INDIVIDUAL EMOTION RECOGNITION AND SUBGROUP ANALYSIS FROM PSYCHOPHYSIOLOGICAL SIGNALS

Lin Zhang, Harald C. Traue, Dilana Hazer-Rau

Medical Psychology, University of Ulm, Ulm, Germany

ABSTRACT

This study involves intra- and inter-individual emotion classifications from psychophysiological signals and subgroup analysis on the influence of gender and age and their interaction on the emotion recognition. Individual classifications are conducted using a selection of feature optimization, classification and evaluation approaches. The subgroup analysis is based on the inter-individual classification. Emotion elicitation is conducted using standardized pictures in the Valence-Arousal-Dominance dimensions and affective states are classified into five different category classes. Advantageous intra-individual rates are obtained via multi-channel classification and the respiration best contributes to the recognition. High inter-individual variances are obtained showing large variability in physiological responses between the subjects. Classification rates are significantly higher for women than for men for the 3-category-class of Valence. Compared to old subjects, young subjects have significantly higher rates for the 3-category-class and 2-category-class of Dominance. Moreover, young men's classification performed the best among the other subgroups for the 5-category-class of Valence/Arousal.

KEYWORDS

Individual Classification, Affective Computing, Emotion Recognition, Biosignal Processing, Physiological Response, Gender/Age Analysis.

1. INTRODUCTION AND RELATED WORK

Individual emotion recognition is increasingly gaining interdisciplinary significance in many human-computer interaction fields including healthcare, educational and mobile applications as well as cognitive intelligent systems such as companion, prevention or elderly support applications. A trans situational investigation using the current dataset of this study shows for instance a correlation between the physiological recognition rate of emotions with the dialog success during human-computer interaction and with individual differences in brain activation in an fMRI study [1]. Human affective states can be assessed based on the analysis of various modalities, including facial expression, body gestures, contexts, speech and physiological signals [2,3,4]. Among these different modalities, psychophysiological signals have various considerable advantages in assessing human affective expressions. As honest signals, they are considered as the most reliable for human emotion recognition as they cannot be easily triggered by any conscious or intentional control [5].

From the perspective of applicability, there are two types of emotion recognition models: subject-dependent and subject-independent models [6,7,8,9]. Subject-dependent models are suitable to a given individual subject or a specific group and are difficult to apply to new samples. Subject-independent models are convenient for more general applications as the classification is tested with unknown data that is data collected from users different from the ones used for the initial training of the classifier [7]. Due to the individual differences of physiological responses and

inter-individual variability, training a subject-independent model with high validity and accuracy is much harder than training a subject-dependent model [8,9]. Subject-independent models need more informative features and more intelligent classification algorithms to deliver good performance [6]. In the application of these models, intra-individual and inter-individual testings are conducted. In Intra-individual classifications, the data collected from a subject is used for training and testing, while in inter-individual classifications, also commonly known as Leave-One-Subject-Out technique, one subject's material is left out in training and afterwards utilized for testing only [10].

Several researchers have also studied the influence of gender and age on the physiological responses during emotional processing. Bradley et al. found gender differences on the motivational-related responses to pictures with various valences [11]. Hazer et al. investigated the effect of emotion elicitation on various age groups using standardized movie clips representing five basic emotions. The results show an influence of age on the film-clip choice, a correlation between age and valence/arousal rating as well as differences in valence and arousal ratings in the different age groups [12]. The studies on brain activities demonstrated that females react more to negative pictures whereas male have greater activities when watching erotic pictures [13,14]. Further, several studies indicate that old subjects have lower physiological responses in terms of electromyography, skin conductance, heart rate changes and finger temperature [15,16]. However, fewer investigations are conducted on the effect of gender and age on the psychophysiological emotion classification [17,18,19]. To the best of our knowledge, studies related to the influence of gender and age on the classification accuracies are very limited despite their important implications in the emotion computing field.

For the psychophysiological emotion recognition, various feature extraction and selection optimization approaches, classification algorithms, and evaluation methods are currently been used [20,21,22]. To adapt to the fast-changing technologies and recent applications, automatic recognition algorithms have been also applied to different biosignals in order to efficiently compute and classify human's affective states [23,24,25]. Thereby, various emotional models are employed to describe the emotional space for the affect recognition based on discrete and dimensional models. Psychophysiological emotion classification allows suitable categorization of affective states, for instance in terms of Valence, Arousal and Dominance subspaces [26,27].

In the following, we present individual classifications and subgroup analysis for the human emotion recognition from psychophysiological signals. The application of various feature selection techniques, classification algorithms and evaluation methods is thereby based on our previous performance evaluation study conducted on a larger dataset [28]. In Hazer-Rau et al., we analyzed the combination of approaches that are most reliable for best identifying different emotional states using intra-individual classifications. In the present study, our previous findings are first adopted on the current dataset and the analysis is extended to include inter-individual classifications as well as feature selection analysis. Finally, based on the inter-individual recognition, the present study also includes a subgroup analysis in which we evaluate the variances of individual classification accuracies to investigate the influence of gender and age differences on the physiological responses and their related classification rates.

2. METHODS

2.1. Experimental Procedure

2.1.1 Subjects Description

The dataset used in this study is based on subjects previously recruited via bulletins distributed on the campus of the University of Magdeburg. The total sample size was n= 20 subjects (9 women, 11 men) between the age of 22 and 76 years old. All subjects were right-handed, healthy and had normal vision or corrected normal vision.

