Person Identification based on Physiological Signals: Conditions and Risks

Peter Bellmann[1], Patrick Thiam[1,2] and Friedhelm Schwenker[1]
[1]*Institute of Neural Information Processing, Ulm University, James-Franck-Ring, 89081 Ulm, Germany*
[2]*Institute of Medical Systems Biology, Ulm University, Albert-Einstein-Allee 11, 89081 Ulm, Germany*

Keywords: Biopotentials, Person Identification, Decision Trees.

Abstract: Person identification is usually based on video signals, DNA samples or fingerprints. In this study, we want to show the effectiveness of other physiological signals for person identification. For this purpose, we evaluate different settings with the SenseEmotion Database. The data set was initially collected for research purposes in the fields of emotion and pain intensity recognition. However, we use the multi-modality of this database to evaluate the effectiveness of different physiological signals, such as the heart activity or skin conductance, for person identification purposes. It is almost impossible for human beings to identify persons by evaluating a set of different fingerprints. Machine learning methods usually outperform humans in both, operation time as well as accuracy, in those tasks. In our study, we show that basic pattern recognition models can be used to identify human beings based on physiological signals. However, our outcomes show that person identification based on physiological signals must be treated with caution. Specifically, our results indicate that it is essential to include physiological signals from different recording sessions, to ensure generalisation ability of the classification model, for the person identification task.

1 INTRODUCTION

Person identification is an everyday phenomenon. The human brain is trained to easily recognise people on audiovisual basis. Old science fiction movies showed security mechanisms, such as retinal scans, as novel ideas (e.g. in Demolition Man[1], in 1993). For many years, fingerprints have been used in criminology. Nowadays, person identification based on fingerprints is a basic feature of many smartphones.

It is common to use video signals for person identification purposes. However, the trained machine learning model has to be able to overcome different circumstances, such as different lighting conditions, persons' movements and others, such as changing glasses for contact lenses.

Therefore, we want to evaluate the task of person identification by considering physiological signals. The measurement of physiological signals, such as the heart rate or muscle activity, is not affected by the aforementioned circumstances, which affect models trained on video signals.

However, one needs adequate sensors in order to record physiological data. Moreover, those sensors have to be attached to the human body. Besides, there are other factors affecting biopotentials, such as an individual's physiological state (e.g. sleeping vs. performing sports activities). However, those factors are not present in the data set, which we are using in the current study.

The remainder of this study is organised as follows. In Section 2, we motivate our idea for choosing the SenseEmotion Database, specifically for person identification purposes, which was initially recorded for the research in the fields of emotion recognition and pain intensity classification. Section 3 provides a description of the SenseEmotion Database. Section 4 is an overview of all experimental settings that are applied in this study. In Section 5, we state and discuss our outcomes on person identification based on physiological signals, with comparison to the outcomes based on video signals. We change our evaluation protocol in Section 6, to show the shortcomings of physiological signals based person identification (PSbPI). Based on those outcomes, we provide a guide for the data recording and classification model design, for the PSbPI task. Finally, in Section 7, we conclude this study.

[1]https://www.imdb.com/title/tt0106697/?ref_=fn_al_tt_1

Bellmann, P., Thiam, P. and Schwenker, F.
Person Identification based on Physiological Signals: Conditions and Risks.
DOI: 10.5220/0008865503730380
In *Proceedings of the 9th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2020)*, pages 373-380
ISBN: 978-989-758-397-1; ISSN: 2184-4313

2 MOTIVATION & RELATED WORK

As we will shortly explain in Section 3, the SenseEmotion Database was initially collected for emotion recognition and pain intensity classification purposes. In previous studies based on that data set (see Sec. 3.2), we found that the pain intensity classification task is highly affected by individual characteristics of the data. Particularly, the leave-one-participant-out pain intensity classification task is challenging. By changing the task to person identification, we are able to analyse the effectiveness of EDA and RSP signals, in such scenarios. Moreover, in our main experiments, in Section 6, we will analyse the effects of including different recording sessions, in person identification tasks. In particular, we will focus on the real application scenario, in which a whole recording session is not seen during the training phase. Thereby, we will analyse the classification performance based on physiological signals in comparison to non-physiological signals.

