were sufficient for a correct classification of weepers and non-weepers in about 85 % of all cases. However, the classifiers employed in this study are likely to provide at least partially redundant information because they aim to extract many of the same AUs. While a standard cross-validation approach was applied to prevent potential overfitting [65], the question still arises to what extent the final model captures meaningful differences between weepers and non-weepers. Here, a time-based model analysis may provide further insights into the validity of our approach, particularly given the slow trajectory of sadness and tears. Furthermore, the performance gains achieved by combinations of different classifiers suggest that sadness-prototypical combinations of AUs alone may not be sufficient to explain the models' success. We thus performed a permutation-based analysis of feature importance to examine which features contributed the most towards the classification of weepers and non-weepers.

3.3.1 Time-Based Analysis

If the machine learning model was driven by meaningful crying-related behaviors, then its ability to correctly distinguish between weepers and non-weepers should exhibit properties that are comparable to the trajectory of sadness and tears observed in the laboratory [11], [15]. Conversely, if the model has mostly learned from spurious correlations [66], then it might exhibit a wildly different behavior when exposed to emotionally neutral samples of the same data. We therefore decided to analyze the evolution of our model performance over time. We expected that model-performance should generally mirror the known temporal evolution of the sadness- and tear-response in this data set. The model should thus fail to distinguish weepers from non-weepers during the baseline-period. Furthermore, weeping predictions may significantly exceed baseline accuracies already before the sad/tear moment.

To examine this hypothesis, we re-applied our model iteratively across shifting time windows. We stacked 50 frames (representing 5 s), calculating maxima, mean, median, kurtosis and skew, into one window and used a shift of 25 frames (2.5 s), i.e. each window containing 2.5 s of the previous window. We combined 11 continuous windows into one group representing 30 s of video, resulting in 61 continuous groups[10] for the sad video and 13 groups for the neutral video. Each group thus contained one new window (2.5 s new data) that differed from the previous group. Starting with the first neutral group, we trained the same SVM with an RBF kernel to classify each window and perform majority voting over the output to classify the group. We then conducted a 5-fold CV at each time interval.

As illustrated by Fig. 4, the time-based analysis of tear-prediction accuracy appeared to be generally in line with the expected pattern of differences between weepers and non-weepers reported by human observers in [8]. We observed the best model performance, i.e. 97 % accuracy, for the window starting at the sad tear moment. In contrast, classification accuracy during the neutral video hovered about the baseline, with a maximum accuracy score of 62.5 %. Somewhat unexpectedly, the second highest peak, with 94 % accuracy, emerged about only 50 s after the sad/tear moment. However, overall, performance during the build-up -and recovery phases appeared to be well in-between that of the neutral video and the sad/tear moment. Together, these results suggest that our model has likely learned to distinguish between weepers and non-weepers based on observable crying-related behaviors.

3.3.2 Permutation-based Feature Importance

Explainable algorithms have gained growing attention in the field of machine learning as the increase in model complexity often implies a decrease in their interpretability [67]. However, understanding the mechanisms of a model's decisions can increase trust in affective computing and provide important insights into the mechanisms underlying human behavior [67]. Ideally, analyses of explainable algorithms should aim to bridge algorithm-generated explanations and established theories [68]. We apply a permutation-based feature importance (PB-FI) to assess which individual features are important for the classification of weepers versus non-weepers. PB-FI describes the model's performance decrease when a single feature is randomly shuffled [69]. The underlying assumption is that the permutation of important features will substantially weaken the model's performance, whereas the permutation of unimportant ones will only have a minuscule effect.

10. Encoders differed considerably from one another with respect to when they started to cry, and how close this time point was to the end of the sad video. We therefore limited the temporal analysis from 90 s prior to tearing up until 90 s after the first tear. Since each group contains 30 s of video information, the resulting performance plot ends 60 s after the sad/tear moment.

The four classifiers (FACET, Affdex, OpenFace, OpenPose) were trained using all feature sets for each time window as described in Sec. 3.3.1. Applying a user-independent 5-Fold CV, we randomly shuffled each individual feature of the test set 50 times. We then observed the average change in the model's accuracy over each feature and each fold for each time window.