2.1.2 Emotion Elicitation

Emotion induction was conducted using standardized stimuli from the International Affective Picture System (IAPS) and extended by the Ulm Picture Set to represent the VAD (Valence, Arousal, Dominance) space. Both picture systems allow a dimensional induction of emotions according to their ratings in the Valence, Arousal and Dominance dimensions [29]. The experimental design is thereby based on a previous experiment [30], adapted to stimulate prolonged emotion induction. Prolonged presentations consisting of 10 pictures with similar rating à 2s each (total of 20s per presentation) are used to intensify the elicitation [31]. A total of 10 sets of these picture-presentations à 20s each were presented to induce a total of 10 different VAD-states. Thus, the induced VAD-space for the 10 sets of picture-presentations included combinations of positive/negative/neutral (+/-/0) Valence (V), positive/negative (+/-) Arousal (A), and positive/negative (+/-) Dominance (D) values. In order to neutralize the user's affective state between two different sets of presentations, 20s of neutral fixation crosses were introduced as baseline inbetween. In total, 100 pictures were used for the emotion induction. While the order of pictures in each presentation-set was fixed, the display of the 10 sets was randomized.

For the classification, picture-presentation with similar ratings in terms of Arousal (+/-) and/or Valence (+/-/0) and/or Dominance (+/-) were combined into one category. In this study, we evaluated in total five different category classes, as presented in Table 1.

Table 1. Overview of the five different category classes used in this study.

2-category-class: 2Cat(D)	Dominance D: + / -
2-category-class: 2Cat(A)	Arousal A: + / -
3-category-class: 3Cat(V)	Valence V: + / - / 0
5-category-class: 5Cat(VA)	VA: 0- / ++ / -+ / +- / --
10-category-class: 10Cat(VAD)	VAD: All picture-presentations

2.1.3. Physiological Signals

The physiological signals analyzed in this study include:

Skin Conductivity (SC): Two electrodes connected to the sensor were positioned on the index and ring fingers. Since the sweat glands are innervated sympathetically, electrodermal activity is a good indicator of the inner tension of a person.

Respiration (RSP): The respiration sensor was used to measure the abdominal breathing frequency, as well as the relative depth of breathing. It was placed tight enough in the abdominal area just above the navel.

Electromyography (EMG): Electrical muscle activity is related to the activity of the sympathetic nervous system. We used two-channel electromyography EMG signals for the zygomaticus major (Zyg.) and the corrugator supercilii, (Corr.) muscles, which are expected to be active during different emotions.

2.2 Data Processing and Classification Analysis

The processing of the physiological signals includes the pre-processing of the raw data and the feature analysis as well as the emotion classification. The emotion recognition analysis includes both intra-individual and inter-individual classifications. These steps are described in the following subsections.

2.2.1. Pre-processing and Feature Analysis

For both intra-individual and inter-individual analysis, the raw data are first pre-processed by extracting the relevant signals and picture-sessions from the whole dataset. Then, the extracted data are further processed and prepared to meet the AuBT (Augsburg Biosignal Toolbox) file format requirements [32]. The Toolbox provides tools to analyze physiological signals for the emotion recognition [33]. It is used in this study for the later signal processing and analysis including the feature extraction and feature selection as well as the emotion classification and the evaluation analysis. For the application of these pre-processing steps and for the optimization of the quality of the signals, we adopt our previously developed automation scripts and filtering techniques, composed and implemented as Matlab-based functions [28].

In the next step, the AuBT toolbox is used to extract features from the physiological signals including skin conductivity and respiration, as well as the electromyography signals, EMG corrugator and EMG zygomaticus. For the intra-individual analysis, all acquired signals are examined both individually as well as in various combinations among each other. While this is similar to the classification analysis conducted in our previous study [28], the inter-individual analysis is conducted using all signals combined together. In both intra-individual and inter-individual, and for each of the resulting signal configuration, the selection of the features is optimized and the resulting selected features are used to train and evaluate a classifier. Feature selection is thereby optimized using the Sequential Forward Selection (SFS) and the Sequential Backward Selection (SBS) algorithms.

2.2.2. Intra-individual Emotion Classification

The intra-individual emotion classification is conducted for four different category-classes, including the 2-category-class of Arousal, the 3-category-class of Valence, the 5-category-class of Valence-Arousal and the 10-category-class of Valence-Arousal-Dominance (as shown in Table 1, excluding the two-category-class of Dominance).

As classification methods, the k-Nearest-Neighbors (kNN) and the Linear Discriminant Analysis (LDA) models are adopted. Finally, the classifiers are evaluated using three different evaluation methods including the normal split, the random split and the one-leave-out methods.

Each of the mentioned approaches is executed using various combinations of strategies and parameters: The SFS and SBS feature selection algorithms are tested using both "break" (stops as

soon as increasing SFS or decreasing SBS results in a feature set, that has a lower recognition rate) and "best" (picks the subset consisting of the first n features – n<20 with the highest recognition rate) strategies. The kNN classifier is applied using k closest training samples in the feature space, with k varying between 3 and 8 nearest neighbors. On the other hand, the statistical LDA classifier requires no parameter input and is used with no variation. In addition to the feature selection, feature reduction based on the Fisher transformation is also applied and the recognition rates are compared to the classification results without reduction of dimensionality (Fisher vs. none).

As for the classifier evaluation methods, the normal split and random split methods are applied using both x= 0.75 and x= 0.90 parameters. In the normal split, the first x(%) of the samples are taken for training and the rest for testing, while in the random split, x(%) of the samples are taken for training and the rest for testing but the data are divided randomly. Further, in the random split method, the procedure is repeated iter times and the average recognition rate of all runs is calculated. The iter parameter is set to both 10 and 20 iterations. Finally, the one leave out method is applied with no variation, using only one sample at a time for testing and the rest to train the classifier. This is repeated for each sample and the final result is the average of all runs.

The presented choice of signal combinations, feature extraction and selection approaches as well as the classification techniques and evaluation methods adopted here are based on a pre-selection from two previous studies [28,34]. In Zhang et al., we conducted a preliminary explicit analysis of various approaches and their combinations in order to evaluate their efficiency in the emotion recognition process [34]. In Hazer-Rau et al. the performance of those various emotion classification approaches was explicitly further evaluated as intra-individual analysis [28]. For the present intra-individual and inter-individual analysis, we adopt the approaches which best performed in the first study and which were evaluated for their performance in the latter study.

