Ubiquitous Computing 2021

Akiyo Worou, Yun Cheng, Lorenz Frey
aworou@student.ethz.ch,chengyu@ethz.ch,lofrey@ethz.ch

1 INTRODUCTION

This paper proposes a random forest classifier to predict a person's emotional state, that is, the level of valence and arousal. The underlying data combines features extracted from sensing modalities such as electrocardiograms (ECG), electroencephalograms (EEG), electrodermal activities (EDA) and facial landmark trajectories (EMO). The ASCERTAIN dataset was used that collected data from 44 participants watching different affective movie clips. In a first step, the data was pre-processed and meaningful features were extracted. In a second step, a RF classifier was implemented and tested that predicts the emotional state (arousal and valence) of a participant using two different cross-validation approaches. Additionally, an evaluation of the most important features for each model was performed.

2 FEATURE EXTRACTION AND DATA ANALYSIS

2.1 ECG

We first plot the Power Spectral Density (PSD) for Lead I ECG signal of the first participant watching Clip 1 from 0 to 40 Hz in Figure 1

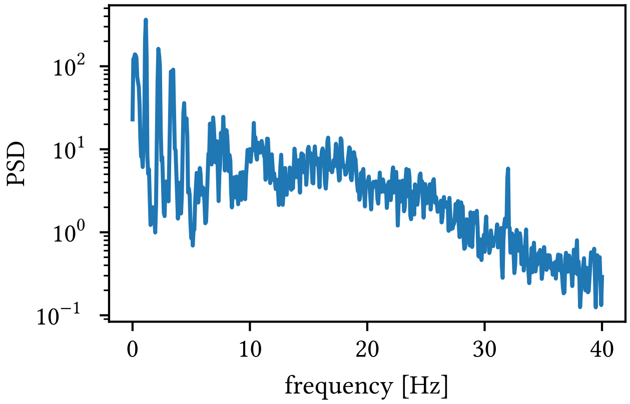


Figure 1: Power Spectral Density of ECG.

To remove noise using low-pass filter, a cutoff frequency between 20 and 30Hz is seen suitable in previous research [5]. Filtering the ECG-signal with different cutoff frequencies in the proposed range, the optimal value $f_{cutoff} = 25Hz$ was chosen by visual inspection. If the cutoff frequency is lower, the T-waves in the ECG signal get amplified considerably, which might hamper the r-peak detection in the subsequent task. On the other hand, if the cutoff is too high, the effect of noise suppression deteriorates. The frequency response of the low pass filter with order 5 is shown in figure 3. An additional notch filter with a narrow stopband around 0.05Hz is used to remove undesired baseline wander. A comparison between the filtered and the unfiltered ECG signal is given in figure 2.

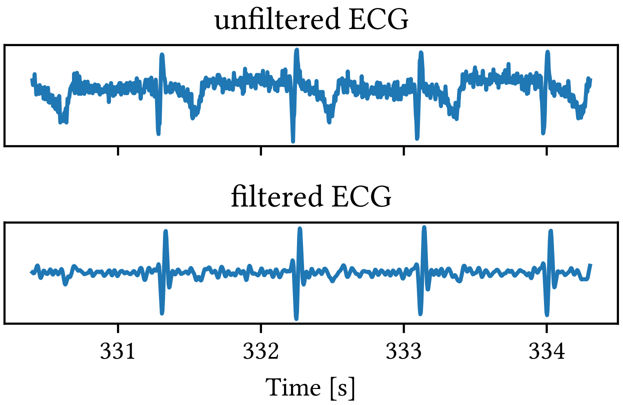


Figure 2: Unfiltered and filtered sequence of ECG-signal.

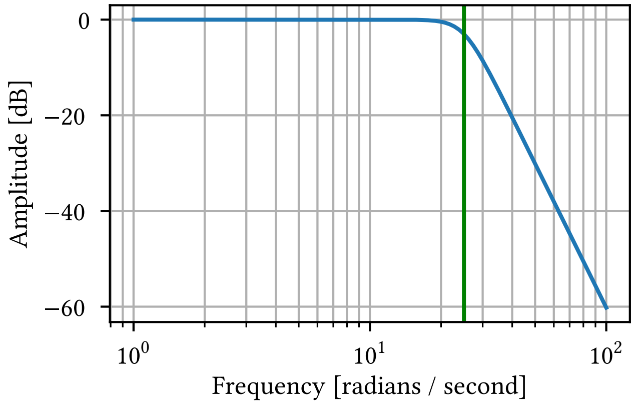


Figure 3: Frequency response of low-pass ECG-filter.

