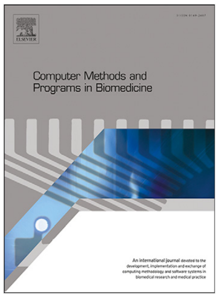
Journal Pre-proof



Stress recognition identifying relevant facial action units through explainable artificial intelligence and machine learning

Giorgos Giannakakis, Anastasios Roussos, Christina Andreou, Stefan Borgwardt, Alexandra I. Korda

PII: S0169-2607(24)00500-5
DOI: https://doi.org/10.1016/j.cmpb.2024.108507
Reference: COMM 108507

To appear in: *Computer Methods and Programs in Biomedicine*

Received date : 14 December 2022
Revised date : 5 November 2024
Accepted date : 6 November 2024

Stress recognition identifying relevant facial action units through explainable artificial intelligence and machine learning

Giorgos Giannakakis[a,b,c], Anastasios Roussos[a,d], Christina Andreou[e], Stefan Borgwardt[e], Alexandra I. Korda[e]

[a] *Institute of Computer Science, Foundation for Research and Technology Hellas (FORTH), N. Plastira 100, Heraklion, 70013, Crete, Greece*
[b] *Department of Electronic Engineering, Hellenic Mediterranean University, Chania, 73133, Greece*
[c] *Institute of Agri-food and Life Sciences, University Research and Innovation Center, Hellenic Mediterranean University, Heraklion, 71003, Greece*
[d] *College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK*
[e] *Translational Psychiatry, Department of Psychiatry and Psychotherapy, University of Luebeck, Ratzeburger Allee 160, Lübeck, 23562, Germany*

Abstract

Background and objective: Facial cues and expressions constitute a component of bodily responses that provide useful information about one's stress levels. According to the Facial Action Coding System, they can be modelled consistently in terms of fundamental facial muscle movements, called facial Action Units (AUs). This article investigates automatic acute stress recognition based on AUs using conventional Machine and Deep Learning techniques.
Methods: We created a new experimental dataset containing 58 participants performing 4 experimental phases and 11 stress and non-stress tasks in which the proposed system performs automatic facial AUs recognition. A computational feature selection method was employed to select a robust relevant AU combinations subset, which integrated with conventional Machine Learning and Deep Learning methods using the Layer-Wise Relevance Propagation algorithm to assess and model the implication of AUs under acute stress conditions. Ordinal modelling was used following the pairwise transformation to establish a common reference based on the personalized values of each participant.
Results: The results indicate that, under acute stress conditions, participants' faces presented significantly more AUs and with greater intensity compared to neutral conditions. The most relevant combination of AUs to each stress type was computationally identified, ranked and selected. The mean yielded classification accuracy of stress condition versus neutral achieved across all experimental tasks was greater than 93%.
Conclusions: There are specific combinations of AUs that are relevant to the stress conditions of each experimental phase leading in each case to better neutral and stress separability.

Keywords: stress, facial action units (AU), machine learning, deep learning, explainable artificial intelligence

1. Introduction

Stress estimation is a demanding task that can be evaluated utilizing multiparametric behavioural and physiological data sources. The majority of the relevant literature focuses on the investigation of stress-related physiological bodily response as expressed in biomarkers (e.g. hair/saliva cortisol, corticotrophin-releasing factor (CRF), adrenocorticotropin (ACTH)) [1, 2] and biosignals (e.g. ECG, EDA, respiration, EMG, etc.) [3, 4, 5, 6, 7, 8]. Apart from the physiological modalities, there is also an emerging research interest in Facial Expression Recognition (FER) technology which can assess facial manifestation of emotions by analysing facial behaviour [9].

Facial expressions are voluntary manifestations and have therefore been criticized for being potentially misleading about a person's actual emotional state, as they can be manipulated or hidden [3]. Various interesting approaches have been developed for FER [10, 11, 12]. However, there needs to be a clear research direction for the facial expression analysis towards affect estimation. Though, over the last years, there has been an attempt to identify objective in- or semi-voluntary facial parameters (such as micro-expressions, blinks, mouth micro-activity) for stress estimation[13, 14, 15, 16].

A consistent and objective method for analysing facial expressions is through the *Facial Action Coding System* (FACS) [17] and its extension, the *Emotion Facial Action Coding System* (EMFACS) [18] which correlates emotions and its corresponding facial parameters. Facial Action units (AUs) [19] are the fundamental actions of individual facial muscles or groups of muscles whose combination provides information about one's affective state. Facial action units (AUs) analysis has been widely adopted for affective computing applications, and many studies have developed various methods addressing the accurate estimation of AUs and their decoding into emotions [19][20]. The difficulty of a precise and reliable AUs estimation model has been addressed widely [19]. However, few studies applied AUs methodologies for modelling stress, so the behaviour and involvement of AUs under stress conditions have yet to be investigated thoroughly. Defining a reliable ground truth in estimating stress levels is still challenging, as it presents great inter-subject and inter-task variability.

In the literature, stress modelling was usually performed with conventional Machine Learning (ML) models. In recent years, deep learning (DL) techniques have been applied to investigate stress, achieving higher accuracy performances, but presenting limitations in the model's generalizability and transparency [21]. Lately, the related literature has adopted deep learning methods for facial expressions/cues/action units and corresponding emotion recognition [22, 23, 24, 25, 26, 27]. In [25], facial AUs were estimated and fed 4 types of artificial neural networks (both shallow and deep learning) (FFNN, RBFNN, RNN and CNN), leading to a stress classification accuracy of 90.2% with the intra-subject methodology with the conventional FFNN and SVM classifiers outperforming all other methods. In the work of Viegas et al. [28], AUs intensities over time were used as features combined with a Random Forest classifier, achieving an average stress recognition accuracy of 75% for person-independent and 93% for person-dependent analysis. In [29], facial AUs intensities along with stress models based on machine learning classification were utilized, leading to a binary classification accuracy between no stress/stress of 74.6%. Besides, in [12], a real-time facial expression recognition system based on a 3D Morphable Model and Deep Convolutional Neural Networks was established and applied effectively to recognize stress states. Jeon et al. [23] used a spatial attention module (providing high weight to the stress-related facial regions), achieving discriminating accuracy of 66.8%. Zhang et al. [24], developed a two-level network (TSDNet) and tested it in a database of 2092 labelled video clips for stress conditions, leading to a detection accuracy of 85.42%.

A model's non-transparency constitutes a crucial concern for the research questions under investigation. Thus, explainable AI systems have been employed to overcome this issue. Conventional ML methods are well established and widely approved, on the other hand, DL methods yield outstanding results and can also be equipped with explainability thanks to the advances in the field of explainable AI [30]. In [31], authors proposed a hybrid explainable AI framework for identifying expressions, incorporating the explainability and the AUs extraction. Explainable AI is a type of feature ranking method applied in deep neural networks, and indicates the features that contribute more to the classification decision.

