Automatic stress detection evaluating models of facial action units

Giorgos Giannakakis[1], Mohammad Rami Koujan[2], Anastasios Roussos[1] and Kostas Marias[1,3]
[1] Institute of Computer Science, Foundation for Research and Technology Hellas (FORTH-ICS), Greece
[2] College of Engineering, Mathematics and Physical Sciences, University of Exeter, UK
[3] Department of Electrical and Computer Engineering, Hellenic Mediterranean University, Greece

Abstract— **Emotional stress detection can be performed analyzing different facial parameters. This paper focuses on the automated identification of facial Action Units (AU) as quantitative indices in order to discriminate between neutral and stress/anxiety state. Thus, a model for automatic recognition of facial action units is proposed being trained in two available annotated facial datasets, the UNBC and the BOSPHORUS datasets. Facial features, both geometric (non-rigid deformations of 3D shape of AAM landmarks) and appearance (Histograms of Oriented Gradients) were extracted. The intensity of each AU was regressed using Support Vector Regression (SVR). The corresponding models of each dataset were fused to a combined model. This combined model was applied to the experimental dataset (SRD'15) containing neutral states and inducing stressful states related to four types of stress. The results indicate that there are specific AU relevant to stress and the AU intensity are significant increased during stress leading to a more expressive human face.**

I. INTRODUCTION

There are different indices associated with the stress response of the human body to stressful stimuli and situations. Among different physiological measures (such as biomarkers, laboratory exams, psychometric scales, etc), the emotional stress can be assessed analyzing different facial parameters. Even though facial expressions can be suppressed or manipulated, however, it is not always possible to be suppressed and they convey significant information related to affective states. Especially, during stress conditions, there are semi-voluntary facial cues or micro-expressions.

Facial expressions are associated with emotions and type of affect, thus the research community pursuits to find reliable techniques to decode their presence or the combination of their presence in the human face. Besides, it is arguable whether the AU intensity may be estimated and how their accurate and reproducible measurement can be performed. There are some coding systems addressing this issue, the most widely adopted of which is the notion that face expresses facial Action Units (AU) which can be coded according to the Facial Action Coding System (FACS) [1] [1]. A recent review summarizes techniques of facial AU analysis, not concluding to specific guidelines but delineating good practices in facial action units analysis [2].

There are some recent studies, investigating facial cues [3] and facial expressions' behaviour during stress conditions [4], [5]. However, there are neither specific guidelines which AU are implicated in stress state nor a consistent model describing the stress manifestation on specific facial expressions.

In this paper, an AU recognition model is established using publicly available training datasets and the combined model is applied to a thorough stress experimental dataset. Besides, the AU implicated in stress are investigated as a response to different types of stressors (related to social exposure, emotional recall, mental strain and stressful videos stimuli) in experimental conditions.

II. METHODS

This study focuses on automatic stress identification from the intensity of facial AU that are estimated from trained SVM models trained in two annotated databases. The procedure has 3 phases: preprocessing (including face detection, AAM facial landmark estimation, face alignment/normalization, face warping), feature extraction (shape and appearance features), AU classification (including PCA on appearance features, Support Vector Regression (SVR) training and AU intensity estimation). The AU recognition procedure flowchart depicting the 3 phases (preprocessing, feature extraction, classification) is shown in Figure 1.

TABLE I
SUMMARY OF THE AU, THEIR FACS NAME AND MUSCULAR BASIS INVESTIGATED IN THIS STUDY

AU	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
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