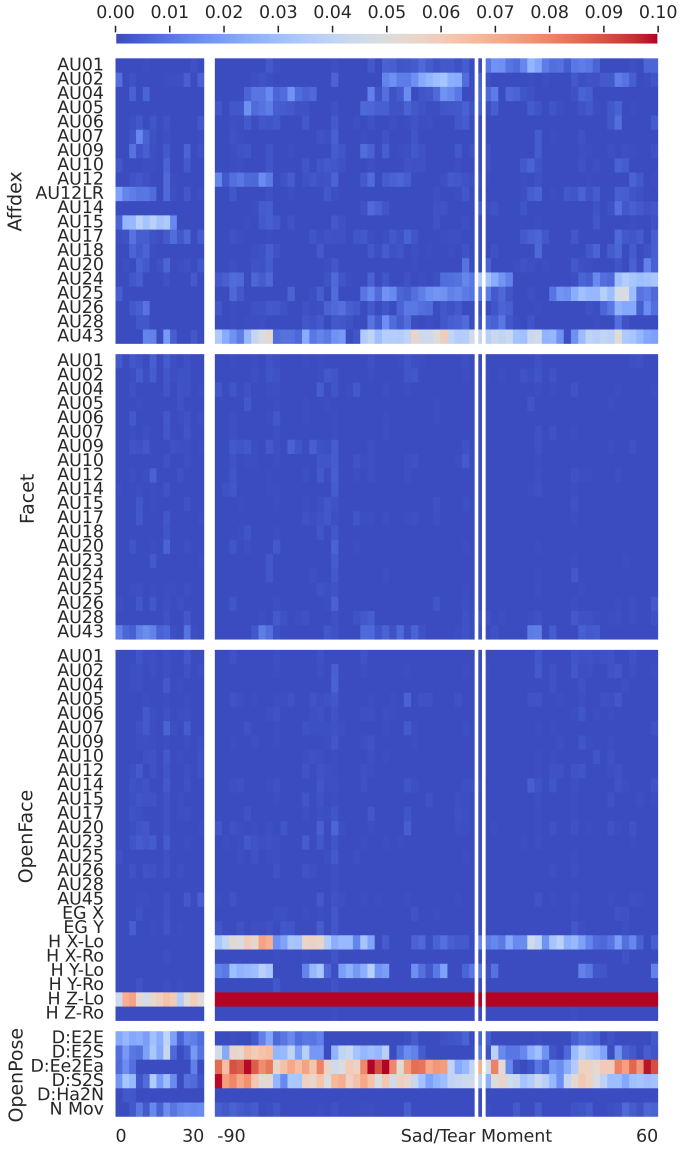


Fig. 5. The average permutation-based feature importance across participants over time measured as the change in accuracy (%) per feature type. Red-shaded cells indicate important features while blue cells indicate unimportant features. The color scale is limited to 0.1 to improve visibility in the mid-range of values. EG X: eye gaze x-axis; EG Y: eye gaze x-axis; H X-Lo: head x-axis location; H X-Ro: head X-axis rotation; H Y-Lo: head y-axis location; H Y-Ro: head Y-axis rotation; H Z-Lo: head y-axis location; H Z-Ro: head Y-axis rotation; D:E2E: distance between both eyes; D:E2S distance between eyes to shoulder; D:Ea2EA: distance between both ears; D:S2S: distance between both shoulders; D:Ha2N: distance between hands and nose; N Mov: movement of the nose compared to the previous frame.

The majority of the features used by our model were facial AUs, including multiple instances of AUs that are considered to be prototypical for sadness by basic emotion theory (AU1, AU4, AU15, AU17; [57], [70]). However, AUs were generally much less important in the PB-FI analysis compared to specific head- and body postures[11] (see Fig. 5). Overall, postural features such as the distance of the head to the camera (H Z-Lo), the distance between both eyes (D:Ea2Ea), and the visual distance between both shoulders (D:S2S) appeared to be the most important features during, before, and after the sad/tear moment. Among the facial AUs, only AU43 (eyes closed) and AU24 (lip pressor) showed visibly elevated feature importances during the sad/tear moment, with AU25 (lips part) indicating slightly elevated importances before and after this event. Furthermore, the heatmap suggests little importance of single AUs as detected by either FACET or OpenFace. For example, when considering the sad/tear moment, the recognition accuracy would drop by 19.0 % when the feature H Z-Lo would not be accessible to the classifier, followed by D:Ee2Ea (6.9 %), AU43 (4.2 %), D:S2S 3.7 %, and AU24 3.4 %. Permuting AU25 leads to a decrease of 1.1 %.

Together, these results suggest that our model was particularly sensitive to changes in the distance of participants to the screen, as reflected by the three most important postural features, eye blinking, and the lips either pressing together or parting. In comparison, most individual AUs were not as important for the performance of our model.

4 Discussion

Due to a lack of reliable training data and a historical neglect of tears as a socio-emotional signal, weeping and emotional tears have previously played only a minor role in affective computing research. The present work took a first step towards remedying this situation by proposing an indirect approach for automatic tear detection. We leveraged four well-established classifiers for facial expression- and posture recognition to extract potential weeping-related features from a new data set of spontaneous dynamic sadness expressions and tears [8]. We then submitted the resulting feature sets to SVMs. Our results show that this approach succeeded in almost perfectly separating weepers from non-weepers during the sad/tear moment. Thus, we could detect whether encoders were merely sad or weeping when watching self-selected sad films.