This study investigated acute stress recognition through facial AUs from both conventional ML and DL approaches. We focused on acute stress, which can be defined as the short-term psychological state, including feelings of fear, worry, and irritability, triggered by the exposure to a variety of distressing or challenging tasks (e.g., mental arithmetic, speech tasks, and distressing films) or life conditions [32]. However, it should be noted that a stressor can cause subjective emotional response depending on the context, previous subject's learning, and appraisals [33]. It has been stated in our previous studies that emotion recognition using AUs/FACS coding and AI methods provides more accurate results than conventional ML methods [29, 25, 28]. The aim of this study is to reveal which facial AU combinations contribute to the recognition of the stress for each experimental stressor and their consistency across algorithms and experimental phases. The most robust and relevant to each stress type AUs were selected, and established a stress model using conventional ML and DL methods. Various stress models were evaluated by tuning and optimizing the AUs inputs, the model type/parameters/architecture, as well as their performance in terms of stress/no stress state classification accuracy.

2. Methods

This study investigated automatic acute stress detection based on facial AUs and machine/deep learning techniques. The flowchart of the proposed automatic acute stress detection system is presented in Fig. 1. The facial videos were input into our system and for each video frame, a pipeline of facial landmarking and AUs intensities extraction (per-frame AUs feature vector) was performed. The clinical protocol and the study procedure are described in the Supplemental Material. The AUs feature vectors were input into the ML/DL module for stress recognition, which in turn outputs a binary classification of the stress state.

2.1. Experimental procedures

An experimental protocol was developed to investigate facial behaviour under acute stress conditions. The experiment included both neutral and stress-inducing tasks from different stressor types. These stressors can be divided into 4 distinct phases: *social exposure*, *emotional recall*, *mental workload*, *stressful videos stimuli*. The experimental phases, tasks durations and induced affective condition are presented in Table 1.

Table 1: Experimental tasks employed in this study

#	Experimental task	Duration (min)	Affective State
Social Exposure			
1	1.1 Neutral (reference)	2	N
2	1.2 Baseline Description	2	N
3	1.3 Interview	2	S
Emotional recall			
4	2.1 Neutral (reference)	2	N
5	2.2 Recall stressful event	2	S
Mental Workload			
6	3.1 Reading letters/numbers (reference)	2	N
7	3.2 Stroop Colour-Word Test (SCWT)	2	S
8	3.1 PASAT task	2	S
Stressful stimuli			
9	4.1 Relaxing video	2	R
10	4.2 Adventure video	2	S
11	4.3 Psychological pressure video	2	S

Note: Intended affective state N:neutral, S:stress, R:relaxed)

The social exposure phase included an interview with a psychologist in which the participants described themselves, in which the psychologist emphasized on negative aspects of their trait. Its corresponding reference state for this phase was the participant saying conventional words (e.g. count from one to

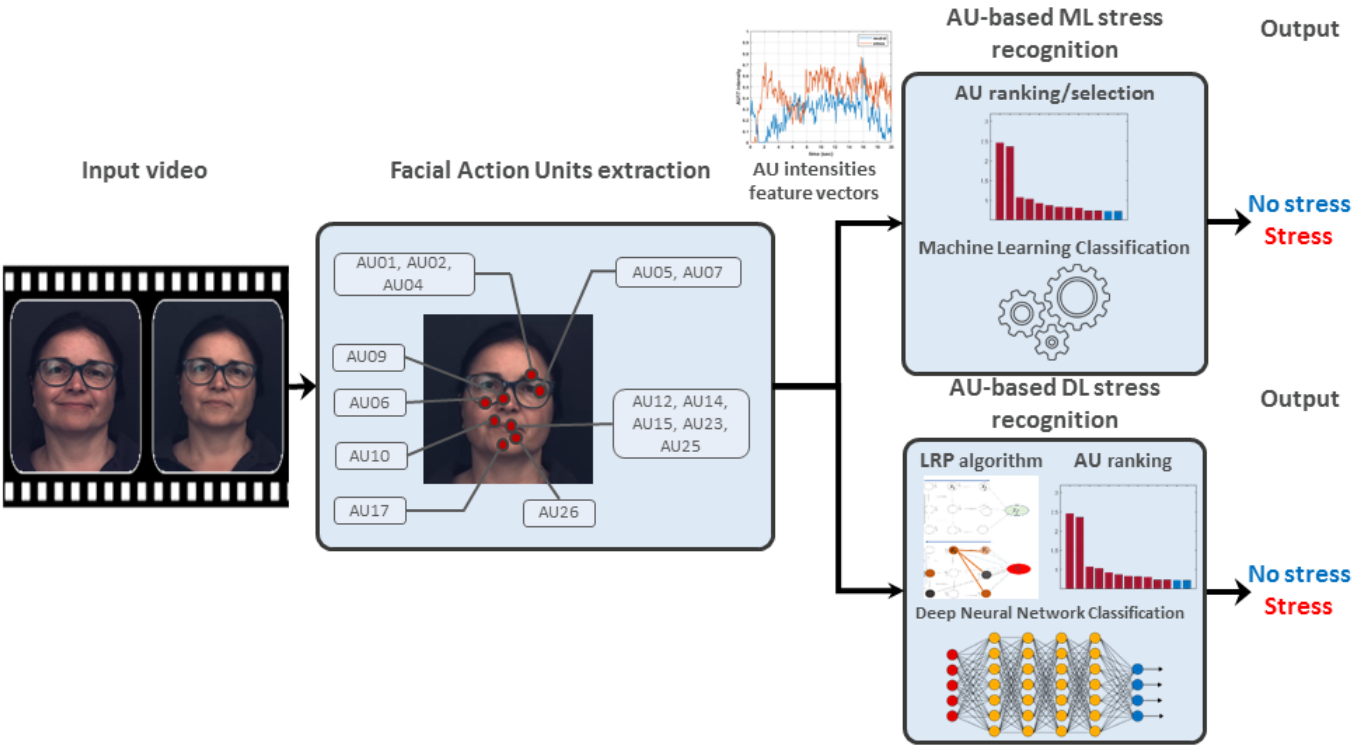


Figure 1: Flowchart of the proposed facial AUs system for stress detection based on feature selection/Machine Learning (upper part of the figure) and on explainable AI/Deep Neural Network classification (lower part of the figure). The input video was fed into the system, the facial landmarks (see Section 2.4) and the timeseries of the facial AUs intensities were extracted (see Section 2.5). These AUs intensities were fed into the AUs ranking/selection mRMR algorithm, Machine Learning Classification (see Section 2.8) (upper part of the figure) or AUs ranking (LRP algorithm), Deep Neural Network classification (see Section 2.9) (lower part of the figure) providing a binary classification of the subject's stress state.

ten, months of the year, etc.). The emotion recall phase included stress elicitation by asking participants to recall and relive a stressful event from their past as if it was currently happening. The mental tasks phase included cognitive load assessment through tasks such as the modified Stroop colour-word task (SCWT) [34], requiring participants to read colour names (red, green, and blue) printed in incongruous ink (e.g., the word RED appearing in blue ink). In the present task, difficulty was increased by asking participants first to read each word and then name the colour of the word. A second mental task used was the Paced Auditory Serial Addition Test (PASAT) [35], which is a neuropsychological test with arithmetic operations for assessing attentional processing.

The stressful videos stimuli phase included video segments presentation in order to induce a relaxed (calming video) and stressful conditions (adventure film scenes, videos with scenes from very high positions (acrophobia), scene with home invasion, car accidents etc.).