To apply a person identification task without any dependency on experienced pain, we focus on the samples resulting from the pain-free stimuli of $32°C$. These samples are intended to represent the participants' pain-free states. Anyway, it has to be mentioned that during those phases were emotional stimuli (negative, neutral and positive). The data subsets (recordings for the left and right forearms, respectively), which we analyse in this study consist of 1200 samples each (30 samples per participant).

There exist different works in the literature, for person identification based on biopotentials. Chan et al. introduced the wavelet distance (WDIST) measure for ECG based person identification (Chan et al., 2008). Their WDIST measure outperformed the so-called *percent residual difference* and the *correlation coefficient* measures in an evaluation of a set of 50 subjects. Thereby, each subject participated in three recording sessions in a non-clinical setting, in which the participants simply held two electrodes using their thumbs and index fingers.

Suresh et al. provided the first EMG based person identification approach evaluated on a set of 49 subjects (Suresh et al., 2011). The subjects participated in three recording sessions, in which they performed wrist motions of 10 seconds, several times each. The authors evaluated the so-called *vector quantization* (Linde et al., 1980) and *Gaussian mixture model* (GMM) approaches (Dempster et al., 1977) leading to the preference of GMM models.

In the late nineties, Poulus et al. published several studies related to EEG based person identification (Poulos et al., 1999; Poulos et al., 1999b; Poulos et al., 1999a), being the first to apply parametric spectral analysis of the EEG signals. In their works, Poulus et al. used a subset of a data set containing continuous EEG recordings of 79 individuals of three minutes each, for different analyses of the so-called *alpha rhythm spectral band*. In (Poulos et al., 1999), the authors followed a non-parametric approach for feature augmentation by extracting *spectral values* from the Fast Fourier Transform (FFT) of the EEG signal. The classification was undertaken by Kohonen's Linear Vector Quantizer (LVQ) (Kohonen, 1989). In (Poulos et al., 1999b), the authors used the LVQ classifier in combination with a parametric spectral analysis by fitting a linear all-pole (AR) model to the EEG spectrum. The coefficients of the AR model were used as features. In (Poulos et al., 1999a), the authors followed the aforementioned parametric approach in combination with a different classifier, using so-called *characteristic convex polygon* models (O'Rourke et al., 1982).

3 SENSEEMOTION DATABASE

The SenseEmotion Database (SEDB) was collected at Ulm University for research purposes in the field of emotion and pain (intensity) recognition (Velana et al., 2016). Forty-five healthy subjects participated in the experiments. Due to missing or erroneous data, five participants were excluded from the data set. The current study is based on the recordings specific to the remaining 40 participants (20 female and 20 male).

3.1 Data Set Description

Pain was induced in form of heat by a Medoc thermode[2], which was placed at the participant's forearm. The pain-free temperature was set to $32°C$ for each participant.

Heat Stimuli Sequences. After an individual calibration phase, which led to three equidistant pain temperature levels, each of the participants was stimulated thirty times with each of the four temperature levels (pain-free, pain, intermediate and tolerance level). The order of the pain stimuli was randomised. Each of the pain levels was held for four seconds. After each pain stimulation, each participant was stimulated with the pain-free level with a random duration length of eight to twelve seconds. The experiments were conducted twice. Once, the heat elicitation thermode was attached to the participant's left forearm,

[2]https://medoc-web.com/products/pathway/

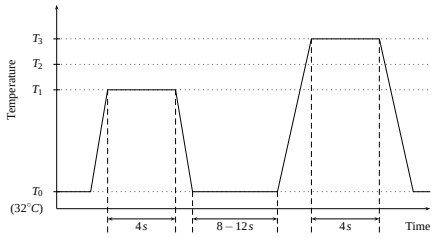

Figure 1: An example for a participant's sequence of different stimuli. In this study, we focus on the T_0-related stimuli.

and once it was attached to the participant's right forearm. Therefore, the SEDB consists of two subsets, which we will simply denote by *left subset* and *right subset*, according to the placement of the thermode. Note that for each participant, the data specific to each forearm were recorded on different days, i.e. during two different sessions. This is especially important for the main outcome of this study, as later discussed in Sections 6 and 7.

