Multimodal Pain Level Recognition using Majority Voting Technique

Amir Salah Mahmoud I. Khalil Hazem Abbas

Computers and Systems Engineering Department, Faculty of Engineering,
Ain Shams University, Cairo, Egypt

***Abstract*—The measurement of subjective pain is still a problem especially with people who have verbal or cognitive impairments. In this work, we analyze the problem of some patients who did not express their pain through their facial muscles, but they were expressing it involuntarily through the autonomic neural system that can be observed in the physiological signals. An ensemble learning algorithm consisting of multimodal models trained on the geometric facial expressions and the physiological signals is proposed. Each model provides a certainty measure and the pain level is assigned by the most certain model. The proposed system is compared with the previous models.**

Keywords: Pain quantification, bio-physiological features, multimodal signals, majority voting, support vector machines, feature extraction and selection.

I. Introduction

Pain is a displeasure feeling that informs us about threats to our bodies. It is a protective mechanism that obliges us for early seeking for treatment and hence plays an important role in the prevention of complications. The sense of pain is reflected on facial expression, some individuals can tolerate pain to a far extent. The threshold of pain (the level of stimulus at which pain starts feeling) is different from one another. In such individuals, the autonomic component reaction of pain aids in recognition of painful stimuli. The autonomic component reaction of painful stimuli includes reflects that mediated by the autonomic nervous system as dilations of skin vessels, tachycardia, the rise of blood pressure and dilatation of pupils.

In this work, an analysis was applied for different visual and bio-physiological features to highly recognize the accuracy level of pain. In this work, pain recognition is divided into two-class problem (pain vs no pain) which is accustomed to and replaced by a continuous pain with different intensity levels. This work introduces a framework for individual autonomous pain estimation from video and bio-physiological channels. We use both videos and bio-physiological signals to recognize the pain independently of the individuals.

The main contribution of this work is to apply majority voting for both videos and bio-physiological signals to reach a higher accuracy compared with other works which used other strategies.

II. Related Work

Current pain assessment in laboratories is still inefficient as it depends on the patient questionnaires or pain scales, which is a great drawback especially for the patients who suffer from dementia or any illness that prevents them to represent clearly their pain. So extensive research is done to accurately measure the subjective pain feeling and to assess it objectively. Many studies till now investigated about the signals that can be an indicator of the pain.

A lot of research already used the multimodal bio-physiological signals such as Electroencephalogram (EEG), blood pressure [1], Electrocardiogram (ECG), heart rate variability [2], Electromyogram (EMG), the Galvanic skin response (GSR) and Changes in skin conductance [3].In other way Lucey et al. [4] used video signals based on face expression to estimate the pain level. Different features have been observed to have a strong relationship with the pain e.g. heart rate and blood pressure. Werner et al [5] applied early fusion for videos and bio-physiological signals. Lopez-Martinez et al [6] used skin conductance (SC) and ECG for personalized pain recognition and achieved accuracy between 82.75% and 70.04%. Kchele et al. [7] used bio-physiological signals to measure the similarity for personalized continuous real-time estimation of pain intensity.

To our knowledge, a few studies only addressed the issue that there is a class of people who have no pain manifestation in their face. These people may suppress their pain in their face, but we may still detect it from the signals sent by the brain to the body that can be processed from the bio-physiological signals. Our results showed that the learning model that takes a majority vote of three different models outperformed one single model based on the video modality as it reduces the overall model accuracy. The first model is trained for the video modality, the second model is trained for the bio-physiological signals and the third is trained in classifying the data fusion of the two modalities.

III. The Proposed Pain Level Recognition System

Current assessment of pain depends on the verbal methods (e.g., pain scales and questionnaires). In this work, we automate this process of pain assessment depending on current machine learning systems. Machine learning systems are systems that are designed to iterate through previously correct assessed samples and try to get a new model that describes the relationship between the inputs and the outputs.