The experimental tasks described in this section are well-established in the literature and they are considered to induce acute stress effectively. We have used similar experimental protocols in our previous studies and datasets [13, 14, 29, 36], providing evidence of effective stress induction manifested in different modalities such as facial cues/blinks/AU [13, 14, 29], head pose [36]. We have also concurrently recorded participants' Heart Rate (HR) derived placing an ECG sensor in symmetric position of the chest (right and left of the sternum, corresponding to the V1 and V2 leads) providing a typical and low SNR ECG recording [37]. HR was statistically significant increased during acute stress conditions in relation to their neutral (reference) experimental task, in all experimental phase pairs except the pair relaxing video-psychological pressure video where there was no significant difference. Besides, the participants' self-reports of perceived emotional arousal and valence ratings in a [1-9] scale were performed using the Self-Assessment Manikin (SAM) scales [38]. In order to validate the induced stress emotion, we evaluated the arousal/valence ratings of each stressful experimental task in relation to the reference (neutral) task using Repeated Measures ANOVA. The results, which are presented in the Appendix, indicate that in all experimental tasks, the perceived arousal was statistically significant increased, and the perceived valence was statistically significant decreased during the stressful tasks in relation to their corresponding neutral experimental tasks.

2.2. Study Dataset

The inclusion criteria of this study were that the participants should be above 18 years old, and the exclusion criteria were a history of heart disease and not signing the study's informed consent. The population of our experimental dataset was 58 adults (24 men, 34 women) aged 26.9±4.8 years. For each participant, 11 tasks (4 neutral, 6 stressed and 1 relaxed state) were

Table 2: Summary of the AUs, their FACS name and muscular basis investigated in this study

AUs	FACS name	Muscular basis
AU1	Inner brow raiser	frontalis (pars medialis)
AU2	Outer brow raiser	frontalis (pars lateralis)
AU4	Brow lowerer	depressor glabellae, depressor supercilii, corrugator supercilii
AU5	Upper lid raiser	levator palpebrae superioris, superior tarsal muscle
AU6	Cheek raiser	orbicularis oculi (pars orbitalis)
AU7	Lid tightener	orbicularis oculi (pars palpebralis)
AU9	Nose wrinkler	levator labii superioris alaeque nasi
AU10	Upper lip raiser	levator labii superioris, caput infraorbitalis
AU12	Lip corner puller	zygomaticus major
AU14	Dimpler	buccinator
AU15	Lip corner depressor	depressor anguli oris (triangularis)
AU17	Chin raiser	mentalis
AU20	Lip stretcher	risorius
AU23	Lip tightener	orbicularis oris
AU25	Lips part	depressor labii inferioris, or relaxation of mentalis or orbicularis oris
AU26	Jaw drop	masseter; relaxed temporalis and internal pterygoid
AU45	Blink	relaxation of levator Palpebrae and Contraction of Orbicularis Oculi

performed. Videos were sampled at 60 fps with a video resolution of 1216×1600 pixels which was subsampled to 30 fps and a resolution of 608×800 pixels. A neutral condition was presented at the beginning of each phase of the experiment. This condition was used as a baseline for the subsequent stressful tasks. All participants provided informed consent. The study was approved by the Research Ethics Committee of FORTH (approval no. 155/12-09-2022).

2.3. FACS coding

The *FACS* [17, 39] categorizes fundamental movements of human facial muscles based on anatomic functions. It consists of 32 distinct fundamental facial muscle movements called Action Units (AUs). In this study, 17 AUs were investigated (due to the fact that only these AUs were included in the training datasets) for their stress state discriminatory ability, which are presented in Table 2.

2.4. Preprocessing

The preprocessing phase included processes of face detection, facial landmarks estimation and removal of facial rigid information. The Viola-Jones detector [40] was used for face detection. The landmark estimation was performed using 3D aware 2D landmarks extracted with the Deng et al.'s Cascade Multi-view Hourglass Model [41]. The 68 facial landmarks derived from this model are projections of their corresponding 3D points on the image plane. For the facial alignment (removal of the rigid information), we applied Procrustes analysis [42] so as to find the 2D similarity transform that aligns the facial landmarks of each frame to the mean face landmarks. This similarity transform is then applied to the input image using Delaunay triangle-based affine warp [43].

2.5. Facial AUs extraction

The facial AUs were extracted utilizing both geometric and appearance features to improve the AUs estimation [19, 29]. For the geometric features, the non-rigid 3D landmarks were estimated as described in Section 2.4. The removal of the rigid information provides reliable information about facial parts' shape deformation coming only from facial expressions and not from head movements. The deformable shapes are modelled using a linear Point Distribution Model (PDM) [44] providing a parametric representation. As geometric features, we used the non-rigid 3D facial landmarks and as appearance features, we used Histograms of Oriented Gradients (HOG) [45] on the aligned/warp face according to the base face shape extracting dimensional histograms of blocks with a cell size of 2x2 and 8x8 pixels. Principal Component Analysis (PCA) was applied to reduce data dimensionality retaining the components explaining the 95% of the total data variability to be used for the appearance features vector.

Both geometric and appearance features constituted the feature matrix that fed a Support Vector Regression model (SVR) [46]. The SVR model's regression on data and its corresponding labels led to the annotated AUs intensities. The training of the SVR model was performed utilizing two available facial datasets (UNBC [47] and BOSPHORUS [48]). In our previous study [29], we thoroughly evaluated the quality of the detected AU of our model in terms of 10-fold cross validation accuracy measure using two available and annotated facial datasets, the UNBC-McMaster Shoulder Pain Expression Archive Database (UNBC) and the Bosphorus database (BOSPHORUS) providing reliable AU performance (79.9% mean AU detection accuracy) as presented in the figure below. The achieved validation performance was similar (depending on the AU) to related studies in the literature [49, 50, 51].

2.6. AUs Feature selection

After extracting the AUs feature matrix, the most significant and relevant AUs to the stress conditions were identified. The mRMR algorithm [52] evaluated and ranked the features according to their relevance to the problem based on maximal relevance and minimum redundancy optimizing in terms of the Mutual Information Quotient (MIQ) criterion [53] as presented in the equation

$$\max_{x_j \in x - S_{m-1}} \left[I\left(x_j; c\right) - \frac{1}{m-1} \sum_{x_j \in S_{m-1}} I\left(x_j; x_i\right) \right],$$

where $m-1$ features are selected from the X feature matrix, forming the feature subset S_{m-1} in order to select the next most relevant feature, c is the class labels and I is the mutual information (MI) function which is presented in the following equation

$$I(x; y) = \iint p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dxdy$$

where $p(x, y)$ is the joint probabilistic density function, $p(x)$ and $p(y)$ are the marginal probabilistic density functions of the variables x,y.

The retained AUs features subset was determined by minimizing the classification error between neutral and stress states of each phase using the 10-fold cross-validation with SVM classifier (linear and RBF kernel).

2.7. Pairwise transformation and normalization

As there is great inter-subject variability in AUs manifestation (as well as to other body stress responses), it is appropriate to consider each participant's personalized values to have a common reference for the subsequent analysis. This neutral task of each experimental phase corresponds to each subject's baseline for each experimental phase. The pairwise transformation provides a mapping taking into account the personalized participants' baseline values, so as to establish a common reference to each feature across subjects providing data normalization.