Figure 1 illustrates an example for a participant's heat stimuli sequence. The recorded signals can be categorised into the following three groups, which we simply call *modalities*.

Biopotentials (BIO). The recordings of the physiological modality consists of four channels, i.e. electrocardiogram (ECG), electrodermal activity (EDA), electromyogram (EMG) and respiration (RSP). ECG measures heart activity. From the ECG signal, one can extract different kind of information, such as the heart beat interval or the heart rate. EDA measures the skin conductance. The EDA sensors were placed at the index and ring fingers. EMG measures muscle activity. In the experiments, the activity of the trapezius muscle (located in the upper back area of the human body) was recorded. An elastic belt system was used to record the breathing activity (respiration).

Videos (VID). A synchronised camera system was installed to record the participants from three different angles. Between each of the three cameras and the participants was a distance of approximately one meter. One frontal camera was placed towards the participants, and two cameras were placed in an angle of approximately 45° to the left and right side, respectively (see Figure 2).

Audio Signals (AUD). Similar to the videos, the audio signals were recorded synchronously through three different sources. A digital wireless headset microphone, a directional microphone and the Microsoft Kinect v2 integrated microphone, recorded the signals.

Figure 3 shows the experimental settings, including examples of extracted facial areas, which were used to compute video features. For more details concerning the SEDB, we refer the reader to (Velana et al., 2016).

Feature Extraction. Feature extraction is beyond the scope of this study. Therefore, we refer the reader to one of our latest works (Thiam et al., 2019b), for complete details on the extraction and normalisation of features. The video features were extracted from windows of length 6.5 sec, for each sample. For the video signals, we extracted three types of features, i.e. geometric features (GEO), head pose features (HPO) and local binary patterns from three orthogonal planes (LBP-TOP). We will denote the LBP-TOP features simply by LBP, for better readability of our tables. Note that all video features were extracted solely from the participants' faces (see Figure 3, bottom). The audio and physiological features were extracted from the same windows but of length 4.5 sec, for each sample. This was especially important for the samples specific to the temperature levels T_1, T_2 and T_3, which we do not consider here. After applying different signal detrending and smoothing techniques to reduce noise and artefacts in the physiological signals, different statistical descriptors, such as *mean*, *standard deviation*, *extreme values*, were extracted, amongst others, from the temporal domain. From the frequency domain, additional features, including amongst others, *bandwidth*, *central* and *mean frequency*, were extracted. From the audio signals, different low-level, as well as, high-level descriptors have been extracted. Commonly used audio low-level descriptors are *Mel Frequency Cepstral Coefficients* (MFCCs) (Davis and Mermelstein, 1980) and features computed by applying the *Relative Spectral Perceptual Linear Predictive Coding* (RASTA-PLP) (Hermansky et al., 1992), which is an extension of *Perceptual Linear Predictive* (PLP) analysis (Hermansky, 1990). Table 1 summarises the feature dimensions for all available channels. Biopotentials are defined by 307 features in total, followed by 980 and 3126 audio and video features, respectively. In (Thiam et al., 2019a), we propose using deep neural networks for autonomous feature learning based on ECG, EDA and EMG signals, on a similar data set.

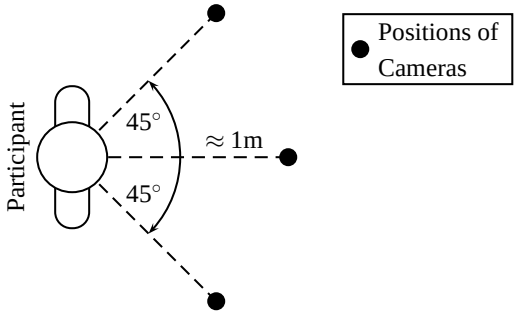

Figure 2: Sketch of the synchronised camera system setup.