For the artifact detection algorithm, the criterion beat difference was computed based on the derivation in the lecture notes. Then, the algorithm proposed in [8] was used to detect invalid RR-intervals. Figure 4 depicts a signal containing artifacts (e.g. the central rr-interval that is roughly 25s long). The total percentage flagged as artifacts, i.e. the ratio between the accumulated time of rr-intervals flagged as artifacts divided by the total time of all signals (each clip, each participant), is 1.19%

Using the last 50 seconds of each recording, we process the data and generate all required features, see details in our attached notebook. To further improve the accuracy of the model, we propose

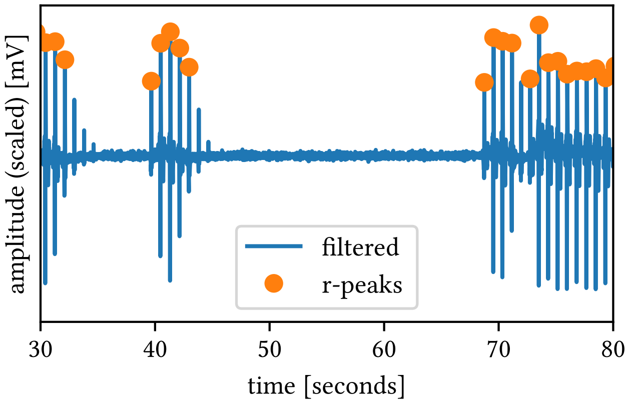


Figure 4: Sequence of filtered ECG signal containing artifacts.

5 more ECG features as follows: *(1) HRV-pNN20/50:* Percentage of successive RR intervals that differ by more than 20/50m, which is shown effectiveness in human emotion detection; *(2) HRV-SD1SD2:* The ratio between SD1 and SD2 of HRV, previous study [15] show this ratio contain meaningful information of ECG; *(3) HRV-RMSSE:* Root mean square of the successive differences in proved to be significant in emotion detection [7]; *(4) HRV-SamEn:* The sample entropy of ECG is also useful for emotion detection problems [12].

2.2 EMO

From the EMO signal we extracted the statistical measurements of 12 features that we used later in our work.

2.3 EDA

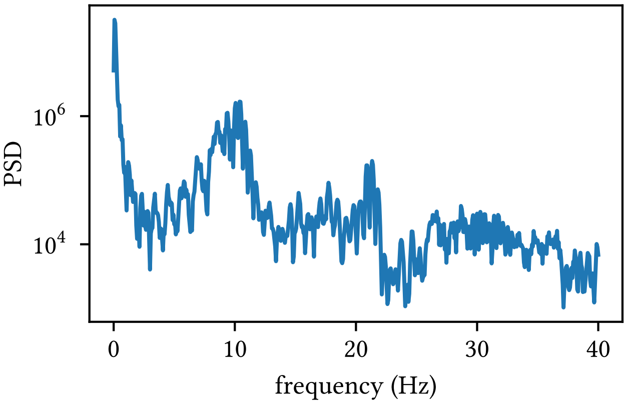


Figure 5: Power Spectral Density of EDA.

We plot the PSD for the EDA signal of the second participant watching Clip 1 in Figure 5. Previous work [3] shows that a cutoff of 1.99Hz is proposed with a filter of order 3 for the EDA analysis. This frequency is used in order to include the phasic and tonic components of the GSR signal as shown in Figure 6. Indeed, the tonic component represents the slow changing baseline levels and individual background characteristics (between 0Hz and 0.16Hz) and the phasic denotes the fast changing element that may be linked to an event (between 0.16Hz and 2.1Hz). One filtered EDA signal is shown in Figure 7.

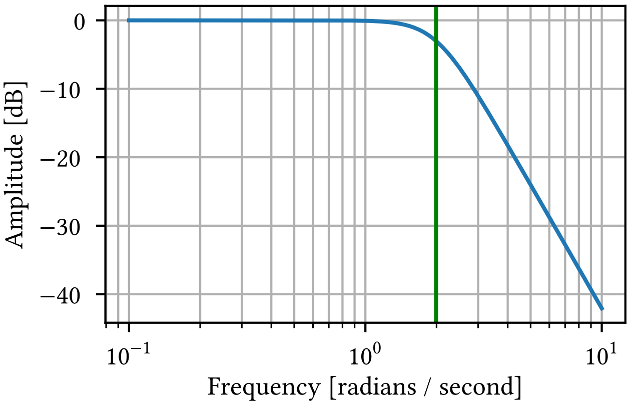


Figure 6: Frequency response of low-pass EDA-filter.