In this case, the problem of stress detection can be viewed as a ranking problem. In order to transform it into a 2-class classification problem (classes: no stress vs stress), we use the pairwise transformation introduced in [54, 55]. The pairwise transformation which maps the features matrix X (consists of AUs intensities (AU1, AU2, ..., AU45) for the different video recordings of this study as described in Section 2.5) and the class labels Y is described by the equation:

$$T : \left\{ \begin{array}{l} X' = X(t_i) - X\left(t_j\right) \\ Y' = \operatorname{sign}\left\{ \mathrm{Y}(t_i) - Y\left(t_j\right) \right\} \end{array} \right\}, \forall \text{ corresponding } i, j$$

where i, j refer to the indices of neutral and stress states respectively with all possible pairs of a specific subject of the feature matrix. The overall transformation procedure is described in Algorithm 1.

Algorithm 1: Pairwise transformation used in this study

Input:
X – feature matrix [cases x features]
Y – classes [1: non-stress, 2: stress]
Output:
X' – pairwise transformed feature matrix
Y' – classes [-1,1]
for each extracted data **do**
X_1, X_2 feature vectors of class Y_1, Y_2 respectively
Find indices i, j of all permutations without repetition of X_1, X_2
for each pair i, j **do**
$X' = X_1(i) - X_2(j)$
if $Y_i > Y_j$ then $Y' = 1$
if $Y_i < Y_j$ then $Y' = -1$
end
end

This transformation creates preference pairs of feature vectors $X(i)-X(j) = [f_1(i)-f_1(j), \ldots, f_M(i)-f_M(j)]$ and their labels $\operatorname{sign}\{Y(i) - Y(j)\}$. If $Y(i) > Y(j)$ then $X(i) > X(j)$ and this preference pair is a positive instance, otherwise, it is a negative instance $X(i) < X(j)$. The preference pairs and their corresponding labels after transformation can be considered as instances and labels in a new classification problem which then can be performed with conventional classification schemes. This step is significant for the subsequent analysis as it addresses the inter-subject variability taking into account the baseline of each subject of the neutral tasks.

2.8. Machine Learning (ML) classification

Classifiers based on conventional Machine Learning (KNN, Naive Bayes (NVB), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), SVM linear, SVM quadratic, SVM cubic, SVM Gaussian and ensemble bagged trees) were employed to evaluate the proposed system in terms of its discriminability between non-stress and stress states using The classification accuracy was evaluated by a 10-fold cross-validation technique.

2.9. Deep Learning (DL) Classification

Using a deep learning approach, we also tested the ability of the two emotional states (neutral and stress) to discriminate. The classifier implemented was a neural network-based classifier presented in our previous work [56]. The network used the hyperbolic tangent sigmoid transfer function and was batch-trained using the Levenberg-Marquardt training algorithm [57] and implemented in MATLAB 2020b. L2-regularization was applied to access possible types of uncertainty. We selected parameters after experimentation; 5 hidden layers (tested 2 to 5), each hidden layer consisting of 2 nodes (tested 2 to 20) and 1.000 epochs. Hyperparameter tuning methods include grid search and the learning rate was 1e-3 (tested 1e-5 to 1e-2) and 1 batch size (tested 1 to 10). The training procedure followed two cycles of 10-times repeated nested cross-validation with 10 folds in the inner cycle and 10 folds in the outer cycle, resulting in 10,000 models. The average accuracy, sensitivity and specificity were calculated across all hold-out datasets of the 10x10 nested cross-validation, repeated 10 times. The hyperparameters of the network were fine-tuned based on the average accuracy, sensitivity and specificity.

To perform localization, we calculated the relevance of the AUs in each condition using the LRP algorithm for multilayer neural networks, as described in Bach et al. [58]. LRP, in its general form, assumes that the classifier can be decomposed into several layers of computation. Such layers can be parts of the feature extraction from the image or parts of a classification algorithm that runs on the computed features. The LRP algorithm has been used on brain MRI prediction models for the identification of biomarkers in schizophrenia and depression [56] and in EEG analysis for the prediction of schizophrenia [59]. The explanation given by LRP would be a map showing which AUs of the original face map contributed to the diagnosis and to what extent (see Figure 3). The algorithm used parameters such as the weights and the activations of the nodes in the hidden layers to back-propagate the classification decision back to the initial space. For the specific deep learning scheme with 5 hidden layers with size 2, the LRP algorithm is presented in the Appendix.

2.10. Evaluation of AUs detection accuracy

We used the classification accuracy, sensitivity and specificity metrics as presented in the following equations

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

where TP, TN, FP and FN are the number of true positives, true negatives, false positives and false negatives respectively.

3. Results

The methodology described in Section 2 was applied to the study's dataset, which is described in 2.2. The time series of AUs intensities were extracted for the AUs described in Section 2, and their average values were calculated for each subject and experimental task.

3.1. AUs-based ML stress recognition

The pairwise transformation was applied as described in Section 2.7, to ensure a common reference model for our analysis. Then, the mRMR ranking algorithm was applied to select the most robust and relevant to the problem feature subset. The top-ranked features were inserted iteratively in the feature subset, evaluating each candidate subset's performance in terms of a 10-fold SVM classification accuracy used as the objective function. The features retained in descending order according to this procedure are presented in Table 3.

It can be observed that in all experimental tasks, a large number of AUs needed to be selected (above 12 AUs) to achieve a good classification performance.

The selected features subset derived from the mRMR algorithm was evaluated for its discriminatory ability between non-stress and stress conditions. A 10-fold cross-validation technique was employed and the classifiers described in 2.8 were used. The classification results are summarized in Tables 4 and 5.

The pairwise transformation, which is a robust method to incorporate the participant's personalized control values and deal with inter-subject variability, significantly increased the discriminatory ability in all experimental phases. The exposure and the mental workload experimental phases presented the higher mean classification accuracies with percentages of 98.7% and 93.6% respectively, as compared with the emotional recall and stressful stimuli, which presented lower classification accuracies (85.6% and 87.3% respectively).

Table 3: Relevant AUs implicated in stress conditions using mRMR feature ranking and selection algorithm. The significant selected features in descending order, as well as the number of retained features are reported.

Experimental phase	**Relevant and ranked AUs**	**# features retained**
exposure	AU45, AU17, AU09, AU23, AU10, AU15, AU04, AU14, AU12, AU26, AU01, AU05	12
emotional recall	AU15, AU04, AU17, AU45, AU05, AU01, AU26, AU23, AU02, AU09, AU07, AU25, AU10, AU20	14
mental workload	AU12, AU01, AU04, AU09, AU10, AU14, AU26, AU07, AU05, AU02, AU15, AU25 AU06, AU20, AU45, AU23	16
stressful stimuli	AU09, AU26, AU06, AU02, AU17, AU04, AU25, AU10, AU01, AU20, AU23, AU07, AU05, AU15, AU14, AU12, AU45	17
All tasks	AU09, AU04, AU14, AU02, AU15, AU06, AU10, AU20, AU01, AU12, AU17, AU07, AU05, AU26, AU25, AU23	16

3.2. AUs-based DL stress recognition

The Deep Learning Classification methodology, as described in Section 2.9, was applied to the feature matrix after pairwise transformation and normalization (Section 2.7). The training procedure followed a 10 times repeated 10x10 nested cross-validation process.