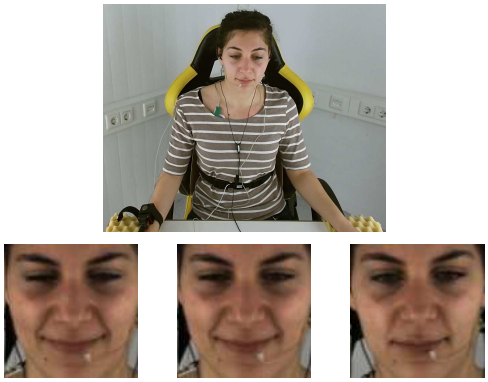

Figure 3: **Top**: Experimental settings. Each participant remained seated with both forearms resting on a desk, throughout the recording sessions. **Bottom**: Examples of extracted facial area used to compute video features.

Table 1: Number of extracted features grouped by biopotentials and video/audio features. GEO, HPO and LBP are features extracted from video signals, based solely on each participant's face and orientation of the head.

BIO	ECG	EDA	EMG	RSP
	115	72	61	59
VID/AUD	GEO	HPO	LBP	AUD
	714	252	2160	980

3.2 Related Work on the SEDB

Based on the collected data, Kessler et al. show the effectiveness of including camera photoplethysmography for pain recognition (Kessler et al., 2017a; Kessler et al., 2017b). Thiam et al. analyse the combination of audio and video channels (Thiam et al., 2017), as well as different multi-modal data fusion approaches, including biopotentials (Thiam and Schwenker, 2017), for pain intensity recognition. Different decision tree based classification ensembles are evaluated in (Bellmann et al., 2018). In one of our latest studies on the SEDB (Bellmann et al., 2019), we introduce an unsupervised data transformation, which improves the accuracy of nearest neighbour classifiers.

From the mentioned previous works, we can conclude that pain intensity recognition works well on the SEDB for a late fusion architecture (Snoek et al., 2005) with random forests (Breiman, 2001), which are combined by a pseudo-inverse aggregation layer (Penrose, 1955; Schwenker et al., 2006).

4 EXPERIMENTAL SETTINGS

This section gives a short overview of the experimental settings, which we apply throughout this work.

4.1 Definitions of Different Tasks

In this study, we consider two different person identification tasks. Those tasks, as well as the applied cross validation approach and the choice of a significance test, are explained in the following.

Binary Classification Task. In the binary classification task, we consider two participants at once. We refer to this task as the *pairwise* task. This scenario is unrealistic and *simple*. Its purpose is primarily to show that it is valid using the SEDB for person identification analyses (Poor classification performance in combination with this task would question the current study).

Multi-class Task. In the multi-class task, we consider all of the participants at once.

Cross Validation. For both tasks, we apply a **l**eave-**o**ne-**s**ample-**o**ut (LOSO) cross validation. Therefore, in the pairwise task, we apply the LOSO cross validation for each possible pair of participants (participant 1 vs. participant 2, participant 1 vs. participant 3, etc.). The final results are then averaged by the number of possible participant pairs, i.e. $40 \cdot 39/2 = 780$.

Significance Tests. In this work, we apply the two-sided Wilcoxon signed-rank test (Wilcoxon, 1945) to test for significant differences in accuracy, at a significance level of $p = 0.05$.

4.2 Classifiers & Performance Measure

In this study, we choose one single type of classification models and one single performance measure. Note that the focus of this study is to emphasise the importance of providing several recording sessions during the data acquisition, which is the first phase of each pattern recognition task.

Classification Models. For both of our classification tasks, we first use one single decision tree (DT) classifier (Breiman et al., 1984) applying the built-in implementation in MATLAB[3], with default parameter settings without optimisation. The Gini index is used as impurity measure. There is no threshold for the maximum tree depth and no pruning of the decision tree. The reasons for the choice of one single DT are twofold. Being a *weak* classifier, we suppose that applying one single DT will lead to significantly different classification performance values across the recorded modalities and channels. In that way, we can sort out some of the modalities and channels, based on initial experiments, as shown later in Section 5. On the other hand, achieving promising classification performance values by the use of one single DT in our

[3]https://www.mathworks.com/products/matlab.html

initial experiments justifies choosing random forests as *strong* classification models, in our main experiments, in Section 6.