In this work as shown in Figure 1, we build a machine learning model based on majority vote technique to reach to a higher accuracy by combining results from each modality by

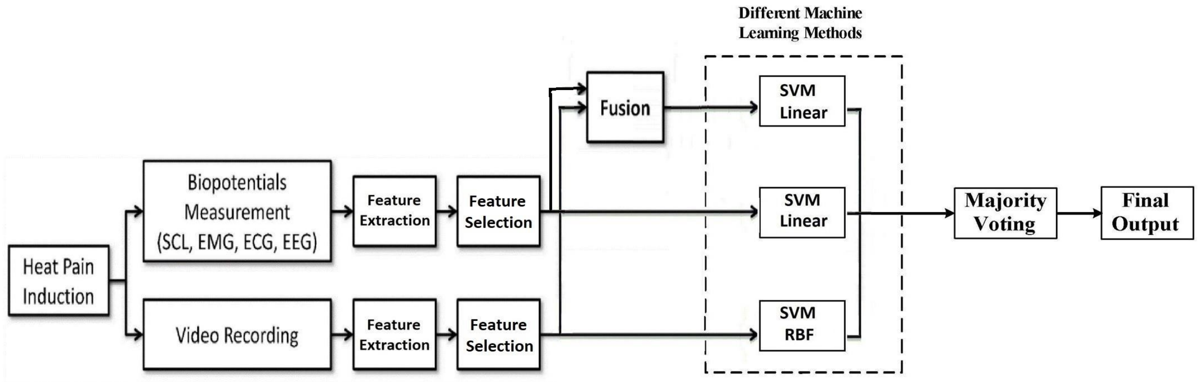


Fig. 1. The proposed machine learning architecture.

its optimum model and to boost the results of the unimodal models. A lot of models are trained on the videos and bio-physiological data, but the most optimum one used was the SVM-RBF model on videos in parallel with SVM linear model on bio-physiological which are found to give the highest accuracy compared with a lot of models tested. Alongside with the mentioned two models, a third model is used, SVM model that is trained on the features after the step of the early fusion. The three models are used together with a majority voting step to correctly assess the pain. We tested the system on 25% of the dataset and then we compared the results with the previous results.

A. Geometric Based Features

The features are extracted from the videos by iterating through the frames and locating the Face using Viola-Jones and then locating the landmarks using the Dlib implementation of One Millisecond Face Alignment with an Ensemble of Regression Trees [[8]]. Geometric features are extracted from the video files by computing 68 facial points that describe the face, nose, and mouth at the frame level, and computing the Euclidean distances between each one and the computed centroid of the 68 points.

Ten statistical measures (mean, standard deviation, min, max, median, range, inter-quartile range, median absolute deviation, mean absolute deviation, trimmed mean) were measured sequentially for each distance resulting of 680 features that encoded the facial points through the entire video. These features are fed to forward feature selection step to reduce the feature set and then the top effective 80 features are fed to a lot of models to be assessed and each one is evaluated. The SVM-RBF model outperformed the others. Percentage of data used for SVM training and prediction are 75% and 25%, respectively.

B. Feature extraction from biomedical signals

We extracted features from the three different bio-physiological signals that are contained in the dataset:

- SCL: Skin glands are affected by sympathetic system via Epinephrine and Norepinephrine (Adrenaline and Noradrenaline). Pain and tension lead to sympathetic overactivity and hence raised Epinephrine and Norepinephrine which are reflected in sweat glands. We apply this idea for measuring SCL (Skin Conductance Level) by placing the two electrodes of the sensors on two fingers, namely the index and the ring one. The electrodermal activity is thus a good indicator of inner tension of the individual and a good representation for the level of pain.
- ECG: This signal measures the heart rate, the heart rhythm and the heart rate variability (HRV). The latter describes the fluctuations in the heart rate and is considered in the literature as an indicator of stress and mental effort. To record the ECG, two electrodes of the Nexus sensor were also used here. An electrode was placed on the right side of the upper body of the test subject just below the collarbone, and the second on the left side in height of the last three ribs.
- EMG: Muscle activity is generally one signs of bio-physiological arousal. An increase in muscle tone is associated with increasing activity of the sympathetic nervous system. A sinking on the other hand stands for an excitement of the parasympathetic nervous system.

The derivatives of the individual EMG signals were made by special 2-channel Nexus device [9]. These are able, through carbon technology paired with special Active Noise Reduction technique to deliver a very clean signal. For the recording of a facial EMGs were two electrodes to the left facial muscles corrugator supercilii and zygomaticus major appropriate. Musculus corrugator supercilii and zygomaticus major muscle belong both to mimic muscles. The former ensures the lowering of the eyebrow and for the frown of the forehead, while the second for pulling the corner of the mouth upwards and back is responsible. From both muscles was in the experiment of it assumed that they are activated during a pain stimulus.