The LRP algorithm, as described in [58], was employed to evaluate the AUs relevances concerning acute stress conditions, providing for each AUs a value on the relevance to stress contribution. The mean relevance of all AUs per state was calculated in each task. The higher the AUs' relevance, the more significant the contribution of the AU in the classification decision. Negative relevance indicates that the AUs mislead the classification of stress versus neutral. The mean relevance of the corrected classified states is presented in Figure 2 and the ranked AUs in reverse order in terms of their mean relevance are presented in Table 6. The training and classification procedure was repeated iteratively by adding each time one feature of the reverse ranked order as determined by the LRP algorithm and evaluating the mean classification accuracy. The finally selected AUs are those that yield the maximum mean classification accuracy for each task as shown in Figure 3. The classification results based on the proposed methodology using a 10 times repeated 10x10 nested cross-validation process are presented in Table 6.

According to the mean relevance scores, as presented in Figures 2 and 3, the exposure experimental phase (16 AUs selected) presented AUs the AU23, AU06, AU12, AU07 and AU17, while the emotional recall phase (15 AU selected) the AU06, AU14, AU01, AU45 and AU12 as the 5 most dominant acute stress-relevant AUs. The mental workload phase (15 AUs selected) presented a prominent relevance of AU01, followed by the AU02, AU05, AU04 and AU15, and the stressful stimuli (all 17 AUs selected) presented a prominent relevance of AU26, followed by the AU25, AU17, AU05, AU23.

Table 4: Stress classification results (classification accuracy of 9 different classifiers) using experimental tasks 10-fold cross-validation

	no transformation														
	exposure			stressful event recall			mental workload			stressful stimuli			all tasks		
Classifiers	**acc** (%)	**sens** (%)	**spec** (%)	**acc** (%)	**sens** (%)	**spec** (%)	**acc** (%)	**sens** (%)	**spec** (%)	**acc** (%)	**sens** (%)	**spec** (%)	**acc** (%)	**sens** (%)	**spec** (%)
KNN	79.5	76.6	81.8	**67.7**	73.5	65.3	66.8	77.0	51.5	52.0	64.9	32.0	63.3	67.7	60.1
NVB	**84.7**	75.7	93.6	61.9	68.5	59.5	65.0	86.1	51.0	54.9	76.2	41.4	65.0	75.4	58.9
LDA	79.5	70.0	84.9	63.3	67.3	63.3	**73.5**	78.5	63.2	64.4	67.2	40.0	68.6	73.1	64.5
QDA	77.9	67.3	85.7	58.4	62.9	55.6	64.8	80.4	49.4	53.1	74.0	36.8	64.9	74.3	58.8
SVM linear	80.2	72.2	85.0	65.1	74.2	64.7	71.3	77.7	61.3	66.7	68.9	63.2	**69.8**	75.8	64.6
SVM quadratic	81.3	74.1	84.7	61.9	65.2	60.5	64.7	73.6	49.8	56.5	65.8	26.8	65.0	72.4	59.5
SVM cubic	76.2	66.4	81.9	60.2	60.3	64.8	62.2	72.2	48.7	53.6	63.7	23.3	62.7	69.2	57.9
SVM gaussian	78.1	67.4	86.5	65.3	63.4	70.2	69.0	74.0	54.8	**67.2**	68.1	45.0	66.4	69.3	63.1
Ens. decision trees	82.5	73.7	88.6	65.8	67.4	66.0	71.8	78.5	64.5	59.8	68.2	36.7	65.2	68.7	61.8

Note: acc: 10-fold mean accuracy, sen: 10-fold mean sensitivity (percentage of stress labels classified correctly), spec: 10-fold mean specificity (percentage of neutral labels classified correctly)

Table 5: Stress classification results (classification accuracy of 9 different classifiers) using experimental tasks 10-fold cross-validation for pairwise transformation.

	pairwise transformation														
	exposure			stressful event recall			mental workload			stressful stimuli			all tasks		
Classifiers	**acc** (%)	**sens** (%)	**spec** (%)	**acc** (%)	**sens** (%)	**spec** (%)	**acc** (%)	**sens** (%)	**spec** (%)	**acc** (%)	**sens** (%)	**spec** (%)	**acc** (%)	**sens** (%)	**spec** (%)
KNN	97.0	98.5	96.2	77.1	78.2	77.4	87.8	89.5	86.9	78.8	83.3	77.7	79.4	82.0	77.4
NVB	95.3	95.8	95.3	78.9	81.8	80.3	86.4	86.8	86.8	74.6	76.4	74.5	75.0	75.1	75.0
LDA	97.9	97.8	98.4	85.8	86.7	88.1	**91.5**	92.1	92.4	86.8	86.6	88.4	81.7	81.8	81.8
QDA	95.8	97.0	95.3	56.2	55.4	57.4	82.2	82.0	83.5	53.1	53.3	52.4	80.3	80.2	80.6
SVM linear	**98.7**	98.8	98.6	**89.1**	89.1	91.3	90.7	91.3	91.0	**87.7**	87.9	88.4	82.2	82.3	82.3
SVM quadratic	98.3	98.5	98.5	78.9	79.4	80.8	84.3	84.6	85.6	80.9	82.2	81.6	79.9	80.4	79.6
SVM cubic	97.9	97.6	98.4	80.3	82.3	79.4	89.9	89.7	91.1	82.7	83.0	83.1	**88.6**	89.7	87.6
SVM gaussian	97.9	97.7	98.3	76.1	82.0	77.9	89.3	88.4	90.8	82.6	82.2	84.3	86.1	86.3	85.9
Ens. decision trees	96.6	97.6	96.0	78.8	81.1	79.6	87.3	88.4	87.0	82.3	84.7	81.3	84.7	85.5	83.9

Note: acc: 10-fold mean accuracy, sen: 10-fold mean sensitivity (percentage of stress labels classified correctly), spec: 10-fold mean specificity (percentage of neutral labels classified correctly)

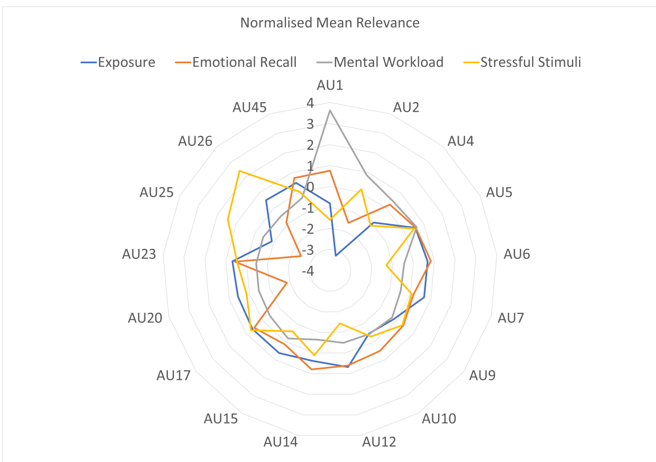


Figure 2: Normalised mean relevance of all AUs for the correct classified experimental phases.