An overview of the physiological signals, feature selection techniques, classification algorithms and evaluation methods applied for the intra-individual analysis is presented in Table 2.

Table 2. Overview of the physiological signals, feature selection, classification and evaluation approaches applied in the intra-individual analysis.

Physiological Signals	- Skin Conductivity (SC) - Respiration (RSP) - SC & RSP - EMG-Corr - EMG-Zyg - EMG-Corr & EMG-Zyg - ALL signals combined
Feature Selection	- Sequential Forward Selection (SFS) - Sequential Backward Selection (SBS)
Classification Algorithms	- k-Nearest-Neighbors (kNN) - Linear Discriminant Analysis (LDA)
Evaluation Methods	- Normal Split - Random Split - One-Leave-Out

2.2.3. Inter-individual Emotion Classification

The inter-individual emotion classification is conducted for all five category-classes shown in Table 1, including the 2-category-class of the Dominance, the 2-category-class of Arousal, the 3-category-class of Valence, the 5-category-class of Valence-Arousal and the 10-category-class of Valence-Arousal-Dominance.

Inter-individual classification rates are obtained by applying the normal split evaluation method with x= 0.95, corresponding to 19 subjects (95%) used for training the classifier and a single remaining individual subject (5%) used as test subject for evaluating the classifier. This is equivalent to the leave-one-subject-out method, leaving one subject's material out during the training and using it later for testing only [10]. In order to calculate the individual recognition rate for each of the 20 subjects, the classification procedure is repeated 20 times to include all the subjects.

Further, the present inter-individual analysis is conducted using all physiological signals combined. This is based on our previous results, showing that the signals' combination leads to the highest recognition rates for all the different category-classes [28].

With exception of the evaluation approach (normal split; x= 0.95) and the physiological signals used (ALL combined), the inter-individual emotion recognition processing, including the feature selection techniques and classification algorithms, is similar to the intra-individual processing described in subsection 2.2.2.

An overview of the physiological signals, feature selection techniques, classification algorithms, and evaluation methods adopted in the inter-individual analysis is presented in Table 3.

Table 3. Overview of the physiological signals, feature selection, classification and evaluation approaches applied in the inter-individual analysis.

Physiological Signals	- ALL signals combined (SC, RSP, EMG-Corr, EMG-Zyg)
Feature Selection	- Sequential Forward Selection (SFS) - Sequential Backward Selection (SBS)
Classification Algorithms	- k-Nearest-Neighbors (kNN) - Linear Discriminant Analysis (LDA)
Evaluation Methods	- Normal Split (with x= 0.95)

2.3. Subgroup Analysis

The subgroup analysis is based on the inter-individual classification analysis. It consists of three parts. The first part is testing the gender differences of the individual classification rates. The second part is related to the effect of age on the individual classification rates. The last third part involves the effect of both gender and age. All three parts consider the inter-individual classification rates obtained as described in Subsection 2.2.3 and derived from the combination of methods presented in Table 3. The statistical analysis is performed using SPSS 22.

In the first analysis part of gender differences, the processing is carried out according to the following steps: First, a descriptive analysis is used in order to describe the overall situation

including the number of subjects, the mean of recognition rates and the associated standard deviation in the two (male and female) groups. The second step is testing the normality of the sample. For this purpose, the Shapiro-Wilk test is used, which is an effective test of normality in frequentist statistics. In the last step, if the samples come from a normally distributed population, which is verified by Shapiro–Wilk test, the t-test is used to compare the means of the two samples, the male and the female group samples. In that case, the independent-samples t-test is used if the variances of the population from two samples are equal (in case of homogeneity of variance); otherwise, the independent-samples t´-test is used if the population variances are not equal (in case of non-homogeneity of variance). If the samples are not from a normally distributed population, the Mann-Whitney U test is used for comparing the mean of the two samples. The Mann-Whitney U test is a nonparametric test suitable for small samples as the case in this study. Further, it does not require the assumption of normal distribution, but is nearly as efficient as the t-test on normal distributions.

The analysis in the second part of age differences is similar to that of the first part. In the second part, the two samples are old group and young group.

In the third part, the differences among the groups of young men, young women, old men and old women are compared and analyzed. Also here, a simple descriptive statistical analysis is first applied to summarize the samples. Then, the normality of the sample is tested using the Shapiro-Wilk test and the homogeneity of variance is additionally tested using the Levene´s test. If the population of samples is normally distributed, then the one-way analysis of variance (one-way ANOVA) is used to analyze the differences among the group means and their associated variations among and between the groups. When significant differences of means among the groups resulting from ANOVA analysis are found, two post-hoc tests, consisting of both Fisher´s least significant difference (LSD) and the Bonferroni tests are used to get deeper details on the significant groups. If the population of samples is not normally distributed, the rank analysis and the Chi-Square test are used to analyze the differences among the groups.   An introduction to the Shapiro-Wilk, the Mann-Whitney U, the Levene´s, the Bonferroni and the Chi-Square tests is described in the book written by Afifi and Azen [35].

3. RESULTS

Feature analysis and classification rates of the various emotion category-classes as well as the subgroup analysis on the different age/gender groups are presented in the following subsections.

3.1. Feature Analysis

An analysis of the amount of features selected for both the intra-individual and inter-individual classifications is presented with regard to the used physiological signals and feature selection methods. The type of features selected depends on the used category-class and covers the range of the features implemented in the AuBT toolbox.