Two other electrodes on the right part of the trapezius muscle, which led the muscle activity in the shoulder or neck area. This

muscle is trapezoidal between spine and shoulder and has the function of the movement of the scapula held. Because a tense neck and shoulder area is an indicator of increased stress level and pain can be considered a stressor also here an increased activity during the pain stimulation suspected.
We extracted five groups of features from each signal (amplitude, frequency, stationarity, entropy, linearity). 135 features are reproduced then applying a forward feature selection for these features to reduce the number of features. The top effective features are used to train a SVM linear model to assess the pain depending on the extracted physiological features. 75 % of the data used to train the SVM and 25% for prediction.

C. Early fusion of features

To combine predictions of different modalities, an early fusion technique is employed. The features extracted from the video signals and the bio-physiological signals are combined to be learned with a new SVM linear model to give predictions on the combined features and model evaluation has been done to assess the model trained on both features.

D. Majority voting and pain quantification

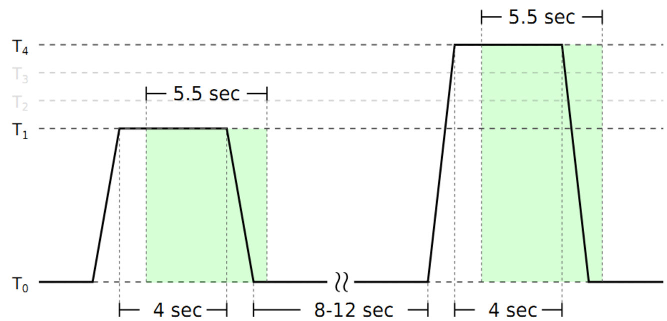
First, we evaluated the models in the video signals and the system gives a low accuracy. When the bio-physiological signals are used, we got a relatively higher accuracy and yet the fusion of the two signals did not give good results. In addition, it is recorded that there is a subset of the subjects who had shown no pain in their facial expressions whereas their bio-physiological signals showed that there is a pain.
So we proposed that a multimodal majority voting technique consisting of three learned models could improve the efficiency of our system. By using the trained models using the video signals, bio-physiological signals and the fusion of the two signals, the experiments have proved that the performance of the majority voting improved the performance of each individual model on its own.

IV. Experimental Results

In this section, the experimental results are shown for the pain level recognition for each modality (Videos modality, Bio-physiological modalities) separated and compared to the early fusion of the two signals. It has been shown that the video modality degraded the results of the bio-physiological modalities. Moreover, the results are shown for the proposed Majority Voting algorithm.

A. BioVid Heat Pain DB

In the experiments, the BioVid Heat Pain dataset [10] is well analyzed. It includes the data from 87 individuals which is divided into 44 males and 43 females. According to age, they were three groups 15-35 years, 36-50 years and 51-65 years in nearly equal proportions.
In experiments, a thermode was used for the elicitation of pain. Pain intensity was calibrated for each individual in a manner that it divided the range between two reference levels (threshold and tolerance) into equally spaced intervals. We inform each individual participant to press button when the pain starts (i.e threshold) and again when pain is bearable (i.e tolerance).
Five intensity levels were equally distributed starting from no pain (baseline, BLN; corresponding to T0) to pain levels P1 to P4 (corresponding to calibrated temperatures from T1 to T4). Pain stimulation was applied 20 times for each intensity in addition to baseline measurements. The total number is 100 stimulations for each participant individually. Each stimulus was applied for 4 seconds followed by 8-12 seconds of recovery phase as shown in Figure 2.



(a)

(b)

Fig. 2. (a) Graph of a 4-second heat stimulus with exemplary intensity T1 (pain threshold) and subsequent break. The length of the break varies randomly between 8 and 12 seconds. (b) Video recording. The front camera is recording the experiment of the individuals [7].

B. Video Modality Experiment

The classification results based on videos alone did not give the best accuracy as there are some patients who did not express their pain through their facial muscles. A lot of models are trained on the videos but the most optimum one used was the SVM-RBF model which gives the best accuracy. For the two-class problem highest pain vs no pain (B vs T4) the Video signals models achieved 67% as shown in Figure 3.

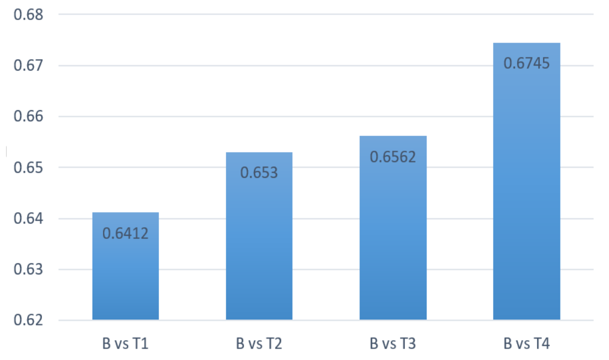


Fig. 3. Classification accuracy between different pain levels using videos alone.