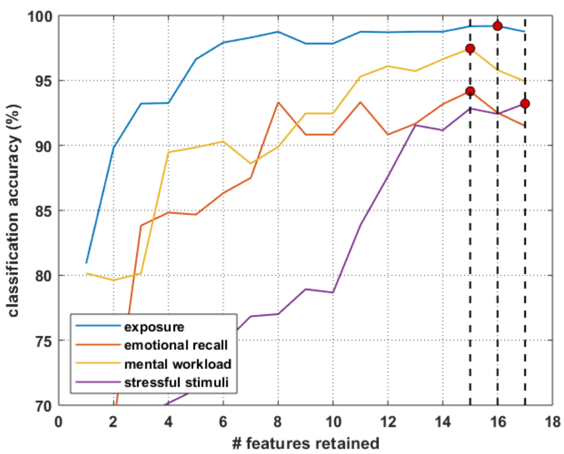


Figure 3: Mean classification accuracy in relation to the retained ranked features derived from the LRP algorithm for each experimental phase. The red dots denote the features selected corresponding to the maximum classification accuracy.

There are some prominent, high positive relevance AUs for recognizing acute stress, specifically, the AU01 Inner brow raiser for the mental workload phase and the AU26 Jaw drop for the stressful stimuli phase. These AUs are distinctively signifi-

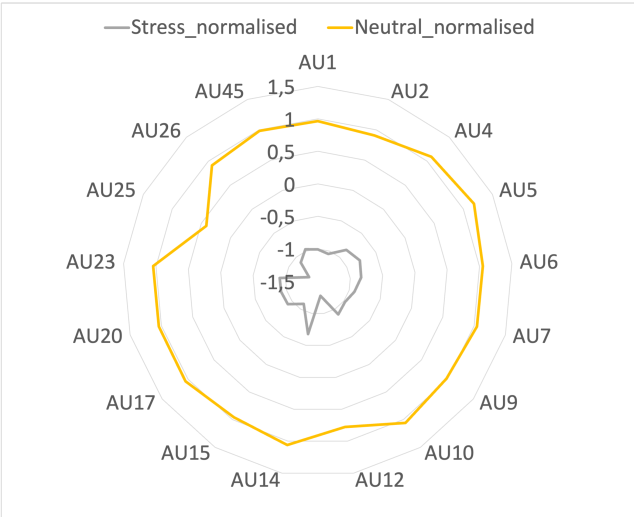


Figure 4: Normalised mean relevance of the AUs for the stressful vs the neutral tasks.

cant for the corresponding phases. However, the results for the AU2 across tasks are not consistent. There is negative relevance in the exposure and emotional recall tasks, zero for the stressful stimuli, and high positive relevance in the mental workload task. In Figure 4, it can be observed that all AUs have substantial importance to the classification decision for both tasks. This fact is consistent with the overall results of this analysis that there are no unique AUs for the identification of stress. Combination of AUs resutl to the best classification accuray. External validation is desirable to verify the robustness of the method.

However, it is obvious that the Deep Learning with explainable AI, LRP methodology improved the performance in relation to the conventional machine learning methods for all experimental phases with classification accuracies ranging from 94-99%. The best performance was observed for the exposure experimental phase with a mean classification accuracy of 99.1%.

4. Discussion

In this study, we investigated the behaviour of facial AUs combinations in recognizing stress conditions for different stressor types. We created a new multistressor dataset of 58 participants extracting facial AUs according to our previous work [29].

The paper investigated the best AUs combination for discriminating acute stress. Towards this direction, we utilized the AI feature ranking/selection method mRMR, which performed well in previous studies [29]. However, explainable AI deep learning methods are considered robust and have been adopted by many recent works. Thus, we employed the explainable AI LRP algorithm as proposed by Bach et al. [58]. This algorithm presented a superior performance in identifying the most robust features and classifying between acute stress/no stress conditions. This methodology improved the performance in relation to the conventional machine learning methods for all experimental phases with classification accuracies ranging from 93% to 99%. However, each proposed classification system strongly depends on the experimental stressor type, and it is not yet possible to derive a ”global stress recognition system” as the performance achieved using all experimental phases together led to a classification accuracy of 83.5%.

It can be observed that, under stress conditions, participants' faces present significantly more AUs and with greater intensity. This can be attributed to the stress-induced facial muscle activity/tension [60] to stress-accompanied signs such as restlessness, irritability, and nervousness [61].

However, no consistent pattern of relevant AUs implicated across all the experimental phases. Each stressor (experimental phase) triggers different AUs in the participant's face. This may be partially attributed to the different stress types elicited during the experimental procedure and to the fact that speech was included in some experimental phases but omitted in others [62]. However, in this study, we identify which AUs combinations were the most relevant to the stress conditions of each experimental phase and should be utilized for each stressor type.

To our knowledge, there are no studies with consistent results regarding which facial AUs are implicated in acute stress conditions. In [28], AU20 was correlated with the stress phase, which was not repeated in other subjects. Most of the studies use AUs to train stress models and to provide accurate, sensitive, and specific results in recognition. Thus, a direct comparison with the stress-relevant AUs estimated in this study is not possible. Regarding the stress recognition accuracies, our method achieves accuracies of 94-99%, which can be considered a good performance in relation to other studies (such as 90.2% [25], 93% [28], 74.6% [29]), however, as the facial videos datasets are different, a direct comparison of the performances is not possible.

In the experimental dataset we created, we compared the conventional ML method with explainable AI to identify combined AUs involved in recognizing acute stress. One of this article's contributions is the adoption of explainable AI tools in the context of face-based stress analysis, which led to interesting conclusions. We addressed the critical issue of each stressor type manifesting itself in different combinations of AUs. Our plan is to investigate more state-of-the-art algorithms using larger samples to investigate the relation of acute stress with AUs with respect to the curse of dimensionality.

A limitation of the study is the dataset's sample medium size, which though is considered relatively higher than other related stress datasets [63], it remains small for the application of the deep learning methods. The results are considered as proof-of-concept and further investigation in larger datasets would permit a more extensive model validation. Even if advanced methods (such as iterative nested cross-validation and cross-validation feature selection) were applied to avoid overfitting and improve the model's generalizability, the nature of deep learning techniques requires large datasets to enhance the model's reliability. Moreover, a combined study with other modalities (such as biosignals, speech etc.) will improve the system's ability to address open challenges.

Table 6: Classification results using Neural Networks with LRP algorithm and pairwise transformation

Experimental phase	Ranked features with higher mean absolute relevance	Mean Accuracy%	Mean Sensitivity%	Mean Specificity%
exposure	AU23 AU06 AU12 AU07 AU17 AU15 AU20 AU05 AU26 AU45 AU14 AU09 AU10 AU01 AU25 AU04	**99.19**	**99.23**	**99.17**
emotional recall	AU06 AU14 AU01 AU45 AU12 AU05 AU23 AU17 AU10 AU09 AU04 AU07 AU15 AU26 AU02	**94.17**	**93.69**	**95.39**
mental workload	AU01 AU02 AU05 AU04 AU15 AU09 AU45 AU17 AU10 AU25 AU06 AU23 AU20 AU07 AU12	**97.46**	**96.86**	**99.23**
stressful stimuli	AU26 AU25 AU17 AU05 AU23 AU09 AU02 AU20 AU14 AU07 AU45 AU10 AU15 AU04 AU06 AU12	**93.22**	**93.13**	**93.86**
All tasks	AU6, AU12,AU10,AU2, AU4, AU1,AU23, AU20,AU15, AU45, AU9,AU14, AU5	**83.50**	**83.80**	**84.24**

5. Acknowledgements

5.1. Statement of ethical approval

The study was approved by the Research Ethics Committee of FORTH (approval no. 155/12-09-2022). All participants provided informed consent.