3.1.1 Intra-individual Analysis

Among all the extracted features, the amount of selected features for the intra-individual classification varies from 3 to 124 features. When the SFS selection method is applied, the number of features selected does not exceed 25 features. This is overall less than the number of features selected when the SBS method is used and which varies between 11 and 124 features. Considering more than one signal, the amount of features selected is larger than considering only one signal to classify the emotions. The largest amount of selected features, varying between 122

and 124 features, is obtained when considering all the signals combined and the SBS feature selection method.

An overview of the number of features selected for the applied category-classes for every signal configuration with respect to the applied feature selection method is summarized in Table 4.

Table 4. Overview of the features selected for every signal configuration with respect to the applied selection method in the case of the intra-individual classification.

	SFS	SBS	SFS	SBS	SFS	SBS	SFS	SBS	SFS	SBS	SFS	SBS	SFS	SBS
	SC		RSP		SC & RSP		EMG-Cor		EMG-Zyg		EMG-Cor & -Zyg		ALL Signals	
2Cat(A)	3	13	14	13	25	83	4	17	19	13	4	31	20	123
3Cat(V)	8	13	7	64	9	82	4	14	3	11	3	37	14	124
5Cat(VA)	3	16	7	63	5	83	5	9	10	18	13	22	3	123
10Cat(VAD)	3	14	4	63	23	83	3	17	5	19	8	30	4	122

3.1.2 Inter-individual Analysis

For the inter-individual analysis, we consider the number of features selected for every subject as well as the number of non-repetitive and repetitive features. A repetitive feature means a feature that is selected for more than one subject in one category-class and using one selection method. The number of non-repetitive features for all subjects is the sum of the selected features from every single subject (for all 20 subjects) minus the number of repetitive features.

When the SFS method is used, the range of selected features for every single subject varies from 2 to 25 features. The numbers of non-repetitive features are 72, 61, 48, 53 and 64 for the 2-category-class of Arousal, the 2-category-class of Dominance, the 3-category-class of Valence, the 5-category-class of Valence-Arousal and the 10-category-class of Valence-Arousal-Dominance, respectively. On the other hand, when the SBS method is utilized, the range of selected features varies between 13 and 122 features, which is larger than the range selected using the SFS method. The numbers of non-repetitive features are 99, 103, 104, 102 and 105 for the 2-category-class of Dominance, the 3-category-class of Valence, the 5-category-class of Valence-Arousal and the 10-category-class of Valence-Arousal-Dominance, respectively.

An overview of the features selected for the applied category-classes with respect to the adopted selection method in the case of the inter-individual classification is presented in Table 5.

Table 5. Overview of the features selected with respect to the applied selection method in the case of the inter-individual classification.

	SFS (break)		SBS (break)	
	Number of non-repetitive Features	Range of Number of Features	Number of non-repetitive Features	Range of Number of Features
2Cat(A)	72	2-20	99	20-44
2Cat(D)	61	4-21	103	20-50
3Cat(V)	48	2-21	104	20-116
5Cat(VA)	53	3-17	102	20-77
10Cat(VAD)	64	3-25	105	13-122

3.2. Classification Analysis

3.2.1 Intra-individual Classification

Table 6 summarizes the highest obtained recognition rates for the intra-individual classifications. The results are illustrated for four different category-classes and for various signal configurations. The adopted combinations of feature selection, classification algorithms and evaluation methods are also presented in the table and are compatible with the results obtained from our previous performance evaluation study [28].

As shown in Table 6, for the 2-category-class of Arousal (A:+/-), the range of classification rates varies from 68.75% to 89.58%. The highest recognition rate of 89.58% is obtained when including respiration and skin conductivity signals in the analysis. This result is obtained using the SBS feature selection, the kNN-none/Fisher classification algorithms and the normal split 0.9 evaluation method. Considering all physiological signals results in a comparable recognition rate of 80%, using the SFS feature selection, the LDA-none classification and the normal split 0.75 evaluation method. Considering only a single physiological channel in this 2-category-class, the respiration signal seems to best contribute to the performance, with a classification rate of 75% obtained using the SFS feature selection, the kNN-none/Fisher classifier and the normal split 0.9 evaluation method. The least recognition rate of 68.75% is obtained when considering only skin conductivity signal in the analysis, and using the SBS feature selection, the LDA-none classifier and the normal split 0.9 evaluation method.

For the 3-category-class defined by the valence dimension (V: +/-/0), the classification rates range from 50% to 76.67%. The highest recognition rate of 76.67% appears when fusing all physiological signals together. This result is obtained using the SFS feature selection, the LDA-none classification algorithm and the normal split 0.75 evaluation method. The second highest recognition rate of 70.83% is obtained when only considering the respiration signal, using the SBS feature selection, the kNN-none classification algorithm and normal split 0.9 evaluation method. The least recognition rate of 50% is obtained when only considering the EMG-Zyg signal channel and applying the kNN-Fisher classification algorithm combined with the SFS feature selection and the normal split 0.9 evaluation method.

For the 5-category-class defined by the Valence and Arousal dimensions (VA: 0-/++/-+/+-/--), the range of classification rates varies from 34% to 50%. The highest recognition rates of 50% appears twice when considering only the respiration signal or the EMG-Corrugator combined with EMG-Zygomaticus muscle signals together. This result is obtained when using the kNN-none classifier combined with the SBS feature selection and the normal split 0.9 evaluation method or using the LDA-none classification algorithm combined with the SFS feature selection and the normal split 0.75 evaluation method, respectively. The least recognition rate of 34% is obtained when only considering the skin conductivity signal and applying the SFS feature selection approach combined with the LDA-none classification algorithm and the normal split 0.75 evaluation method.

For the 10-category-class of Valence-Arousal-Dominance dimensions, the range of classification rates varies from 25% to 36%. The highest recognition rate of 36% is obtained when applying the skin conductivity and respiration signals together. This result is obtained when using the SFS feature selection combined with the LDA-none classification algorithm and the normal split 0.75 evaluation method. The least recognition rate of 25% appears twice when considering only skin conductivity with kNN-none/Fisher classifier or only EMG-Zygomaticus signal with kNN-none classification algorithm. Both least rates are obtained when using the SFS feature selection technique and the normal split 0.9 evaluation method.