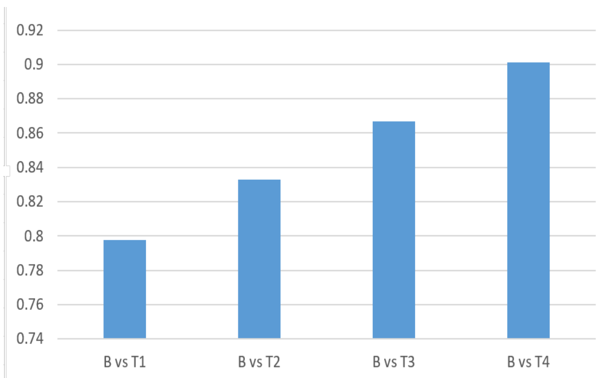


Fig. 4. Classification accuracy between different pain levels using Bio-physiological only.

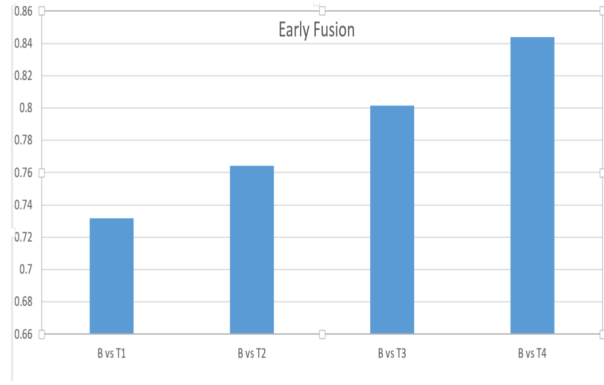


Fig. 5. Classification accuracy between different pain levels using Early Fusion.

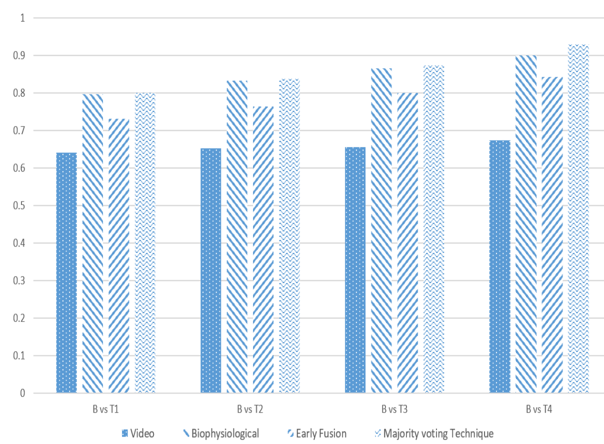


Fig. 6. Classification accuracy between different pain levels.

C. Bio-physiological Modalities Experiment

The classification results based on bio-physiological alone give us better accuracy than videos only as it represents a good indicator of different pain levels. Among the used classifier models used to assess the discrimination, the linear SVM model was found to have the best results achieving 90.11% for the two-class problem pain vs no pain (B vs T4) as shown in Figure 4.

D. Early Fusion Experiment

When Early fusion is used between video and bio-physiological signals, a slight decrease in the accuracy could be noticed to achieve only 0.8437% for the two-class problem pain vs no pain (B vs T4) as shown in Figure 5.

E. The Proposed Algorithm Experiment

The classification results are summarized in Table I. In this table, the classification accuracy of each modality on its own. Accuracy is given by:

$$Accuracy(ACC) = \frac{\sum TruePositive + \sum TrueNegative}{TotalPopulation} \tag{1}$$

The video showed the least overall accuracy, and when the bio-physiological signals are fused with the video signals the bio-physiological signals are slightly decreased.

TABLE I. CLASSIFICATION ACCURACY FOR EACH MODALITY.