5.2. Funding

This research received no external funding.

5.3. Conflicts of Interest

The authors declare no conflict of interest.

References

[1] G. P. Chrousos, Stress and disorders of the stress system, Nature reviews endocrinology 5 (7) (2009) 374.

[2] L. Marques-Feixa, Águeda Castro-Quintas, H. Palma-Gudiel, S. Romero, A. Morer, M. Rapado-Castro, M. Martín, I. Zorrilla, H. Blasco-Fontecilla, M. Ramírez, M. Mayoral, I. Mendez, N. San Martín-Gonzalez, M. Rodrigo-Yanguas, J. Luis Monteserín-García, L. Fañanás, M. José Muñoz, E. Anglada, A. Mas, M. José Lobato, P. Santamarina, S. Gadea, M. Laborde, C. Moreno, L. Gayubo, M. Marín-Vila, Secretory immunoglobulin a (s-iga) reactivity to acute psychosocial stress in children and adolescents: The influence of pubertal development and history of maltreatment, Brain, Behavior, and Immunity 103 (2022) 122–129. doi:https://doi.org/10.1016/j.bbi.2022.04.010.

[3] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, M. Tsiknakis, Review on psychological stress detection using biosignals, IEEE Transactions on Affective Computing (2019).

[4] G. Giannakakis, K. Marias, M. Tsiknakis, A stress recognition system using hrv parameters and machine learning techniques, in: 2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), IEEE, 2019, pp. 269–272.

[5] L. Gonzalez-Carabarin, E. Castellanos-Alvarado, P. Castro-Garcia, M. Garcia-Ramirez, Machine learning for personalised stress detection: Inter-individual variability of eeg-ecg markers for acute-stress response, Computer Methods and Programs in Biomedicine 209 (2021) 106314. doi:https://doi.org/10.1016/j.cmpb.2021.106314.

[6] R. Zhou, C. Wang, P. Zhang, X. Chen, L. Du, P. Wang, Z. Zhao, M. Du, Z. Fang, Ecg-based biometric under different psychological stress states, Computer Methods and Programs in Biomedicine 202 (2021) 106005. doi:https://doi.org/10.1016/j.cmpb.2021.106005.

[7] S. Pourmohammadi, A. Maleki, Stress detection using ecg and emg signals: A comprehensive study, Computer Methods and Programs in Biomedicine 193 (2020) 105482. doi:https://doi.org/10.1016/j.cmpb.2020.105482.

[8] J. Rodríguez-Arce, L. Lara-Flores, O. Portillo-Rodríguez, R. Martínez-Méndez, Towards an anxiety and stress recognition system for academic environments based on physiological features, Computer Methods and Programs in Biomedicine 190 (2020) 105408. doi:https://doi.org/10.1016/j.cmpb.2020.105408.

[9] B. Fasel, J. Luettin, Automatic facial expression analysis: a survey, Pattern recognition 36 (1) (2003) 259–275.

[10] M. Pantic, L. J. M. Rothkrantz, Automatic analysis of facial expressions: The state of the art, IEEE Transactions on pattern analysis and machine intelligence 22 (12) (2000) 1424–1445.

[11] Z. Zeng, M. Pantic, G. I. Roisman, T. S. Huang, A survey of affect recognition methods: Audio, visual, and spontaneous expressions, IEEE transactions on pattern analysis and machine intelligence 31 (1) (2008) 39–58.

[12] M. R. Koujan, L. Alharbawee, G. Giannakakis, N. Pugeault, A. Roussos, Real-time facial expression recognition "in the wild" by disentangling 3d expression from identity, in: 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), 2020, pp. 24–31.

[13] G. Giannakakis, M. Pediaditis, D. Manousos, E. Kazantzaki, F. Chiarugi, P. Simos, K. Marias, M. Tsiknakis, Stress and anxiety detection using facial cues from videos, Biomedical Signal Processing and Control 31 (2017) 89–101.

[14] A. I. Korda, G. Giannakakis, E. Ventouras, P. A. Asvestas, N. Smyrnis, K. Marias, G. K. Matsopoulos, Recognition of blinks activity patterns during stress conditions using cnn and markovian analysis, Signals 2 (1) (2021) 55–71.

[15] F. Bevilacqua, H. Engström, P. Backlund, Automated analysis of facial cues from videos as a potential method for differentiating stress and boredom of players in games, International Journal of Computer Games Technology 2018 (2018).

[16] C. Daudelin-Peltier, H. Forget, C. Blais, A. Deschênes, D. Fiset, The effect of acute social stress on the recognition of facial expression of emotions, Scientific Reports 7 (1) (2017) 1036.

[17] P. Ekman, Facial action coding system (FACS), A human face (2002).

[18] E. A. Clark, J. Kessinger, S. E. Duncan, M. A. Bell, J. Lahne, D. L. Gallagher, S. F. O'Keefe, The facial action coding system for characterization of human affective response to consumer product-based stimuli: A systematic review, Frontiers in Psychology 11 (2020). doi:10.3389/fpsyg.2020.00920.

[19] B. Martinez, M. F. Valstar, B. Jiang, M. Pantic, Automatic analysis of facial actions: A survey, IEEE Transactions on Affective Computing (2017).

[20] R. Zhi, M. Liu, D. Zhang, A comprehensive survey on automatic facial action unit analysis, The Visual Computer 36 (5) (2020) 1067–1093.

[21] N. Sharma, T. Gedeon, Objective measures, sensors and computational techniques for stress recognition and classification: A survey, Computer Methods and Programs in Biomedicine 108 (3) (2012) 1287–1301. doi:https://doi.org/10.1016/j.cmpb.2012.07.003.

[22] X. Wang, T. Zhang, C. P. Chen, Pau-net: Privileged action unit network for facial expression recognition, IEEE Transactions on Cognitive and Developmental Systems (2022).