Our results suggest that successful weeping detection may require a combination of features comprising various AUs as well as postural cues[12], extracted by means of two or more off-the-shelf classifiers. While the results of most single classifiers were at chance level, Affdex achieved nearly 77 % accuracy. While this result still fell short of reaching statistical significance, it is possible that larger data sets may allow for better performance by Affdex in the future. Nevertheless, combining Affdex-features with those from an additional open-access classifier (OpenPose) significantly outperformed the baseline at a level of about 85 % accuracy. In line with previous efforts pointing towards the importance of objective coding approaches of crying behaviors [53], this result suggests that postural cues may play a crucial role in distinguishing weeping from sadness. Furthermore,

11. Exact values and all raw data can be obtained from the first author upon request.

12. Apart from body posture per se, this includes also head pose and eye gaze.

our classifier comparisons suggest that Affdex may be able to contribute somewhat more weeping-related information than FACET. This may appear surprising given that FACET has previously been shown to outperform Affdex with respect to the classification of basic emotion expressions [46]. Nonetheless, Affdex could still be more reliable when it comes to capturing weeping-related facial AUs.

Our exploratory analysis of performance patterns per encoder suggests that each of the four classifiers contributed a certain amount of unique information. This result is most evident for the combination of OpenFace and OpenPose. Individually, the feature sets from both classifiers performed insufficiently in order to detect weeping. However, their patterns of (in-)correctly classified encoders were rather complementary, and both systems performed very well when combined with Affdex. Affdex was part of the best two-system and three-system combinations, suggesting a strong contribution of Affdex to the overall success of the classification task. Finally, the model constructed from FACET's features correctly classified two "edge cases" that were missed by all other single-system models. However, it contributed somewhat less to the success of the more aggregated models. Nevertheless, our approach appeared to benefit greatly from including different information sources.

The time-based performance analyses support the validity of our approach for automatic detection of weeping and tears. As expected, the best model performance was at chance level when encoders were watching a control video about owls [8], [56]. Notably, this video still elicited some (non-sadness) self-reported emotional responses from the encoders, including a certain amount of amusement and interest. Thus, while the emotion-inducing materials were neutral with respect to sadness, they still depicted some variance with in terms of the extracted features.

Finally, our analysis of feature importances suggests that AUs that were previously defined as prototypical for sadness by basic emotion theory (e.g., [70]) only play a subordinate role for automatic weeping detection and prediction. Surprisingly, changes in the distance to the camera and other postural features were substantially more important than any individual facial AUs. Among the AUs that did appear to matter more, self-regulatory behaviors such as eye-blinking and lip-pressing (and its opposite) seemed more relevant. Here, eye-blinking could be interpreted as directly associated with (attempts to control) the appearance of tears in the eyes [25]. Likewise, lip-pressing may point towards an effort to suppress the weeping- or sadness response. Similarly, our per-classifier performance results suggest that facial expressions alone may not be sufficient for accurate weeping prediction in this setting. Together, these findings appear to be in line with recent criticisms of the explanatory value of basic emotion theory (e.g., [71], [72], [73], [74]).

While the present work focused on classifying weepers and non-weepers during the sad/tear moment, additional analyses suggest that this distinction may be possible already a minute before (or after[13]) the onset of the tears. Future work could use a similar approach to predict weeping in other (mixed) emotional contexts before the occurrence of teardrops on the face.

4.1 Limitations

Although the present work was successful in detecting (non-)weeping based on facial AUs and postural features, some limitations remain. First, our model did not directly detect visible signs of tears; hence, it is likely not sufficiently tear-specific to reliably pinpoint the presence of tears in an image. In fact, our model should fail to detect artificially evoked "onion tears" [1], or digitally added tears [5], [7]. Instead, the model appears to detect a broader set of weeping-related behaviors that culminate in tears. This makes the current approach suitable for predicting when someone is going to cry - before the presence of any tears on the face. Furthermore, the results cannot be explained by spurious correlations. Specifically, it failed to separate weepers from non-weepers during the baseline period, and peak model performance coincided with the sad/tear moment. Our model performance therefore appears to be highly consistent with predictions from the crying literature (e.g., [1], [15], [31]).

The present work has been limited by the small size of available data sets featuring spontaneous emotional tears [5], [8]. Furthermore, all participants in the present data set were female and White [8]. Therefore, the stimulus set was not diverse with respect to gender and race. This would likely result in biases if the present algorithms were to be applied to future non-white and/or non-female data sets, without additional training. Due to the limited size of the present data set, large absolute gains in accuracy were required to demonstrate any statistically significant effects, which in turn depended on a relatively small number of subjects. Given these constraints, we decided to employ traditional SVMs rather than more data-hungry deep learning approaches (e.g., Convolutional Neural Networks). SVMs are a well-established method for small data sets [75], and we obtained excellent results for the required binary classification. Nevertheless, once larger data sets become available, deep learning approaches may achieve better results for the purposes of direct tear detection. At present, a *shallow* SVM-approach [76] appeared to provide a more robust and explainable starting point.