	Video	bio-physiological	Early Fusion	Majority voting Technique
B vs T1	64.12	79.76	73.16	80.12
B vs T2	65.3	83.29	76.42	83.89
B vs T3	65.62	86.70	80.13	86.45
B vs T4	67.45	90.11	84.37	92.01

For the two-class problem pain vs no pain (B vs T4) the Video signals models achieved 67% accuracy while the bio-physiological achieved 90%. The early fusion features achieved 84% accuracy.
After taking this into consideration, the machine learning model based on the majority voting of the three models outperformed the three individual models and achieved 92.01% accuracy as shown in Figure 6.
In a Receiver Operating Characteristic (ROC) curve the true positive rate(TPR) is plotted in function of the false positive rate(FPR) for different cut-off points for each two-class from (BLN vs T1) to (BLN-T4).

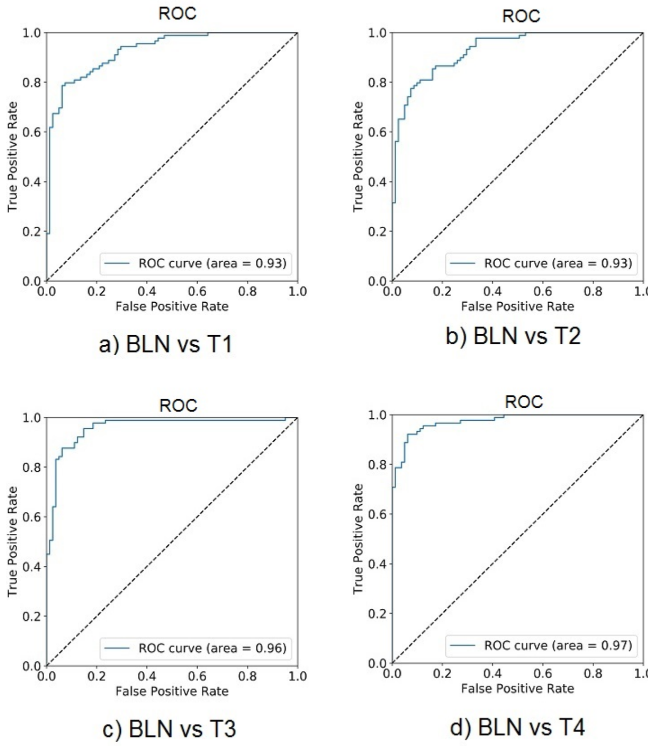


Fig. 7. ROC Curve for each two-class model pain vs no pain from T1 to T4.

True positive rate is given by:

$$TPR(T) = \int_{T}^{\infty} f_1(x)dx \tag{2}$$

False positive rate is given by:

$$FPR(T) = \int_{T}^{\infty} f_0(x)dx \tag{3}$$

where T is the threshold where the instance is classified as pain if $X > T$ and no pain otherwise,$f_1(x)$ is the probability density if the instance actually belongs to pain, and $f_0(x)$ if otherwise. Area under the curve is given by:

$$\begin{aligned} AUC &= \int_{\infty}^{-\infty} TPR(T)FPR^{'}(T)dT \\ &= \int_{\infty}^{-\infty}\int_{\infty}^{-\infty} I(T^{'} > T)f_1(T^{'})f_0(T)dT^{'}dT \\ &= P(X_1 > X_0) \end{aligned} \tag{4}$$

Here X_1 is the score for a positive pain and X_0 is the score for a positive no pain as shown in Fig 7, The ROC for BLN vs T4 reached 97% and BLN vs T3 reached 96% in the other side BLN vs T1 and BLN vs T2 reached the same value 93%.

F. Comparative Experiment

In this section, the experimental results in Table II have shown that the proposed algorithm results outperformed the previous work [11], [6], [5], [12] using the same dataset. Gruss et al [11] used all Bio-physiological signals in classification after applying manual and automatic feature selection without using the videos and reached to accuracy between 79.29% - 90.94%. Lopez-Martinez et al [6] used skin conductance (SC) and ECG for personalized pain recognition only and ignored other Bio-physiological signals so they achieved accuracy between 70.04% - 82.75%.Werner et al [5] applied random forest with early fusion for videos and bio-physiological signals to get better results but they didn't take in considration that some individuals may suppress their pain in their face. Better accuracy has been achieved by the proposed algorithm because of using the majority voting technique which consists of three learned models. This technique improved the efficiency of our system and increased the accuracy.

V. CONCLUSIONS

This paper presents a model that can efficiently discriminate between pain and no pain classes. It has been shown to robustly increase the accuracy, in contrast to only using the video signals or the bio-physiological signal. Further investigation is needed about the degrading effect of the video signals to accurately detect the pain for those who conceal their pain.