Table 6. Intra-individual classification results with the highest rates obtained for the different category-classes und using various signal combinations.

Cat\Signals	SC	RSP	SC & RSP	EMG-Cor	EMG-Zyg	EMG-Cor & Zyg	ALL signals
	68.75%	75%	89.58%	73.33%	72.5%	74.17%	80%
2Cat(A)	SBS	SFS	SBS	SFS	SBS	SBS	SFS
	LDA-none	kNN-none/Fisher	kNN-Fisher	LDA-none	LDA-none	LDA-none	LDA-none
	normal split 0.9	normal split 0.9	normal split 0.9	normal split 0.75	normal split 0.75	normal split 0.75	normal split 0.75
	66.67%	70.83%	66.67%	56.67%	50%	58.33%	76.67%
3Cat(V)	SBS	SBS	SFS	SFS	SFS	SFS	SFS
	kNN-Fisher	kNN-none	LDA-none	LDA-none/Fisher	kNN-Fisher	LDA-none/Fisher	LDA-none
	normal split 0.9	normal split 0.9	normal split 0.75	nomal split 0.75	normal split 0.9	normal split 0.9	normal split 0.75
	34%	50%	40%	48%	36%	50%	46%
5Cat (VA)	SFS	SBS	SFS	SBS	SFS	SFS	SFS
	LDA-none	kNN-none	LDA-none	LDA-none	LDA-none	LDA-none	LDA-none/Fisher
	normal split 0.75	normal split 0.9	normal split 0.75	normal split 0.75	normal split 0.75	normal split 0.75	normal split 0.75
	25%	26%	36%	26%	25%	32%	30%
10 Cat(VAD)	SFS	SFS	SFS	SFS	SFS	SFS	SBS
	kNN-none/Fisher	LDA-none	LDA-none	LDA-none	kNN-none	LDA-none	kNN-none/Fisher
	normal split 0.9	normal split 0.75	normal split 0.75	normal split 0.75	normal split 0.9	normal split 0.75	normal split 0.9

3.2.2 Inter-individual Classification

Table 7 illustrates the classification rates resulting from the inter-individual analysis. The mean of each individual classification rate of all 20 subjects is calculated for the five different category-classes. For each category-class, a total of eight different mean rates values are calculated representing the mean classification rates resulting from the different applied combinations of approaches. The combination of methods consists of: the LDA-none, LDA-Fisher, kNN-none and kNN-Fisher classifiers combined with both the SFS and the SBS feature optimization techniques. Further, the best mean rate results for each category-class are highlighted in the table.

Table 7. Inter-individual classification results in terms of mean rates ± standard deviation (std) for the five different category classes. The highest mean classification rates obtained from the applied combination of feature selection and classification approaches is highlighted.

Categories	Classification	SFS	SBS
2Cat(A)	LDA-none	62.5 ± 12.83	**69.17 ± 15.44**
	LDA-Fisher	55.83 ± 19.89	67.92 ± 18.59
	kNN-none	63.96 ± 11.00	61.88 ± 16.79
	kNN-Fisher	58.96 ± 13.06	61.25 ± 13.72
2Cat(D)	LDA-none	66.91 ± 12.79	**70.12 ± 18.96**
	LDA-Fisher	66.43 ± 14.76	55.48 ± 21.87
	kNN-none	59.05 ± 11.40	59.05 ± 12.94
	kNN-Fisher	67.02 ± 15.22	66.67 ± 17.39
3Cat(V)	LDA-none	**60.75 ± 14.60**	42.92 ± 14.62
	LDA-Fisher	57.83 ± 10.98	38.33 ± 17.40
	kNN-none	49.17 ± 14.78	47.5 ± 12.99
	kNN-Fisher	52.92 ± 16.72	43.75 ± 17.91
5Cat(VA)	LDA-none	**37 ± 13.80**	29.5 ± 12.34
	LDA-Fisher	31.5 ± 14.96	26.5 ± 13.08
	kNN-none	32 ± 10.56	26.5 ± 8.13
	kNN-Fisher	35 ± 15.39	28 ± 12.81
10Cat(VAD)	LDA-none	**21.5 ± 9.88**	15 ± 9.46
	LDA-Fisher	17.5 ± 10.70	10.5 ± 8.26
	kNN-none	**21.5 ± 11.82**	18 ± 9.51
	kNN-Fisher	19 ± 13.34	12.5 ± 6.39

For both 2-category-classes, of Arousal (A: +/-) and of Dominance (D: +/-), the best mean rate of 69.17% and of 70.12% respectively, are obtained using the LDA-none classification algorithm and the SBS feature optimization technique. As for the 3-category-class of Valence (V: +/-/0), the highest mean of recognition rate of 60.75% is obtained using the LDA-none classifier and the SFS feature selection. Similar for the 5-category-class of Valence-Arousal (VA: 0-/++/-+/+-/--), the highest recognition rate of 37% is also obtained using the LDA-none classifier and the SFS feature selection technique. Finally, for the 10-category-class including all picture-presentations, the highest mean of recognition rate of 21.5% is obtained using both the LDA-none and the kNN-none classification algorithms each combined with the SFS feature selection technique.

In summary, our best mean values of the classification results show individual recognition rates of 69.17%, 70.12%, 60.75%, 37% and 21.5% obtained for the 2-category-class of Arousal, 2-category-class of Dominance, 3-category-class of Valence, 5-category-class of Valence-Arousal and the 10-category-class of Valence-Arousal-Dominance, respectively. Two feature selection techniques and a variation of four classification algorithms are thereby applied. The best-mean rates were all obtained when applying the LDA classification algorithm, combined with either the SBS feature selection method for both the 2-category classes and combined with the SFS feature selection method for the higher (3-, 5- & 10-) category classes. All inter-individual classification results are obtained using the normal split evaluation method with 0.95 factor.