Performance Measure. Both of our defined tasks constitute equally distributed classification tasks. In combination with the LOSO cross validation, we measure the performance by the ratio of the number of correct classified samples to the total number of samples, i.e.

$$accuracy = \frac{|\{x \in X : t(x) = l(x)\}|}{|X|}.$$

Thereby $X \subset \mathbb{R}^d, d \in \mathbb{N}$, denotes the current data set. The true label of x is denoted by $l(x)$ (participant ID). Moreover, by $t(x)$ we denote the output of the decision tree/random forest t (which is trained on $X \setminus \{x\}$, in the LOSO cross validation). Solely for the pairwise task, we provide additional values for the standard deviation, arising from the different accuracies across all possible pairs of participants.

5 RESULTS FOR EACH SUBSET

In this section, we provide the results for both settings, i.e. the pairwise and the multi-class setting. First, we evaluate the performance specific to each modality and channel. Subsequently, we evaluate the performance specific to a bio-visual combination.

5.1 Evaluation of Modalities

First, we evaluate the performance of each modality, i.e. BIO, VID and AUD. Table 2 depicts the results for the pairwise setting, whereas Table 3 states the results for the multi-class setting.

Table 2: Pairwise Task. Averaged accuracies and standard deviations in %. The best performing modality is underlined. Chance level accuracy is at 50%. BIO outperforms the other modalities significantly, according to a two-sided Wilcoxon signed-rank test with $p = 0.05$.

Modality	Left Subset	Right Subset
BIO	$\underline{98.91 \pm 1.91}$	$\underline{99.29 \pm 1.40}$
VID	97.76 ± 2.37	97.71 ± 2.63
AUD	83.10 ± 8.73	83.72 ± 8.36

Table 3: Multi-Class Task. LOSO accuracies in %. The best performing modality is underlined. Chance level accuracy is at 2.5%. BIO outperforms the other modalities significantly, according to a two-sided Wilcoxon signed-rank test with $p = 0.05$.

Modality	Left Subset	Right Subset
BIO	$\underline{97.46}$	$\underline{98.21}$
VID	87.02	82.81
AUD	35.03	35.50

The accuracy values from Table 2 are higher than those ones stated in Table 3. The pairwise setting constitutes a binary classification task, whereas the multi-class setting constitutes a 40-class classification task. AUD is the worst performing modality, significantly worse than the other modalities. This is exactly what we expected, since there were no verbal interactions during the experiments. The only available audio signals are moaning or (heavy) breathing noises. For both tasks, the biopotentials lead to the best results.

5.2 Evaluation of Channels

In this section, we evaluate the performance of each single channel. Table 4 states the results for the pairwise task.

Channels EMG, ECG and RSP lead to high performance values, with RSP being the best performing channel. EDA is the worst performing physiological channel. The best performing video channel is LBP. Video channels GEO and HPO perform even worse than the audio channel (see Table 2). Therefore, for the video modality, only the LBP features are considered for the rest of this study. Table 5 depicts the channel results for the multi-class task.

The video features (LBP) are significantly outper-

Table 4: Pairwise Task. Averaged accuracies and standard deviations in %. The best performing channel is underlined. Chance level accuracy is at 50%. The horizontal line separates physiological features from video-based features.

Channel	Left Subset	Right Subset
EMG	98.50 ± 2.04	98.60 ± 2.02
ECG	98.64 ± 1.98	99.04 ± 1.59
RSP	$\underline{98.91 \pm 1.82}$	$\underline{99.24 \pm 1.46}$
EDA	97.53 ± 2.72	97.63 ± 2.74
GEO	79.11 ± 10.5	79.54 ± 10.6
HPO	77.23 ± 10.5	78.66 ± 11.0
LBP	97.76 ± 2.37	97.75 ± 2.56

Table 5: Multi-Class Task. LOSO accuracies in %. The best performing channel is underlined. The horizontal line separates physiological features from video-based features. Each of the physiological channels outperforms the LBP channel significantly, according to a two-sided Wilcoxon signed-rank test with $p = 0.05$.