REFERENCES

[1] Marcella Saccò, Michele Meschi, Giuseppe Regolisti, Simona Detrenis, Laura Bianchi, Marcello Bertorelli, Sarah Pioli, Andrea Magnano, Francesca Spagnoli, Pasquale Gianluca Giuri, et al. The relationship between blood pressure and pain. *The Journal of Clinical Hypertension*, 15(8):600–605, 2013.

[2] Gunnar Jess, Esther M Pogatzki-Zahn, Peter K Zahn, and Christine H Meyer-Frieem. Monitoring heart rate variability to assess experimentally induced pain using the analgesia nociception index: A randomised volunteer study. *European Journal of Anaesthesiology (EJA)*, 33(2):118–125, 2016.

[3] Hanne Storm. Changes in skin conductance as a tool to monitor nociceptive stimulation and pain. *Current Opinion in Anesthesiology*, 21(6):796–804, 2008.

[4] Patrick Lucey, Jeffrey F Cohn, Kenneth M Prkachin, Patricia E Solomon, Sien Chew, and Iain Matthews. Painful monitoring: Automatic pain monitoring using the unbc-mcmaster shoulder pain expression archive database. *Image and Vision Computing*, 30(3):197–205, 2012.

[5] Philipp Werner, Ayoub Al-Hamadi, Robert Niese, Steffen Walter, Sascha Gruss, and Harald C Traue. Automatic pain recognition from video and biomedical signals. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 4582–4587. IEEE, 2014.

[6] Daniel Lopez-Martinez and Rosalind Picard. Multi-task neural networks for personalized pain recognition from physiological signals. In *Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), 2017 Seventh International Conference on*, pages 181–184. IEEE, 2017.

[7] Markus Kächele, Patrick Thiam, Mohammadreza Amirian, Friedhelm Schwenker, and Günther Palm. Methods for person-centered continuous pain intensity assessment from bio-physiological channels. *IEEE Journal of Selected Topics in Signal Processing*, 10(5):854–864, 2016.

[8] Vahid Kazemi and Sullivan Josephine. One millisecond face alignment with an ensemble of regression trees. In *27th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, United States, 23 June 2014 through 28 June 2014*, pages 1867–1874. IEEE Computer Society, 2014.

TABLE II. COMPARISON WITH THE PREVIOUS WORK

Algorithm	**Features**	**BLN vs T1**	**BLN vs T2**	**BLN vs T3**	**BLN vs T4**
SVM [11]	All Bio-physiological signals (EMG,ECG,SCL,..)	79.29	81.29	84.94	90.94
MT-NN[6]	SCL+ECG	51.72	58.62	70.04	82.75
Random Forest[5]	All video (facial expression, head movement features) + Bio-Physiological	49.6	60.5	72.0	80.6
Random Forest[12]	All bio-physiological signals (EMG,ECG,SCL,..)	–	–	–	83.1
The Proposed Algorithm Majority Voting (SVM + SVM RBF +Early fusion)	All video + Bio-physiological signals	80.12	83.89	86.45	92.01

[9] Steffen Walter, Sascha Gruss, Hagen Ehleiter, Junwen Tan, Harald C Traue, Philipp Werner, Ayoub Al-Hamadi, Stephen Crawcour, Adriano O Andrade, and Gustavo Moreira da Silva. The biovid heat pain database data for the advancement and systematic validation of an automated pain recognition system. In *Cybernetics (CYBCONF), 2013 IEEE International Conference on*, pages 128–131. IEEE, 2013.

[10] Lin Zhang, Steffen Walter, Xueyao Ma, Philipp Werner, Ayoub Al-Hamadi, Harald C Traue, and Sascha Gruss. biovid emo db: A multimodal database for emotion analyses validated by subjective ratings. In *Computational Intelligence (SSCI), 2016 IEEE Symposium Series on*, pages 1–6. IEEE, 2016.

[11] Sascha Gruss, Roi Treister, Philipp Werner, Harald C Traue, Stephen Crawcour, Adriano Andrade, and Steffen Walter. Pain intensity recognition rates via biopotential feature patterns with support vector machines. *PloS one*, 10(10):e0140330, 2015.

[12] Markus Kächele, Patrick Thiam, Mohammadreza Amirian, Philipp Werner, Steffen Walter, Friedhelm Schwenker, and Günther Palm. Multimodal data fusion for person-independent, continuous estimation of pain intensity. In *Engineering Applications of Neural Networks*, pages 275–285. Springer, 2015.