3.3. Subgroup Analysis

In terms of gender differences, the mean rates of the inter-individual classification for women (47.22%) was found to be significantly higher than that of men (31.06%) for the 3-category-class of Valence (V: +/-/0). The statistical analysis is conducted using the Mann-Whitney U Test (U = -2.594, p = 0.009). The classification results for this gender analysis are obtained when using the LDA-Fisher classification algorithm and the SBS feature selection technique for the computation of the individual recognition rates. The corresponding result is summarized in Table 8. For the other category-classes, no significant results could be found between the gender differences and the classification rates.

Table 8 Significance analysis between gender differences and inter-individual classification rates obtained for the 3-category-class of Valence.

Category-class & Classification Approach	Gender	Nr.	Mean ± Std.	Test
3Cat(V)	w	9	47.22 ± 7.21	Mann-Whitney U
SBS, LDA-Fisher	m	11	31.06 ± 20.10	U= -2.594; p= 0.009

As for the age analysis, the mean rates of the inter-individual classification of young subjects are found to be significantly higher than the mean rates of old subjects for both the 3-category-class of Valence (V: +/-/0) and the 2-category-class of Dominance (D: +/-). For the 3-category-class of Valence, the statistical analysis is conducted using the Mann-Whitney U Test (U = -2.179, p = 0.029) and the classification results are obtained using the SFS feature optimization technique and the LDA-Fisher classification algorithm. For the 2-category-class of Dominance, the statistical analysis is conducted using the Independent-Samples t-Test and the classification results are obtained when using the SFS feature selection technique combined with either the LDA-Fisher classifier (F = 0.68, p = 0.048) or the kNN-Fisher classification algorithm (F = 1.51, p = 0.018).
A summary of these significant age analysis results is shown in Table 9, including the obtained classification rates.

Table 9. Significance analysis between age differences and inter-individual classification rates obtained for the 3-category-class of Valence and the 2-category-class of Dominance.

Category-class & Classification Approach	Age	Nr.	Mean ± Std.	Test
3Cat(V) SFS, LDA-Fisher	young	10	63.34 ± 9.78	Mann-Whitney U
	old	10	52.33 ± 9.56	U = -2.179, p = 0.029
2Cat(D) SFS, LDA-Fisher	young	10	72.85 ± 11.07	Independent-Samples t-Test
	old	10	60.00 ± 15.67	F = 0.68, p = 0.048
2Cat(D) SFS, kNN-Fisher	young	10	74.76 ± 15.71	Independent-Samples t-Test
	old	10	59.28 ± 10.44	F = 1.5, p = 0.018

When both age and gender are considered for the subgroup analysis, the one-way ANOVA analysis indicates significant differences among the groups of the young-men, the young-women, the old-men and the old-women (F = 5.235, p = 0.01) for the 5-category-class of Valence-Arousal. The classification results are obtained when using the SBS feature selection and the kNN-none classifier. The results of this subgroup analysis are illustrated in Table 10.

Table 10. Significance analysis between age & gender differences and inter-individual classification rates obtained for the 5-category-class of Valence-Arousal.

Category/Method	Age-Gender	Nr.	Mean ± Std.	Test
5Cat(VA) SBS, kNN-none	young-men	5	36.00 ± 5.48	one-way ANOVA ; LSD F = 5.235, p= 0.010
	young-women	5	22.00 ± 4.47	
	old-men	6	23.33 ± 8.16	
	old-women	4	25.00 ± 5.77	

Further, the result of the post-hoc test (LSD) presented in Table 11 reveals that the mean of inter-individual classification rates of young men (36%) is significantly higher than that of the other groups.

Table 11. Significance analysis for the age & gender differences from the LSD post-hoc test.

	young-men	young-women	old-men	old-women
young-men		0.003[**]	0.004[**]	0.019[*]
young-women			0.731	0.487
old-men				0.687
old-women				

4. DISCUSSION

For the intra-individual classifications, we conducted similar analysis as in our previous studies and obtained comparable findings [28,34]. First, the classification results for all category-classes are above the probability for random hits due to chance (50%, 33.3%, 20% and 10%

respectively). Further, the best recognition rates of 89.58%, 76.67%, 50% and 36% for the 2-, 3-, 5- and 10-category-classes respectively, are obtained when considering more than one physiological signal. This is also consistent with our previous finding stating that multi-channels recognition improves the results compared to single channel performance. Further, when considering only one signal, the highest recognition is always obtained with the respiration, showing again this channel contributes the most to the classification rates. Further, except for the 2-category-class, the best results are derived from the LDA-none classifier combined with the SFS feature selection and the 0.75 normal split evaluation method. This is also in accordance with our previous performance evaluation study showing that the mentioned approaches together present a relatively robust combination leading to the best results.

As for the extended inter-individual classification analysis, the averages of each individual classification rate are also above the probability for random hits due to chance, but the performance of the inter-individual classification is not as good as the intra-individual one. Moreover, the standard deviations of the inter-individual classifications show high values, stating that large inter-individual variability is present between the subjects. These findings are reasonable due to the inter-individual differences in physiological responses and are consistent with several previous studies. They explain the fact why it is harder to get high-performing inter-individual classification compared to intra-individual classifications [8,9].

Further, feature analysis related to the amount of features selected for both the intra-individual and inter-individual classifications is presented for all the applied category-classes and shows large variations with respect to the adopted physiological signals and feature selection methods. An extended analysis related to the type of features used and selected for both intra- and inter-individual classifications and for the different combination of approaches is not explicitly presented here as it goes beyond the frame of this paper.