Channel	Left Subset	Right Subset
EMG	$\underline{98.13}$	97.53
ECG	97.54	$\underline{98.30}$
RSP	96.69	98.04
EDA	96.44	96.77
LBP	87.87	84.68

formed by the physiological channels. While RSP seems to perform best for the binary task, EMG leads to the best results for the left data subset, whereas ECG leads to the best results for the right data subset. EDA stays the worst performing biosignal, for the multi-class task. Thermal stimuli make the EDA sensor more unreliable, due to sweating, since EDA measures skin conductance. We consider recordings specific to heat stimuli with $32^{\circ}C$, in this work. However, each of those stimuli was directly following a heat stimulus with a higher temperature (see Sec. 3). Heat implies participant's perspiration, which directly affects the EDA signals and hence complicates the task of person identification.

5.3 Evaluation of Combined Modalities

In this part, we combine all physiological channels with the best video channel, i.e. LBP. Table 6 shows the results for both tasks, the pairwise and the multi-class tasks.
From Table 6, we can conclude that combining the biopotentials with video signals leads at most to the same performance, which is based solely on the biopotentials. Therefore, the performance based on the physiological channels is not improved by the addition of video signals, especially in the multi-class task (see Table 6, bottom part).

5.4 Discussion of initial Experiments

The physiological signal based accuracy values, which are reported in Tables 2, 3, 4, 5 and 6 seem to be surprisingly good, since we used only one DT as classification model. However, we applied the LOSO evaluation protocol, which provides an optimistic classification performance approximation. It is most likely that the biopotentials significantly outperform the video based classification due to the number of extracted features. While the physiological channels consist of 307 features (115 ECG, 72 EDA, 61 EMG, 59 RSP) in total, the LBP-TOP channel has a dimensionality of 2160. Therefore, an unpruned DT is more likely to overfit the data specific to the video based channel, during the training phase.

Table 6: Averaged/LOSO accuracies in %. The horizontal line separates the pairwise task (top) from the multi-class task (bottom). The standard deviation values for the pairwise task are left out, for better readability.

Modality	Left Subset	Right Subset
BIO	98.91	99.29
BIO-LBP	98.91	99.29
BIO	97.46	98.21
BIO-LBP	94.15	95.50

6 RESULTS ON MIXED SUBSETS

In the previous section, we considered the two subsets of the SEDB as two separate data sets. Since exactly the same subjects participated in the data acquisition experiments, we now use both subsets at once, within the following two settings. We train our classification model on the left subset and test it on the right subset, and vice versa.

6.1 Evaluation of the multi-Class Task

In this section, we do not apply any cross validation. The accumulated accuracy values arise from one single testing iteration. Moreover, instead of using one single decision tree, we now use random forests with 300, 500 and 1000 decision trees, respectively. Table 7 states the results with the left subset defined as the training data, and the right subset defined as the test data. In contrast, Table 8 depicts the results with the right subset as the training data, and the left subset as the test data.

6.2 Discussion

The results, which are stated in Tables 5, 7 and 8, as well as results stemming from the same experimental settings (training on the left subset and testing on the right subset, and vice versa) specific to the pain intensity classification task, lead to the following conclusions.

Table 7: Accuracies in %. Training data: left subset. Test data: right subset. The best performing channel is underlined. Chance level accuracy is at 2.5%. The horizontal line separates physiological features from video-based features. L: Number of decision trees in the random forest ensemble.

Channel	$L = 300$	$L = 500$	$L = 1000$
EMG	9.96	10.55	9.79
ECG	15.66	15.15	16.43
RSP	7.83	8.51	8.26
EDA	5.53	7.40	6.21
LBP	<u>97.19</u>	<u>96.94</u>	<u>97.36</u>

Table 8: Accuracies in %. Training data: right subset. Test data: left subset. The best performing channel is underlined. Chance level accuracy is at 2.5%. The horizontal line separates physiological features from video-based features. L: Number of decision trees in the random forest ensemble.