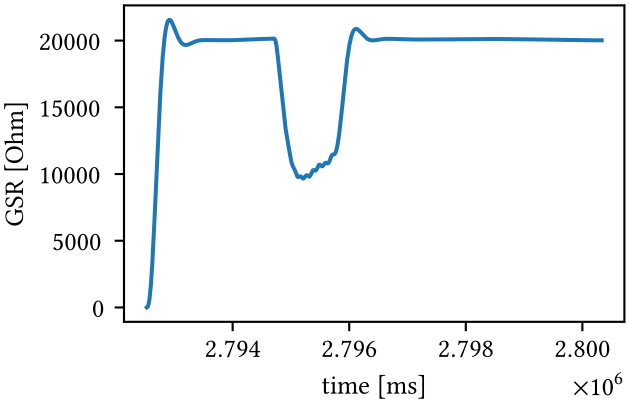


Figure 7: Filtered EDA-signal.

Using the last 50 seconds of each recording, we process the data and generate all required features, see details in our attached notebook. To further improve the accuracy of the model, we propose 5 more EDA features as follows (we use the mean of following features as our final features): *(1) SCR_Height:* the SCR amplitude of the signal including the Tonic component; *(2) SCR_Amplitude:* The SCR amplitude of the signal excluding the Tonic component; *(3) SCR_RiseTime:* the time taken for SCR onset to reach peak amplitude within the SCR; *(4) SCR_Recovery/time:* the samples at which(when) SCR peaks recover (decline) to half amplitude.

2.4 Valence and Arousal

The Pearson correlation between Valence and Arousal is approximately -0.02. This means that there is a very small or almost no correlation between Valence and Arousal. For a better analysis of the Valence and the Arousal we plot both distributions and the bivariate distribution as shown in Figure 8 . The Valence data is well balanced with 54% ratio for low values whereas the Arousal data is unbalenced with 34% for low values. The bivariate plot confirms the non correlation of the two variables. Indeed, for arousal = 5 we can have valence in {-2,2} almost uniformly. It is the same for arousal = 4 and valence in {-1,1}. It means that for a high or arousal the valence can be low or high. In fine, having an information about Valence does not tell us more about Arousal.

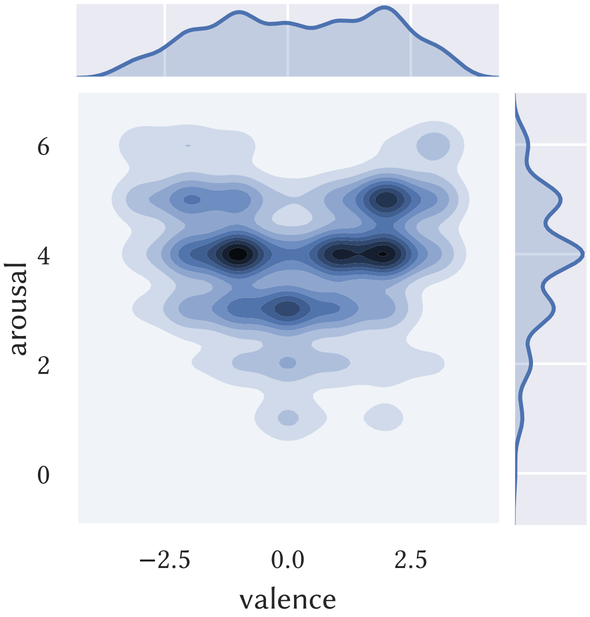


Figure 8: Valence and Arousal bivariate distribution

3 CLASSIFICATION

For the classification task, we have trained a basic random Forest to predict Valence and Arousal, using the features extracted before. For both Valence and Arousal we trained two random forests models using one-clip out validation and one-participant out validation, the results are shown in Table 1 and confusion matrix are plot in 9. We further implement ZeroR and report its performance in Table 1 as a comparison with our method.

With the random forest we can easily get the importance of each features used in our training. We decided to use our best models for both Valence and Arousal for this features selection. Indeed the models using one-participant out validation are the bests. We then easily get the 10 most import features for both Valence and Arousal best models (See details in notebook including the feature importance plot).

For Valence those features ordered by degree of importance are : *EEG feature 77, EEG feature 42, EEG feature 66, EEG feature 20, EEG feature 22, EEG feature 33, EEG feature 75, EEG feature 44, SCR_Risetime, SCR_Recovery and SCR_RecoveryTime* .