[23] T. Jeon, H. B. Bae, Y. Lee, S. Jang, S. Lee, Deep-learning-based stress

recognition with spatial-temporal facial information, Sensors 21 (22) (2021) 7498.
[24] H. Zhang, L. Feng, N. Li, Z. Jin, L. Cao, Video-based stress detection through deep learning, Sensors 20 (19) (2020) 5552.
[25] M. Gavrilescu, N. Vizireanu, Predicting depression, anxiety, and stress levels from videos using the facial action coding system, Sensors 19 (17) (2019) 3693.
[26] S. Jaiswal, M. Valstar, Deep learning the dynamic appearance and shape of facial action units, in: 2016 IEEE winter conference on applications of computer vision (WACV), IEEE, 2016, pp. 1–8.
[27] G. Giannakakis, M. R. Koujan, A. Roussos, K. Marias, Automatic stress analysis from facial videos based on deep facial action units recognition, Pattern Analysis and Applications 25 (3) (2022) 521–535.
[28] C. Viegas, S.-H. Lau, R. Maxion, A. Hauptmann, Towards independent stress detection: A dependent model using facial action units, in: 2018 International Conference on Content-Based Multimedia Indexing (CBMI), IEEE, 2018, pp. 1–6.
[29] G. Giannakakis, M. R. Koujan, A. Roussos, K. Marias, Automatic stress detection evaluating models of facial action units, in: 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), 2020, pp. 817–822.
[30] H. Ge, Z. Zhu, Y. Dai, B. Wang, X. Wu, Facial expression recognition based on deep learning, Computer Methods and Programs in Biomedicine 215 (2022) 106621. doi:https://doi.org/10.1016/j.cmpb.2022.106621.
[31] M. Deramgozin, S. Jovanovic, H. Rabah, N. Ramzan, A hybrid explainable ai framework applied to global and local facial expression recognition, in: 2021 IEEE International Conference on Imaging Systems and Techniques (IST), IEEE, 2021, pp. 1–5.
[32] V. S. Ramachandran, et al., Encyclopedia of the human brain, Tech. rep., Academic Press (2002).
[33] H. Ursin, H. R. Eriksen, The cognitive activation theory of stress, Psychoneuroendocrinology 29 (5) (2004) 567–592.
[34] F. Scarpina, S. Tagini, The stroop color and word test, Frontiers in Psychology 8 (2017). doi:10.3389/fpsyg.2017.00557.
[35] M. Nikravesh, Z. Jafari, M. Mehrpour, R. Kazemi, Y. Amiri shavaki, S. Hossienifar, M. P. a. Azizi, The paced auditory serial addition test for working memory assessment: Psychometric properties, Medical Journal of the Islamic Republic Of Iran 31 (1) (2017). arXiv:http://mjiri.iums.ac.ir/article-1-4138-en.pdf, doi:10.14196/mjiri.31.61.
URL `http://mjiri.iums.ac.ir/article-1-4138-en.html`
[36] G. Giannakakis, D. Manousos, V. Chaniotakis, M. Tsiknakis, Evaluation of head pose features for stress detection and classification, in: 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), 2018, pp. 406–409.
[37] M. Puurtinen, J. Viik, J. Hyttinen, Best electrode locations for a small bipolar ecg device: Signal strength analysis of clinical data, Annals of biomedical engineering 37 (2009) 331–336.
[38] M. M. Bradley, P. J. Lang, Measuring emotion: the self-assessment manikin and the semantic differential, Journal of behavior therapy and experimental psychiatry 25 (1) (1994) 49–59.
[39] P. Ekman, W. Friesen, Facial action coding system (FACS): manual (1978).
[40] P. Viola, M. J. Jones, Robust real-time face detection, International Journal of computer vision 57 (2) (2004) 137–154.
[41] J. Deng, Y. Zhou, S. Cheng, S. Zaferiou, Cascade multi-view hourglass model for robust 3d face alignment, in: FG, 2018.
[42] T. Cootes, C. Taylor, Statistical models of appearance for computer vision, Tech. rep., University of Manchester (2004).
[43] D. Watson, Contouring: a guide to the analysis and display of spatial data, Vol. 10, Elsevier, 2013.
[44] T. Baltrušaitis, M. Mahmoud, P. Robinson, Cross-dataset learning and person-specific normalisation for automatic action unit detection, in: Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on, Vol. 6, IEEE, 2015, pp. 1–6.
[45] P. Carcagnì, M. Del Coco, M. Leo, C. Distante, Facial expression recognition and histograms of oriented gradients: a comprehensive study, SpringerPlus 4 (1) (2015) 645.
[46] A. J. Smola, B. Schölkopf, A tutorial on support vector regression, Statistics and Computing 14 (3) (2004) 199–222.
[47] P. Lucey, J. F. Cohn, K. M. Prkachin, P. E. Solomon, I. Matthews, Painful data: The unbc-mcmaster shoulder pain expression archive database, in: Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, IEEE, 2011, pp. 57–64.
[48] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, L. Akarun, Bosphorus database for 3d face analysis, in: European Workshop on Biometrics and Identity Management, Springer, 2008, pp. 47–56.
[49] J. He, D. Li, B. Yang, S. Cao, B. Sun, L. Yu, Multi view facial action unit detection based on cnn and blstm-rnn, in: 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), IEEE, 2017, pp. 848–853.
[50] Z. Shao, Z. Liu, J. Cai, L. Ma, Deep adaptive attention for joint facial action unit detection and face alignment, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 705–720.
[51] Z. Shao, Z. Liu, J. Cai, L. Ma, Jaa-net: joint facial action unit detection and face alignment via adaptive attention, International Journal of Computer Vision 129 (2021) 321–340.
[52] C. Ding, H. Peng, Minimum redundancy feature selection from microarray gene expression data, Journal of bioinformatics and computational biology 3 (02) (2005) 185–205.
[53] G. Gulgezen, Z. Cataltepe, L. Yu, Stable and accurate feature selection, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2009, pp. 455–468.
[54] R. Herbrich, T. Graepel, K. Obermayer, Support vector learning for ordinal regression, in: 1999 Ninth International Conference on Artificial Neural Networks ICANN 99., Vol. 1, 1999, pp. 97–102 vol.1. doi:10.1049/cp:19991091.
[55] J. Fürnkranz, E. Hüllermeier, Pairwise preference learning and ranking, in: European conference on machine learning, Springer, 2003, pp. 145–156.
[56] A. Korda, A. Ruef, S. Neufang, C. Davatzikos, S. Borgwardt, E. Meisenzahl, N. Koutsouleris, Identification of voxel-based texture abnormalities as new biomarkers for schizophrenia and major depressive patients using layer-wise relevance propagation on deep learning decisions, Psychiatry Research: Neuroimaging 313 (2021) 111303.
[57] J.-L. Zhang, On the convergence properties of the levenberg–marquardt method, Optimization 52 (6) (2003) 739–756.
[58] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, W. Samek, On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation, PloS one 10 (7) (2015) e0130140.
[59] A. Korda, E. Ventouras, P. Asvestas, M. Toumaian, G. Matsopoulos, N. Smyrnis, Convolutional neural network propagation on electroencephalographic scalograms for detection of schizophrenia, Clinical Neurophysiology 139 (2022) 90–105.
[60] L. M. Mayo, M. Heilig, In the face of stress: Interpreting individual differences in stress-induced facial expressions, Neurobiology of stress 10 (2019) 100166.
[61] J. R. Geddes, N. C. Andreasen, New Oxford textbook of psychiatry, Oxford University Press, USA, 2020.
[62] G. Giannakakis, D. Manousos, P. Simos, M. Tsiknakis, Head movements in context of speech during stress induction, in: 13th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2018), Xi'An, China, 2018.
[63] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, I. Patras, Amigos: A dataset for affect, personality and mood research on individuals and groups, IEEE Transactions on Affective Computing 12 (2) (2018) 479–493.