Channel	$L = 300$	$L = 500$	$L = 1000$
EMG	10.18	11.45	11.03
ECG	14.84	15.44	15.18
RSP	7.29	8.91	8.40
EDA	5.60	5.60	6.02
LBP	<u>99.07</u>	<u>99.24</u>	<u>99.15</u>

Pain Intensity Classification. In one of our previous works, we applied the same *transfer* of the SEDB subsets for the pain intensity classification task (Thiam et al., 2019b), i.e. we used the left subset as the training set and the right subset as the test set, and vice versa. The results in the aforementioned study show that there was no significant difference in performance compared to the task, in which each of the SEDB data subsets was analysed separately (Note that we do not provide accuracy values here, because we illustrated the results for the transfer task in our previous work solely as box plots).
Moreover, all afore-mentioned related works on the SEDB show that the best signal for pain intensity recognition is the EDA channel. On the other hand, our study shows that EDA is the worst performing physiological channel, worse than ECG, EMG and RSP, for the person identification task. This is, most likely due to the fact that, the participants were stimulated with a heat thermode, which is of course an expected cause for perspiration, leading to less reliable EDA data for the person identification task.

LBP Features based Person Identification. The results, which are stated in Tables 7 and 8 show that LBP-TOP features perform well in person identification tasks based on video sequences. The accuracies lie all above 96% in a setting with forty participants, when applying random forests instead of one single decision tree.

Biopotentials-based Person Identification. The results, which are depicted in Tables 7 and 8 show that the transfer for the person identification task based on physiological channels is not straightforward. The performance drops dramatically. The accuracy values range approximately between two times chance level (EDA) and six times chance level (ECG), reaching a maximum of only 16.43% (see Table 7, right column). What does that mean for physiological signals based person identification (PSbPI): does it or does it not work? If we had concluded this paper right after Section 5, we would certainly say that it works well. However, our results in the current section show that one has to be careful when drawing conclusions. The participants are well identified within one recording session. However, when different recording sessions are involved, the performance drops dramatically. The performance is affected by the psychological and physiological state of each participant during the session.
Therefore, in the PSbPI task, one should record different sessions for each participant and evaluate the sessions transfer performance to get a realistic generalisation estimation. To design a strong classification model, which is not overfitted to one single recording session, one should train the classification model on all available recording sessions. This is especially important when one has to identify participants from new (unseen) recording sessions.

7 CONCLUSION

The results of our work show that person identification based on physiological signals, i.e. electrocardiogram (ECG), electrodermal activity (EDA), electromyogram (EMG) and respiration (RSP), can outperform person recognition based on audio and video signals. This is especially the case when first, the data is recorded within one session, and second, a *weak* classification model is designed (we used one single decision tree in that part of the experiments).
In addition, our findings show that including data samples from different recordings constitutes a challenging task for physiological signals based person identification. We considered the task, in which only data specific to one of the recording sessions was known and used, to train the classification model. While the classification models performed well in combination with the extracted video features, we noted a dramatic drop in classification performance for all physiological channels.
This is an interesting observation, which leads to the main conclusion of this study. In order to build a reliable classification model, which is trained on physiological signals for the person identification task, one has to record different sessions for each participant. In general, to be able to provide appropriate research analyses of physiological signals, independently from the classification task (pain level or emotion recognition, person identification, etc.), one should include several recording sessions for each test subject. In real-world applications, the designed model should be trained on all available recording sessions, to learn as much as possible of each participant's variety of psychological and physiological states.

ACKNOWLEDGEMENTS

During the research, Peter Bellmann was supported by a scholarship of the Landesgraduiertenförderung Baden-Württemberg at Ulm University.
The research leading to these results has also received fundings from the Federal Ministry of Education and Research (BMBF project: SenseEmotion).
We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Tesla K40 GPU used for this research.

REFERENCES

Bellmann, P., Thiam, P., and Schwenker, F. (2018). *Multi-classifier-Systems: Architectures, Algorithms and Applications*, pages 83–113. Springer International Publishing, Cham.

Bellmann, P., Thiam, P., and Schwenker, F. (2019). Using a quartile-based data transformation for pain intensity classification based on the senseemotion database. In *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pages 310–316.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth.

Chan, A. D. C., Hamdy, M. M., Badre, A., and Badee, V. (2008). Wavelet distance measure for person identification using electrocardiograms. *IEEE Trans. Instrumentation and Measurement*, 57(2):248–253.

Davis, S. B. and Mermelstein, P. (1980). Comparison of parametric representation for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics Speech and Signal Processing*, 28(4):357–366.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

Hermansky, H. (1990). Perceptual linear predictive (plp) analysis of speech. *Journal of the Acoustical Society of America*, 87(4):1738–1752.