For Arousal the features nature are more diversified and we have the following features ordered by degree of importance : *EEG feature 55, One of the low frequency Power Spectral Density, the percentage of time with negative first derivative of the skin resistance, HRV_below, EEG feature 33, HR_below, EEG feature 64, HRV_SampleEn, SCR_Recovery and SCR_RiseTime*

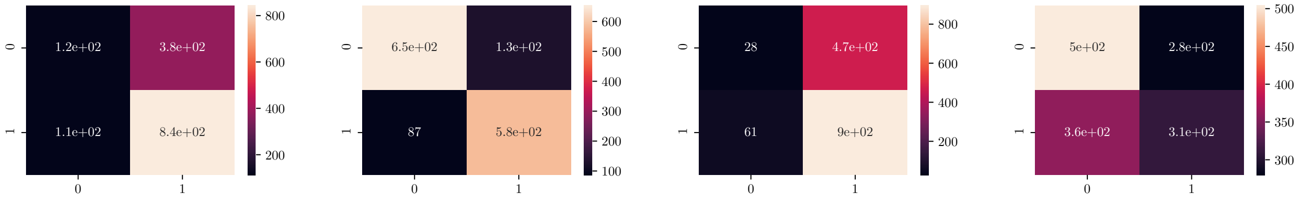
4 DISCUSSION

In order to check the impact of particular features and sensing modalities on our classification, the feature importance based on mean decrease in impurity was calculated for each of our models. For the valence models with leave-one-clip and leave-one-participant out, the EEG features are the most important ones, followed by a group of EDA features. Apparently, wearing sensors for EEG measurements is far from unobtrusive since it requires the subject to wear some kind of headgear. On the other hand, EDA features are better suited for unobtrusive and continuous tracking in real life. According to [6], the electrodes for the skin resistance measurements can be attached invisibly at the foot and can hence improve user acceptance. In the arousal model with leave-one-participant out, there is no clear order of the ten most important features across the different sensing modalities. The same holds true for the leave-one-clip out arousal model, where all sensing modalities contribute to a similar extent to the variance in the dataset.

To evaluate the generalization ability of our proposed classifier, we test the prediction scores mean and standard deviation from different leave-one-out cross validation schema, which are shown in Table 1. We can find that: *(i)* For both valence and arousal predictions, leave-participant-out scheme acquires better results in all evaluation metrics than leave-clip-out scheme. *(ii)* We also calculate the standard deviation of all evaluation metrics, for valence prediction, the standard deviation of leave-clip-out is bigger than the one of leave-participant-out, while for all other three metrics, leave-clip-out is better than leave-participant-out. For arousal prediction, the leave-clip-out standard deviation of accuracy and precision is bigger than the ones of leave-participant-out, and for other two metrics, leave-clip-out works better.

We use the best model found above to evaluate the performance for very high/low valence/arousal prediction. The results reveals that our model works better for very high/low prediction compared with non-very high/low ones for both valence and arousal (see notebook for details).

Two more additional sensing modalities are outlined to further improve classification accuracy as follows. **(i) Vocal characteristics:** Vocal characteristics: The voice is a valuable indicator of a subject's emotions. According to [2], the affective state of a subject can be identified with an average accuracy of $70 - 80\%$. Therefore, it is likely that the accuracy of the classifier can be improved if the dataset is complemented with meaningful features extracted from voice recordings such as rate, pitch average, pitch range, intensity, voice quality, pitch changes, articulation [2]. A possible limitation



(a) Valence, leave-participant-out **(b) Valence, leave-clip-out** **(c) Arousal, leave-participant-out** **(d) Arousal, leave-clip-out**

Figure 9: Confusion matrix

Model	Leava-X-out	Accuracy		Precision		Recall		F1-score	
		mean	*std*	*mean*	*std*	*mean*	*std*	*mean*	*std*
valence (ZeroR)	participant	0.54	-	0.27	-	0.50	-	0.34	-
	clip	0.53	-	0.26	-	0.50	-	0.30	-
valence (Ours)	participant	0.85	0.14	0.85	0.12	0.86	0.07	0.84	0.10
	clip	0.56	0.17	0.39	0.08	0.51	0.0	0.35	0.06
arousal (ZeroR)	participant	0.66	-	0.34	-	0.51	-	0.40	-
	clip	0.65	-	0.32	-	0.50	-	0.39	-
arousal (Ours)	participant	0.66	0.11	0.60	0.12	0.56	0.08	0.54	0.10
	clip	0.63	0.15	0.42	0.16	0.50	0.04	0.41	0.07

Table 1: model performance in cross-validation, comparison to ZeroR.

is the variety of vocal characteristics between gender, age, culture. **(ii) Eye tracking:** Diameter of the pupils as well as movements of the eye are indicators for emotions. In [1], a classification was performed to distinguish between three different classes of emotions, where a maximal accuracy of 80 % was achieved, based on eye tracking data. The eye tracking features might insofar be a valuable addition to the EMO data as the activation of eye movement or pupil width happens on a level of lower consciousness compared to the given EMO features, which makes those features less prone to conscious manipulation. However, lithing conditions can hamper the collection of meaningful data, especially for the pupil width.