While the feature-importance analysis highlights postural and "self-regulatory" behaviors over more "prototypical" facial features of sadness, it is possible that the high degree of redundancy between facial AUs in our model impacted on the results. Most of the facial AUs were measured by three different systems (Affdex, FACET, OpenFace), whereas only a limited number of postural features was considered. Nonetheless, the most important postural features (H Z-Lo, D:Ee2Ea, D:S2S) may be similarly redundant since they were directly affected by changes in the distance to the screen, e.g., when participants were leaning forward or slumping backward into their seat. Furthermore, the AUs that did appear to be important for weeping classification were likewise measured more than once, with results from the feature-importance analysis being largely consistent with those from the classifier-level analyses. Hence, it is unlikely that additional analyses of feature importances at higher

13. We did not extend beyond this period because some encoders wept near to the end of the sadness-inducing film.

levels of aggregation would reveal a fundamentally different picture of results. Overall, while the distance-related features appeared to be particularly important, neither OpenFace, OpenPose, nor the combination of both were sufficient to achieve a recognition performance that was significantly above the baseline. In particular, the combination of distance features and AUs provided by OF only achieved a baseline-level classification performance. Together, these results suggest that using distance features alone would not be sufficient to predict weeping in this data set. However, the relative prominence of distance-related features might be specific to the emotion induction method employed in this particular dataset (watching sad films). Once new tear datasets become available, future work should therefore aim to compare classifier performance across different types of tear elicitation conditions.

Finally, we used a single, not multiple, data set to test our approach, showing that features from at least two classifiers are needed to achieve good results. Combining the two open-source classifiers (OpenFace, OpenPose) still fell short of significantly outperforming the baseline. It falls to future research to test how well the present findings generalize to other data sets or even real-time applications. The cross-validation results suggest that person-independent weeping recognition will likely be feasible in other use cases featuring spontaneously elicited emotional tears. In combination with open-source classifiers, a single commercial classifier (Affdex) may be sufficient for obtaining near-perfect results.

4.2 Future Directions

The current results point to several avenues for future research. For example, understanding the antecedents and consequences of tears has been identified as a major challenge in psychological crying research [1], [11]. The ability of our model to distinguish between weepers and non-weepers shortly before and after the onset of emotional tears suggests that further analyses of fine-grained nonverbal behaviors (e.g., via feature importance) could provide new insights into the mechanisms of crying processes. From a more applied perspective, the ability to predict weeping and tears could enhance a broad range of human-computer interaction scenarios, such as the capacity of intelligent agents in therapy to act empathically (e.g., [32], [77], [78]) or to respond to feelings of loss of control or disengagement (e.g., [79], [80], [81]).

In this work, there were a small number of edge cases as well as a somewhat larger number of (non-)weepers that appeared to be easier to classify. The results may prove useful for researchers who want to select suitable items from the PDSTD [8]. In the future, in-depth case studies may help reveal relevant information with respect to (1) the types of pre-weeping behaviors that are perceived as most typical by human observers, (2) the time point at which it is possible for humans to accurately detect crying, and (3) the usefulness of machine data in making those guesses.

We hope that the current indirect approach to infer weeping from nonverbal behavior may soon be complemented by more direct means for automatically detecting the presence of tears. The current results, based on *shallow* machine learning, suggest that the task of automatically detecting emotional tears might not be as daunting as previously believed. Apart from larger data-sets and deep learning approaches, automatic tear detection could be aided by thermal imaging to help pinpoint the moment when tears first begin to become visible on the face. Future work may combine direct and indirect detection methods for automatically predicting and verifying the presence of emotional tears. In the long term, such advances could help pave the way towards more natural and affect-sensitive systems [82].

5 Conclusion

Our research demonstrates that spontaneous weeping can be successfully inferred from fine-grained nonverbal behaviors. Towards this end, more attention should be paid to postural cues and non-prototypical AUs. We also show that it is possible to predict emotional tears prior to any visible signs on the face. Together these findings open up the possibility of future studies that combine direct and indirect tear-detection methods. The present study is a first step towards addressing this gap in the growing toolbox of research on affective computing.

Acknowledgments

This work was partially funded by the Klaus-Tschira-Stiftung.

References

[1] A. Vingerhoets, *Why only humans weep: Unravelling the mysteries of tears*. Oxford University Press, 2013.

[2] A. Gračanin, L. M. Bylsma, and A. J. J. M. Vingerhoets, "Why only humans shed emotional tears: Evolutionary and cultural perspectives," *Human Nature*, vol. 29, no. 2, pp. 104–133, Jun. 2018. [Online]. Available: http://link.springer.com/10.1007/s12110-018-9312-8

[3] J. H. Zickfeld *et al.*, "Tears evoke the intention to offer social support: A systematic investigation of the interpersonal effects of emotional crying across 41 countries," *Journal of Experimental Social Psychology*, vol. 95, p. 104137, Jul. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0022103121000378

[4] A. Gračanin, E. Krahmer, M. Balsters, D. Küster, and A. J. J. M. Vingerhoets, "How weeping influences the perception of facial expressions: The signal value of tears," vol. 45, pp. 83–105, 2021. [Online]. Available: https://doi.org/10.1007/s10919-020-00347-x

[5] S. J. Krivan and N. A. Thomas, "A Call for the Empirical Investigation of Tear Stimuli," *Frontiers in psychology*, vol. 11, pp. 52–52, Jan. 2020, publisher: Frontiers Media S.A. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/32082220