Finally, the objective of the subgroup analysis was to investigate the influence of gender and age differences on the physiological recognition rates. From the perspective of gender, women have higher classification rates and smaller deviations than men for the 3-category-class defined by the Valence dimension. This means that the physiological emotional reactions on the valence axis are distinguished easier for females than for males. One possible explanation is that women are more expressive than men. Women exhibit more positive expressions in response to a positive stimulus and more negative expressions in response to a negative stimulus, so that the variations of the EMG-Corrugator and EMG-Zygomaticus signals of women are greater than those of men [36]. Further, the present finding is also supported by Wilhelm et al.' study, stating that women have more respiratory reactivity than men in response to high arousal / negative stimulus, but smaller respiratory reactivity than men in response to high arousal / positive stimulus [19]. From the perspective of age, the averages of classification rates of young subjects are higher than those of old subjects for the 3-category-class of Valence and the 2-category-class of Dominance. This means that it is easier to recognize the physiological emotional reactions of young people than those of older people on the axis of Valence or Dominance. This finding can be explained and supported by several previous studies, which demonstrate that the magnitude of physiological changes related to emotions are much smaller for elderly subjects than for young subjects [15,16]. Moreover, the present study also explores the effect of gender and age combined on the recognition rates. The average of classification rates of young men are found to be significantly higher than the rates of the other groups for the 5-category-class. Similar results -though only associated to the arousal dimension- states that young men are much sensitive to arousal resulting in large physiological changes [16].

5. CONCLUSION

In the present study we present intra- and inter-individual emotion classifications using psychophysiological biosignals and investigate the effect of age and gender and their interaction on the classification rates of emotions. The best recognition rates are obtained using similar combination of feature selection techniques, classification algorithms and evaluation approaches as obtained in our previous performance evaluation study. The highest intra-individual recognition rates are obtained using multi-channel classifications fusing more than one physiological signal. Further, the respiration signal is found to contribute the most to the intra-individual classification. Moreover, we conduct extended inter-individual classifications to evaluate the individual differences during the emotion recognition process. The results show large inter-individual variability in physiological responses between the subjects as stated in various studies. Finally, the subgroup analysis explores the effect of gender and age on the classification rates. We found that classification rates of women are significantly higher than the rates of men for the 3-category-class of Valence. Also, young subjects have significantly higher rates than old subjects for the 3-category-class of Valence and the 2-category-class of Dominance. Moreover, the emotion recognition of young men performed the best among the other subgroups for the 5-category-class of Valence/Arousal. Based on the recognition differences related to gender and age, subject-independent models could be improved to provide better classification results. Also, collecting data from a larger sample of subjects is reasonable to validate the current results. Finally, including further modalities such as facial expression, body gestures, audios and videos with the psychophysiological signals would be beneficial to the emotion recognition in various human computer interaction applications.

REFERENCES

[1] D. Rösner, D. Hazer-Rau, C. Kohrs, T. Bauer, S. Günther, H. Hoffmann, L. Zhang, & A. Brechmann, "Is There a Biological Basis for Success in Human Companion Interaction?" In: Kurosu M. (Ed), Human-Computer Interaction, HCI 2016, Part I, Lecture Notes in Computer Science LNCS, vol. 9731, p. 77–88, Springer, 2016.

[2] R.A. Calvo, S. D'Mello, "Affect Detection: An interdisciplinary Review of Models, Methods, and Their Applications." IEEE Transactions on affective computing, 1(1), 2010.

[3] P. Patel, A. Chaudhari, R. Kale, M.A.Pund, "Emotion recognition from speech with gaussian mixture models & via boosted GMM" International Journal of Research In Science & Engineering IJRISE, Vol. 3, Issue 2, 2017.

[4] I. Mohino-Herranz, R. Gil-Pita, S. Alonso-Diaz, M. Rosa-Zurera, "MFCC Based Enlargement of the Training Set for Emotion Recognition in Speech." Signal & Image Processing: An International Journal (SIPIJ), vol.5, No.1, 2014.

[5] A. Pentland, & S. Pentland, Honest signals: how they shape our world. Bradford Books. MIT Press, 2008.

[6]   N. Jatupaiboon, S. Pan-Ngum, & P. Israsena, "Subject-dependent and subject-independent emotion classification using unimodal and multimodal physiological signals." Journal of Medical Imaging and Health Informatics, 5(5), 1020-1027, 2015.

[7]   M. Ali, F. Al Machot, A.H. Mosa, & K. Kyamakya, "CNN Based Subject-Independent Driver Emotion Recognition System Involving Physiological Signals for ADAS." In Advanced Microsystems for Automotive Applications, 125-138, 2016.

[8]   J. Chen, B. Hu, Y. Wang, P. Moore, Y. Dai, L. Feng, & Z. Ding, "Subject-independent emotion recognition based on physiological signals: a three-stage decision method." BMC medical informatics and decision making, 17(3), 167, 2017.

[9]   F. Al Machot, M. Ali, S. Ranasinghe, A.H. Mosa, & K. Kyandoghere, "Improving Subject-independent Human Emotion Recognition Using Electrodermal Activity Sensors for Active and Assisted Living", 2018.

[10]   R. Böck, S. Glüge, A. Wendemuth, K. Limbrecht, S. Walter, D. Hrabal, & H.C. Traue, "Intraindividual and interindividual multimodal emotion analyses in human-machine-interaction." In Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2012 IEEE International Multi-Disciplinary Conference, pp. 59-64, 2012.

[11]   M.M. Bradley, M. Codispoti, D. Sabatinelli, & P.J. Lang, "Emotion and motivation II: sex differences in picture processing." Emotion, 1(3), 300, 2001.