REFERENCES

[1] Amparo Alonso-Betanzos, Paweł Tarnowski, Marcin Kołodziej, Andrzej Majkowski, and Remigiusz Jan Rak. 2020. Eye-Tracking Analysis for Emotion Recognition. *Computational Intelligence and Neuroscience* 2020 (2020), 2909267. https://doi.org/10.1155/2020/2909267

[2] DASC Alu, Elteto Zoltan, and Ioan Cristian Stoica. 2017. Voice based emotion recognition with convolutional neural networks for companion robots. *Science and Technology* 20, 3 (2017), 222–240.

[3] Giuseppina Schiavone Amit Acharyya Walter de Raedt Arvind Gautam, Neide Simoes-Capela and Chris Van Hoof. 2018. A Data Driven Empirical Iterative Algorithm for GSR Signal Pre-Processing. *IEEE Xplore* PP (12 2018), 1165. https://doi.org/10.23919/EUSIPCO.2018.8553191

[4] Gary G Berntson, Karen S Quigley, Jaye F Jang, and Sarah T Boysen. 1990. An approach to artifact identification: Application to heart period data. *Psychophysiology* 27, 5 (1990), 586–598.

[5] Ivaylo Christov, Tatyana Neycheva, Ramun Schmid, Todor Stoyanov, and Roger Abächerli. 2017. Pseudo-real-time low-pass filter in ECG, self-adjustable to the frequency spectra of the waves. *Medical & biological engineering & computing* 55, 9 (2017), 1579–1588.

[6] Franz Gravenhorst, Bernd Tessendorf, Amir Muaremi, Cornelia Kappeler-Setz, Bert Arnrich, and Gerhard Tröster. 2012. Mobile System for Unobtrusive Monitoring of Electrodermal Activity in Daily Life.

[7] Ross Harper and Joshua Southern. 2019. End-To-End Prediction of Emotion From Heartbeat Data Collected by a Consumer Fitness Tracker. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 1–7.

[8] Karen Hovsepian, Mustafa Al'Absi, Emre Ertin, Thomas Kamarck, Motohiro Nakajima, and Santosh Kumar. 2015. cStress: towards a gold standard for continuous stress assessment in the mobile environment. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. 493–504.

[9] Mojtaba Khomami Abadi, Ramanathan Subramanian, Seyed Mostafa Kia, Paolo Avesani, Ioannis Patras, and Nicu Sebe. 2015. DECAF: MEG-based Multimodal Database for Decoding Affective Physiological Responses". *IEEE Transactions on Affective Computing* PP (01 2015), 1. https://doi.org/10.1109/TAFFC.2015.2392932

[10] Jonghwa Kim and Elisabeth André. 2008. Emotion recognition based on physiological changes in music listening. *IEEE transactions on pattern analysis and machine intelligence* 30, 12 (2008), 2067–2083.

[11] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. 2011. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing* 3, 1 (2011), 18–31.

[12] Dae-Young Lee and Young-Seok Choi. 2020. Multiscale distribution entropy analysis of heart rate variability using differential inter-beat intervals. *IEEE Access* 8 (2020), 48761–48773.

[13] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. 2011. A multimodal database for affect recognition and implicit tagging. *IEEE transactions on affective computing* 3, 1 (2011), 42–55.

[14] Ramanathan Subramanian, Julia Wache, Mojtaba Khomami Abadi, Radu L. Vieriu, Stefan Winkler, and Nicu Sebe. 2018. ASCERTAIN: Emotion and Personality Recognition Using Commercial Sensors. *IEEE Transactions on Affective Computing* 9, 2 (2018), 147–160. https://doi.org/10.1109/TAFFC.2016.2625250

[15] Cheng Xiefeng, Yue Wang, Shicheng Dai, Pengjun Zhao, and Qifa Liu. 2019. Heart sound signals can be used for emotion recognition. *Scientific reports* 9, 1 (2019), 1–11.