[6] M. J. H. Balsters, E. J. Krahmer, M. G. J. Swerts, and A. J. J. M. Vingerhoets, "Emotional tears facilitate the recognition of sadness and the perceived need for social support," *Evolutionary Psychology*, vol. 11, no. 1, p. 147470491301100, Jan. 2013. [Online]. Available: http://journals.sagepub.com/doi/10.1177/147470491301100114

[7] D. Küster, "Artificial tears in context: Opportunities and limitations of adding tears to the study of emotional stereotypes, empathy, and disgust," Tilburg, the Netherlands, Oct. 2015. [Online]. Available: http://rgdoi.net/10.13140/RG.2.2.12060.18567

[8] D. Küster, M. Baker, and E. G. Krumhuber, "PDSTD - The Portsmouth Dynamic Spontaneous Tears Database," *Behavior Research Methods*, Dec. 2021. [Online]. Available: https://doi.org/10.3758/s13428-021-01752-w

[9] D. Küster, "d-kuester/Extract-Tears-Action-Photoshop: Photoshop action designed to duplicate tears from one image to another," Jun. 2018. [Online]. Available: https://zenodo.org/record/4561858

[10] M. H. Alkawaz, A. H. Basori, D. Mohamad, and F. Mohamed, "Realistic Facial Expression of Virtual Human Based on Color, Sweat, and Tears Effects," *The Scientific World Journal*, vol. 2014, p. e367013, Jul. 2014, publisher: Hindawi. [Online]. Available: https://www.hindawi.com/journals/tswj/2014/367013/

[11] A. J. J. M. Vingerhoets and L. M. Bylsma, "The riddle of human emotional crying: A challenge for emotion researchers," vol. 8, no. 3, pp. 207–217, 2016. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1754073915586226

[12] L. S. Pauw, D. A. Sauter, G. A. van Kleef, and A. H. Fischer, "Stop crying! the impact of situational demands on interpersonal emotion regulation," vol. 33, no. 8, pp. 1587–1598, 2019. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/02699931.2019.1585330

[13] A. J. Vingerhoets and J. Scheirs, "Sex differences in crying: Empirical findings and possible explanations," in *Gender and Emotion: Social Psychological Perspectives*, A. H. Fischer, Ed. Cambridge University Press, Mar. 2000, pp. 143–165, google-Books-ID: tS1C8Sl5ysEC.

[14] L. M. Bylsma, A. Gračanin, and A. J. J. M. Vingerhoets, "A clinical practice review of crying research," vol. 58, no. 1, pp. 133–149, 2021.

[15] L. S. Sharman, G. A. Dingle, A. J. J. M. Vingerhoets, and E. J. Vanman, "Using crying to cope: Physiological responses to stress following tears of sadness." vol. 20, no. 7, pp. 1279–1291, 2020. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/emo0000633

[16] A. J. Vingerhoets, N. van de Ven, and Y. van der Velden, "The social impact of emotional tears," *Motivation and Emotion*, vol. 40, no. 3, pp. 455–463, 2016.

[17] C. 't Lam, A. Vingerhoets, and L. Bylsma, "Tears in therapy: A pilot study about experiences and perceptions of therapist and client crying," vol. 20, no. 2, pp. 199–219, 2018. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/13642537.2018.1459767

[18] K. L. Capps, K. Fiori, A. S. J. Mullin, and M. J. Hilsenroth, "Patient crying in psychotherapy: Who cries and why?" vol. 22, pp. 208–220, 2013.

[19] L. ten Brinke, S. MacDonald, S. Porter, and B. O'Connor, "Crocodile tears: Facial, verbal and body language behaviours associated with genuine and fabricated remorse." vol. 36, no. 1, pp. 51–59, 2012. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/h0093950

[20] B. Seibt, T. W. Schubert, J. H. Zickfeld, and A. P. Fiske, "Touching the base: heart-warming ads from the 2016 U.S. election moved viewers to partisan tears," *Cognition and Emotion*, vol. 33, no. 2, pp. 197–212, Feb. 2019. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/02699931.2018.1441128

[21] R. R. Provine, K. A. Krosnowski, and N. W. Brocato, "Tearing: Breakthrough in human emotional signaling," vol. 7, no. 1, p. 147470490900700, 2009. [Online]. Available: http://journals.sagepub.com/doi/10.1177/147470490900700107

[22] C. Darwin, "The expression of emotions in animals and man," *London: Murray*, vol. 11, 1872.

[23] J. Breuer and S. Freud, *Studies on hysteria*. Hachette UK, 2009.

[24] O. Hasson, "Emotional tears as biological signals," *Evolutionary Psychology*, vol. 7, no. 3, p. 147470490900700302, 2009.

[25] D. Küster, "Social effects of tears and small pupils are mediated by felt sadness: an evolutionary view," *Evolutionary Psychology*, vol. 16, no. 1, p. 1474704918761104, 2018.