Hermansky, H., Morgan, N., Bayya, A., and Kohn, P. (1992). Rasta-plp speech analysis technique. In *Proceedings of the 1992 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 121–124.

Kessler, V., Thiam, P., Amirian, M., and Schwenker, F. (2017a). Multimodal fusion including camera photoplethysmography for pain recognition. In *2017 International Conference on Companion Technology (ICCT)*, pages 1–4.

Kessler, V., Thiam, P., Amirian, M., and Schwenker, F. (2017b). Pain recognition with camera photoplethysmography. In *IPTA*, pages 1–5. IEEE.

Kohonen, T. (1989). *Self-Organization and Associative Memory, Third Edition*, volume 8 of *Springer Series in Information Sciences*. Springer.

Linde, Y., Buzo, A., and Gray, R. (1980). An algorithm for vector quantizer design. *IEEE Transactions on Communications*, 28(1):84–95.

O'Rourke, J., Chien, C., Olson, T., and Naddor, D. (1982). A new linear algorithm for intersecting convex polygons. *Computer Graphics and Image Processing*, 19(4):384–391.

Penrose, R. (1955). A generalized inverse for matrices. In *Proceedings of the Cambridge Philosophical Society*, volume 51, pages 406–413.

Poulos, M., Rangoussi, M., and Alexandris, N. (1999). Neural network based person identification using EEG features. In *ICASSP*, pages 1117–1120. IEEE Computer Society.

Poulos, M., Rangoussi, M., Chrissikopoulos, V., and Evangelou, A. (1999a). Parametric person identification from the eeg using computational geometry. In *ICECS'99. Proceedings of ICECS '99. 6th IEEE International Conference on Electronics, Circuits and Systems (Cat. No.99EX357)*, volume 2, pages 1005–1008 vol.2.

Poulos, M., Rangoussi, M., Chrissikopoulos, V., and Evangelou, A. (1999b). Person identification based on parametric processing of the eeg. In *ICECS'99. Proceedings of ICECS '99. 6th IEEE International Conference on Electronics, Circuits and Systems (Cat. No.99EX357)*, volume 1, pages 283–286 vol.1.

Schwenker, F., Dietrich, C. R., Thiel, C., and Palm, G. (2006). Learning of decision fusion mappings for pattern recognition. *International Journal on Artificial Intelligence and Machine Learning (AIML)*, 6:17–21.

Snoek, C., Worring, M., and Smeulders, A. W. M. (2005). Early versus late fusion in semantic video analysis. In *ACM Multimedia*, pages 399–402. ACM.

Suresh, M., Krishnamohan, P. G., and Holi, M. S. (2011). Gmm modeling of person information from emg signals. In *2011 IEEE Recent Advances in Intelligent Computational Systems*, pages 712–717.

Thiam, P., Bellmann, P., Kestler, H. A., and Schwenker, F. (2019a). Exploring deep physiological models for nociceptive pain recognition. *Sensors*, 19(20).

Thiam, P., Kessler, V., Amirian, M., Bellmann, P., Layher, G., Zhang, Y., Velana, M., Gruss, S., Walter, S., Traue, H. C., Kim, J., Schork, D., André, E., Neumann, H., and Schwenker, F. (2019b). Multi-modal pain intensity recognition based on the senseemotion database. *IEEE Transactions on Affective Computing*, pages 1–1.

Thiam, P., Kessler, V., Walter, S., Palm, G., and Scwenker, F. (2017). Audio-visual recognition of pain intensity. In *Multimodal Pattern Recognition of Social Signals in Human-Computer-Interaction*, pages 110–126.

Thiam, P. and Schwenker, F. (2017). Multi-modal data fusion for pain intensity assessement and classification. In *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6.

Velana, M., Gruss, S., Layher, G., Thiam, P., Zhang, Y., Schork, D., Kessler, V., Meudt, S., Neumann, H., Kim, J., Schwenker, F., André, E., Traue, H. C., and Walter, S. (2016). The senseemotion database: A multimodal database for the development and systematic validation of an automatic pain- and emotion-recognition system. In *MPRSS*, volume 10183 of *Lecture Notes in Computer Science*, pages 127–139. Springer.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83.