[12]   D. Hazer, X.Y. Ma, S. Rukavina, S. Gruss, S. Walter, & H. C. Traue, "Emotion elicitation using film clips: Effect of age groups on movie choice and emotion rating." . In: Stephanidis C. (Ed.), Human-Computer Interaction, HCI 2015, Part I, Communications in Computer and Information Science CCIS, vol. 528, p. 110–116, Springer, 2015.

[13]   C. Lithari, C.A. Frantzidis, C. Papadelis, A.B. Vivas, M.A. Klados, C. Kourtidou-Papadeli, C. Pappas, A.A. loannides, & P.D. Bamidis, "Are females more responsive to emotional stimuli? A neurophysiological study across arousal and valence dimensions." Brain topography, 23(1), 27-40, 2010.

[14]   D. Sabatinelli, T. Flaisch, M.M. Bradley, J.R. Fitzsimmons, & P.J. Lang, "Affective picture perception: gender differences in visual cortex?" Neuroreport, 15(7), 1109-1112, 2004.

[15]   R.W. Levenson, L.L. Carstensen, W.V. Friesen, & P. Ekman, "Emotion, physiology, and expression in old age." Psychology and aging, 6(1), 28, 1991.

[16]   L. Burriss, , D.A. Powell, , & J. White, "Psychophysiological and subjective indices of emotion as a function of age and gender." Cognition and emotion, 21(1), 182-210, 2007.

[17]   T. Vogt, & E. André, "Improving automatic emotion recognition from speech via gender differentiation." In Proc. Language Resources and Evaluation Conference (LREC 2006), Genoa, 2006.

[18]   S. Rukavina, S. Gruss, H. Hoffmann, J.W. Tan, S. Walter, & H.C. Traue, "Affective computing and the impact of gender and age." PloS one 11.3, 2016.

[19]   F.H. Wilhelm, J.A. Rattel, M. Wegerer, M. Liedlgruber, S. Schweighofer, S.D. Kreibig, V. Kolodyazhniy, & J. Blechert, "Attend or defend? Sex differences in behavioral, autonomic, and respiratory response patterns to emotion–eliciting films." Biological psychology 130 : 30-40, 2017.

[20]   S. Gupta, J. Jaafar, W.F. wan Ahmad, A. Bansal, "Feature Extraction using MFCC." Signal & Image Processing: An International Journal (SIPIJ), vol.4, No.4, 2013.

[21] R.W. Picard, "Affective computing: challenges." International Journal of Human-Computer Studies, 59(1), 55-64, 2003.

[22] S. A. Fattah, Md. A. Iqbal, M. A. Jumana, A.B.M. S. Ud Doulah, "Identifying the Motor Neuron Disease in EMG Signal using Time and Frequency Domain Features with Comparison." Signal & Image Processing: An International Journal (SIPIJ), vol.3, No.2, 2012.

[23] Y.L. Hsu, J.S. Wang, W.C. Wang, C.H. Hung, "Automatic ECG-Based Emotion Recognition in Music Listening." IEEE Transactions on Affective Computing, 2017.

[24] J.A. Healey and R.W. Picard, "Detecting stress during realworld driving tasks using physiological sensors," IEEE Transactions on Intelligent Transportation Systems, Vol. 6, 156-166, 2005.

[25] W. Wei, Q. Jia, Y. Feng, G. Chen, " Emotion Recognition Based on Weighted Fusion Strategy of Multichannel Physiological Signals." Computational Intelligence and Neuroscience, Article ID: 5296523, 2018.

[26] J. Kim, "Bimodal emotion recognition using speech and physiological changes," in Robust Speech Recognition and Understanding, M. Grimm and K. Kroschel, Eds. I-Tech Education and Publishing, Vienna, Austria, pp. 265–280, 2007.

[27] C. Frantzidis, C. Bratsas, C. Papadelis, E. Konstandinidis, C. Pappas, and P. Bamidis, "Towards emotion aware computing: An integrated approach using multi-channel neuro-physiological recordings and affective visual stimuli," IEEE Trans. on Information Technology in Biomedicine, Vol. 14, No. 3, pp. 589, 2010.

[28] D. Hazer-Rau, L. Zhang, , & H. C. Traue, "Performance evaluation of various emotion classification approaches from physiological signals." International Journal of Artificial Intelligence and Applications (IJAIA), Vol.9, No 4, pp. 31-41, 2018.

[29] P. Lang, M. Bradley & B. Cuthbert, "International affective picture system (iaps): Affective ratings of pictures and instruction manual," University of Florida, Gainesville, FL, USA, Tech. Rep., 2008.

[30] S. Walter, J. Kim, D. Hrabal, S.C. Crawcour, H. Kessler, & H.C. Traue, "Transsituational individual-specific biopsychological classification of emotions." Systems, Man, and Cybernetics: Systems, IEEE Transactions on, 43(4), 988-995, 2013.

[31] G. Valenza, A. Lanata, and E.P. Scilingo, "The Role of Nonlinear Dynamics in Affective Valence and Arousal Recognition". IEEE Transactions on Affective Computing, ParePrints. 99, 2011.

[32] J. Wagner, Augsburg biosignal toolbox (Aubt) – User Guide. University of Augsburg, 2006.

[33] J. Wagner, J. Kim, E. André, "From Physiological Signals to Emotions: Implementing and Comparing Selected Methods for Feature Extraction and Classification." In IEEE International Conference on Multimedia & Expo (ICME), 2005.

[34] L. Zhang, S. Rukavina, S. Gruss, H.C. Traue, D. Hazer, "Classification analysis for the emotion recognition from psychobiological data." In: International Symposium on Companion-Technology, ISCT, IEEE, 2015.

[35] A. A. Afifi, & S. P. Azen, Statistical analysis: a computer oriented approach. Academic press, 2014.

[36] A.M. Kring, & A.H. Gordon, "Sex differences in emotion: expression, experience, and physiology." Journal of personality and social psychology, 74(3), 686, 1998.