[26] D. Küster, "Hidden tears and scrambled joy: On the adaptive costs of unguarded nonverbal social signals," in *Social Intelligence and Nonverbal Communication*, R. J. Sternberg and A. Kostić, Eds. Springer International Publishing, pp. 283–304. [Online]. Available: https://doi.org/10.1007/978-3-030-34964-6_10

[27] K. D. Elsbach and B. A. Bechky, "How Observers Assess Women Who Cry in Professional Work Contexts," *Academy of Management Discoveries*, vol. 4, no. 2, pp. 127–154, Jun. 2018. [Online]. Available: http://journals.aom.org/doi/10.5465/amd.2016.0025

[28] K. Ito, C. W. Ong, and R. Kitada, "Emotional tears communicate sadness but not excessive emotions without other contextual knowledge," *Frontiers in Psychology*, vol. 10, p. 878, 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2019.00878

[29] L. I. Reed, P. Deutchman, and K. L. Schmidt, "Effects of tearing on the perception of facial expressions of emotion," vol. 13, no. 4, p. 147470491561391, 2015. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1474704915613915

[30] C. W. Ong and K. Ito, "Can't fight seeing sadness in tears: Measuring the implicit association between tears and sadness," p. bjso.12503, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/bjso.12503

[31] A. Gračanin, A. J. J. M. Vingerhoets, I. Kardum, M. Zupčić, M. Šantek, and M. Šimić, "Why crying does and sometimes does not seem to alleviate mood: a quasi-experimental study," *Motivation and Emotion*, vol. 39, no. 6, pp. 953–960, Dec. 2015. [Online]. Available: https://doi.org/10.1007/s11031-015-9507-9

[32] D. DeVault *et al.*, "SimSensei kiosk: A virtual human interviewer for healthcare decision support," in *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*, ser. AAMAS '14. International Foundation for Autonomous Agents and Multiagent Systems, 2014, pp. 1061–1068. [Online]. Available: http://dl.acm.org/citation.cfm?id=2617388.2617415

[33] R. W. Picard, *Affective computing*. MIT press, 2000.

[34] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," *IEEE Transactions on Affective Computing*, pp. 1–1, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9039580/

[35] F. Noroozi, C. A. Corneanu, D. Kaminska, T. Sapinski, S. Escalera, and G. Anbarjafari, "Survey on Emotional Body Gesture Recognition," *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 505–523, Apr. 2021. [Online]. Available: https://ieeexplore.ieee.org/document/8493586/

[36] M. Pantic, N. Sebe, J. F. Cohn, and T. Huang, "Affective multimodal human-computer interaction," in *Proceedings of the 13th Annual ACM International Conference on Multimedia*, ser. MULTIMEDIA '05. New York, NY, USA: Association for Computing Machinery, 2005, p. 669–676. [Online]. Available: https://doi.org/10.1145/1101149.1101299

[37] M. Soleymani, M. Pantic, and T. Pun, "Multimodal Emotion Recognition in Response to Videos," *IEEE Transactions on Affective Computing*, vol. 3, no. 2, pp. 211–223, Apr. 2012. [Online]. Available: http://ieeexplore.ieee.org/document/6095505/

[38] F. Noroozi, M. Marjanovic, A. Njegus, S. Escalera, and G. Anbarjafari, "Audio-Visual Emotion Recognition in Video Clips," *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 60–75, Jan. 2019. [Online]. Available: https://ieeexplore.ieee.org/document/7945502/

[39] J. Chen, Z. Chen, Z. Chi, and H. Fu, "Facial Expression Recognition in Video with Multiple Feature Fusion," *IEEE Transactions on Affective Computing*, vol. 9, no. 1, pp. 38–50, Jan. 2018. [Online]. Available: https://ieeexplore.ieee.org/document/7518582/

[40] R. Subramanian, J. Wache, M. K. Abadi, R. L. Vieriu, S. Winkler, and N. Sebe, "ASCERTAIN: Emotion and Personality Recognition Using Commercial Sensors," *IEEE Transactions on Affective Computing*, vol. 9, no. 2, pp. 147–160, Apr. 2018. [Online]. Available: https://ieeexplore.ieee.org/document/7736040/

[41] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras, "AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups," *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 479–493, Apr. 2021. [Online]. Available: https://ieeexplore.ieee.org/document/8554112/

[42] R. Kosti, J. M. Alvarez, A. Recasens, and A. Lapedriza, "Context based emotion recognition using emotic dataset," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 11, pp. 2755–2766, 2020.

[43] A. Dhall, G. Sharma, R. Goecke, and T. Gedeon, "EmotiW 2020: Driver Gaze, Group Emotion, Student Engagement and Physiological Signal based Challenges," in *Proceedings of the 2020 International Conference on Multimodal Interaction*. Virtual Event Netherlands: ACM, Oct. 2020, pp. 784–789. [Online]. Available: https://dl.acm.org/doi/10.1145/3382507.3417973

[44] B. Schuller, M. Valstar, F. Eyben, G. McKeown, R. Cowie, and M. Pantic, "Avec 2011–the first international audio/visual emotion challenge," in *Affective Computing and Intelligent Interaction*, S. D'Mello, A. Graesser, B. Schuller, and J.-C. Martin, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 415–424.

[45] A. Kappas, E. Krumhuber, and D. Küster, "Facial behavior," in *In: Hall, Judith A.; Knapp, Mark L. (Ed.), Nonverbal communication (S. 131-166). Berlin: de Gruyter, 2013.* de Gruyter, 2013, pp. 131–166.

[46] D. Dupré, E. G. Krumhuber, D. Küster, and G. J. McKeown, "A performance comparison of eight commercially available automatic classifiers for facial affect recognition," *PLOS ONE*, vol. 15, no. 4, p. e0231968, Apr. 2020. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0231968

[47] E. G. Krumhuber, D. Küster, S. Namba, and L. Skora, "Human and machine validation of 14 databases of dynamic facial expressions," *Behavior research methods*, vol. 53, no. 2, pp. 686–701, 2021.

[48] D. Küster, E. G. Krumhuber, L. Steinert, A. Ahuja, M. Baker, and T. Schultz, "Opportunities and challenges for using automatic human affect analysis in consumer research," *Frontiers in*

Neuroscience, vol. 14, p. 400, Apr. 2020. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnins.2020.00400/full

[49] A. Kleinsmith and N. Bianchi-Berthouze, "Affective body expression perception and recognition: A survey," *IEEE Transactions on Affective Computing*, vol. 4, no. 1, pp. 15–33, 2012.

[50] H. Gunes and M. Piccardi, "Affect recognition from face and body: Early fusion vs. late fusion," in *2005 IEEE International Conference on Systems, Man and Cybernetics*, vol. 4. Waikoloa, HI, USA: IEEE, 2005, pp. 3437–3443. [Online]. Available: http://ieeexplore.ieee.org/document/1571679/

[51] ——, "Bi-modal emotion recognition from expressive face and body gestures," *Journal of Network and Computer Applications*, vol. 30, no. 4, pp. 1334–1345, 2007, special issue on Information technology. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1084804506000774

[52] L. Kessous, G. Castellano, and G. Caridakis, "Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis," *Journal on Multimodal User Interfaces*, vol. 3, no. 1, pp. 33–48, 2010.

[53] H. Znoj, "When remembering the lost spouse hurts too much: first results with a newly developed observer measure for tears and crying related coping behavior," in *The (non) expression of emotions in health and disease*, Vingerhoets, A.J.J.M., van Bussel, F.J., and Boelhouwer, A.J.W., Eds., 1997, pp. 337–352, publisher: Tilburg University Press Tilburg, Netherlands.

[54] T. Hollenstein and D. Lanteigne, "Models and methods of emotional concordance," *Biological Psychology*, vol. 98, pp. 1–5, Apr. 2014. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0301051113002597

[55] G. Marcus, "Deep Learning: A Critical Appraisal," *arXiv:1801.00631 [cs, stat]*, Jan. 2018. [Online]. Available: http://arxiv.org/abs/1801.00631

[56] M. Baker, "Blood, sweat and tears: The intra- and interindividual function of adult emotional weeping," 2019, unpublished doctoral thesis. [Online]. Available: https://researchportal.port.ac.uk/portal/files/26687007/Blood_sweat_and_tears_Final.pdf

[57] P. Ekman, W. V. Friesen, and J. C. Hager, *Facial action coding system (FACS)*, 2nd ed. Salt Lake City, Utah, USA: Research Nexus Division of Network Information Research Corporation, 2002.

[58] D. McDuff, A. Mahmoud, M. Mavadati, M. Amr, J. Turcot, and R. e. Kaliouby, "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*, ser. CHI EA '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 3723–3726. [Online]. Available: https://doi.org/10.1145/2851581.2890247

[59] G. Littlewort *et al.*, "The computer expression recognition toolbox (cert)," in *Face and gesture 2011*. IEEE, 2011, pp. 298–305.

[60] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, pp. 59–66.

[61] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh, "Openpose: Realtime multi-person 2d pose estimation using part affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[62] D. Dupré, E. G. Krumhuber, D. Küster, and G. J. McKeown, "A performance comparison of eight commercially available automatic classifiers for facial affect recognition," *Plos one*, vol. 15, no. 4, p. e0231968, 2020.

[63] E. G. Krumhuber, D. Küster, S. Namba, D. Shah, and M. G. Calvo, "Emotion recognition from posed and spontaneous dynamic expressions: Human observers versus machine analysis." *Emotion*, pp. 447–451, 2019.

[64] P. Dente, D. Küster, L. Skora, and E. Krumhuber, "Measures and metrics for automatic emotion classification via facet," in *Proceedings of the Conference on the Study of Artificial Intelligence and Simulation of Behaviour (AISB)*, 2017, pp. 160–163.

[65] G. C. Cawley and N. L. Talbot, "On over-fitting in model selection and subsequent selection bias in performance evaluation," *The Journal of Machine Learning Research*, vol. 11, pp. 2079–2107, 2010, publisher: JMLR. org.

[66] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why should I trust you?": Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: Association for Computing Machinery, Aug. 2016, pp. 1135–1144. [Online]. Available: https://doi.org/10.1145/2939672.2939778

[67] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable ai: A review of machine learning interpretability methods," *Entropy*, vol. 23, no. 1, 2021. [Online]. Available: https://www.mdpi.com/1099-4300/23/1/18

[68] D. Wang, Q. Yang, A. Abdul, and B. Y. Lim, "Designing Theory-Driven User-Centric Explainable AI," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow Scotland Uk: ACM, May 2019, pp. 1–15. [Online]. Available: https://dl.acm.org/doi/10.1145/3290605.3300831

[69] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[70] P. Ekman, "Basic emotions," in *Handbook of Cognition and Emotion*, T. Dalgleish and M. Power, Eds. John Wiley & Sons, 1999, pp. 45–60.

[71] C. Crivelli and A. J. Fridlund, "Facial displays are tools for social influence," *Trends in Cognitive Sciences*, vol. 22, no. 5, pp. 388–399, May 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1364661318300299

[72] A. S. Cowen, K. Manokara, X. Fang, D. Sauter, J. A. Brooks, and D. Keltner, "Facial movements have over twenty dimensions of perceived meaning that are only partially captured with traditional methods," Jun 2021. [Online]. Available: psyarxiv.com/hc93t

[73] E. G. Krumhuber and K. R. Scherer, "Affect bursts: Dynamic patterns of facial expression." vol. 11, no. 4, pp. 825–841. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0023856

[74] E. Krumhuber and A. Kappas, "More what duchenne smiles do, less what they express," *Perspectives on Psychological Science*, in press. [Online]. Available: https://osf.io/hqavb

[75] C.-W. Hsu, C.-C. Chang, C.-J. Lin, and others, "A practical guide to support vector classification."

[76] A. Barredo Arrieta *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 58, pp. 82–115, Jun. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253519308103

[77] G. Stratou *et al.*, "A demonstration of the perception system in SimSensei, a virtual human application for healthcare interviews," in *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2015, pp. 787–789. [Online]. Available: http://ieeexplore.ieee.org/document/7344661/

[78] M. de Gennaro, E. G. Krumhuber, and G. Lucas, "Effectiveness of an empathic chatbot in combating adverse effects of social exclusion on mood," *Frontiers in Psychology*, vol. 10, 2020. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2019.03061

[79] T. Schultz *et al.*, "I-CARE-an interaction system for the individual activation of people with dementia," *Geriatrics*, vol. 6, no. 2, p. 51, May 2021. [Online]. Available: https://doi.org/10.3390%2Fgeriatrics6020051

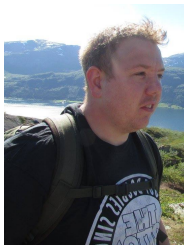
[80] L. Steinert, F. Putze, D. Küster, and T. Schultz, "Towards engagement recognition of people with dementia in care settings," in *Proceedings of the 2020 International Conference on Multimodal Interaction*, ser. ICMI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 558–565. [Online]. Available: https://doi.org/10.1145/3382507.3418856

[81] L. Steinert, F. Putze, D. Küster, and T. Schultz, "Audio-Visual Recognition of Emotional Engagement of People with Dementia," in *Proc. Interspeech 2021*, 2021, pp. 1024–1028.

[82] A. Jones *et al.*, "Empathic Robotic Tutors for Personalised Learning: A Multidisciplinary Approach," in *Social Robotics*, A. Tapus, E. André, J.-C. Martin, F. Ferland, and M. Ammi, Eds. Cham: Springer International Publishing, 2015, pp. 285–295.

Dennis Küster is a senior researcher at the University of Bremen, department of Computer Science. He received his PhD from Jacobs University Bremen, Germany, in 2008, where he studied the relationship between emotional experience and social context. His current research interests focus on emotions and tears, their measurement, elicitation, and functions - including interpersonal interactions mediated by avatars or robots.

Lars Steinert received the MSc degree in business administration from the University of Bremen, Germany, in 2018. He is working towards the PhD degree in the Cognitive Systems Lab at the University of Bremen, Germany, in the field of multi-modal affective computing. His research interests include machine learning and audio-visual recognition of affective states, with a focus on People with Dementia.



Marc Baker is a lecturer at the University of Portsmouth, United Kingdom. He obtained his PhD in psychology at the University of Portsmouth for his thesis on the intraindividual and interindividual functions of adult emotional tears. His research currently focuses on adult emotional weeping, facial thermography of categorical emotions, contextual effects of emotion perception, and individual differences in emotion induction. He has co-authored several papers on adult tears and thermal imaging.

Nikhil Bhardwaj received his BSc degree in computer science at the University of Bremen, Germany, in 2019 and is currently working towards the MSc degree in computer science at the University of Bremen, Germany. His research interests include machine learning, natural language processing and visual recognition of affective states.

Eva Krumhuber is associate professor in the Department of Experimental Psychology at University College London, with research interests in the domain of emotion and facial expression. She obtained her doctoral degree in Social Psychology from Cardiff University for which she won the Hadyn Ellis Prize for Outstanding Dissertation. Much of her work is concerned with empirical investigations into the socio-cognitive and affective processes of human perception and behaviour.