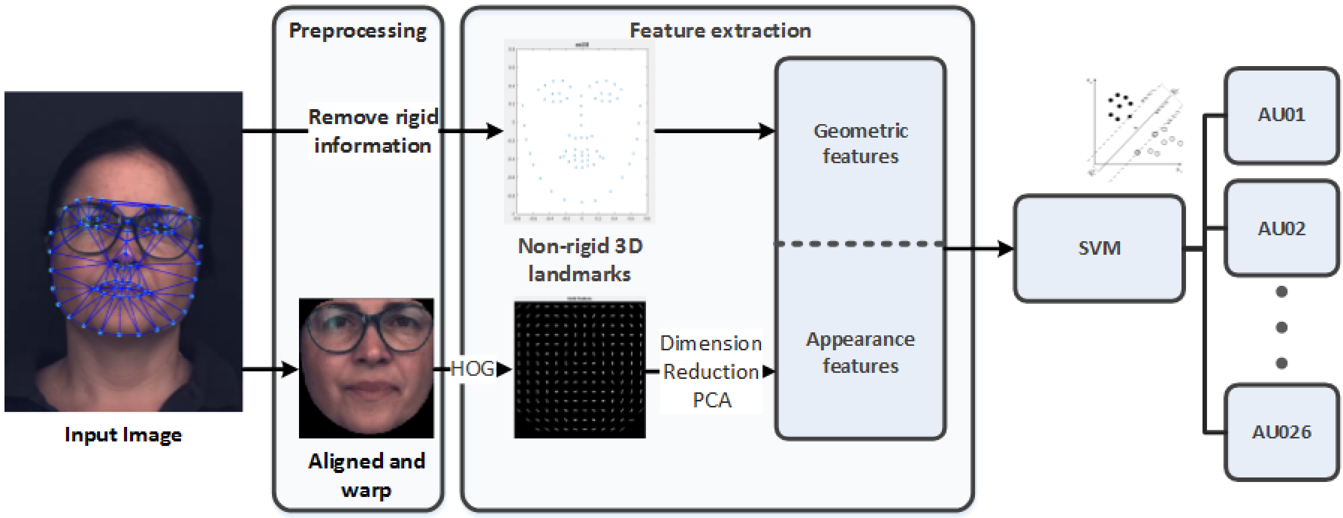

Fig. 1. Flowchart of the system for facial AU presence following the phases of preprocessing, feature extraction (geometric and appearance), SVM classification and AU estimation.

A. Facial Action Coding System

The Facial Action Coding System (FACS) [1][6] is a system that initially developed by a Swedish anatomist and updated by Ekman and Friesen in 1978 [6] and 2002 [1] respectively. It systematically categorizes human facial muscle movements and expressions based on anatomic functions.

Additionally, it encodes actions related to eye gaze, head pose and other actions. The AU that are investigated in this study in order to reveal associations with types of stress are presented in Table I.

B. Preprocessing

The face area is detected from the input image using the Viola-Jones detector described in [7]. Then, the facial landmarks are estimated (68 points mark-up) using Active Appearance Models (AAM) [8]. Then, the preprocessing phase is applied which includes the removal of rigid information and the face alignment/normalization. This step is significant for the subsequent analysis as aligning the faces into a common reference frame will lead to features that correspond to the same facial areas, thus conveying the same semantic information [2]. The alignment and warping to the mean base (neutral) facial shape was performed with Delaunay triangle-based affine warp [9].

C. Feature extraction

In this study, both geometric and appearance features were extracted in order to enhance the estimation of the AU under investigation.

For the geometric features, the non-rigid 3D landmarks were estimated because they provide the most reliable information as in this way the shape displacements are caused only from facial expressions and not head inclinations, nods, etc. A linear model called a Point Distribution Model (PDM) [10], [11] provides a parametric representation of the deformable shapes given by the expression

$$x = a \cdot R(\bar{X} + \Phi p) + T$$

where $x : (x_i, y_i)$ is the 2D landmarks, a is the scale factor, R the rotation matrix, $T : (t_x, t_y)$, the 3D mean value of the PDM in the 3D reference frame, and p the non-rigid shape parameters. The non-rigid 3D landmarks are used as geometric features. Regarding the appearance features, we use Histograms of Oriented Gradients (HOG) [12] 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. In order to reduce high dimensionality, we apply Principal Component Analysis (PCA) and the retained components (explaining the 95% of the total data variability) form the appearance features vector.

Both geometric and appearance features form a vector that is the input in the Support Vector Regression model (SVR) [13]. The SVR performs regression on the data and its corresponding labels which are annotated AU intensities.

D. Training datasets

In this study, we used two available facial datasets in order to train the SVR model, the UNBC-McMaster Shoulder Pain Expression Archive Database (UNBC) [14] and the Bosphorus database (BOSPHORUS) [15].

The UNBC dataset contains 200 sequences across 25 subjects (total 48,398 images) annotated according to FACS code and their corresponding AU intensity in a scale [0 5] where 0 corresponds to AU non-existence whereas 5 corresponds to the maximum AU intensity. It contains annotated information for the AU [6, 7, 9, 10, 12, 25, 26]. The BOSPHORUS dataset contains 105 subjects with total 4666 facial images annotated according to FACS code and their corresponding AU intensity in a scale [0 5]. It contains annotated information for the AU [1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, 43]. The common features from

TABLE II

EXPERIMENTAL TASKS EMPLOYED IN THIS STUDY

	Experimental task	Duration (min)	Affective State
Social Exposure			
1.1	Neutral (reference)	1	N
1.2	Interview	2	S
Emotional recall			
2.1	Neutral (reference)	1	N
2.2	Recall anxious event	1	S
2.3	Recall stressful event	1	S
Stressful images/Stroop task			
3.1	IAPS stressful images	2	S
3.2	Stroop Colour-Word Test (SCWT)	2	S
Stressful videos			
4.1	Neutral (reference)	1	N
4.2	Calming video	2	R
4.3	Adventure video	2	S
4.4	Psychological pressure video	2	S

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

the two training datasets are combined and provided with other non-common features to the SVM model.

E. Features relevance

The most relevant/important AU features are investigated and selected in order to state their relevance with stress and to improve the performance of the stress model. The ranking of feature importance was performed using the minimum Redundancy Maximum Relevance (mRMR) selection algorithm [16]. This algorithm evaluates the features' importance ranking based on maximal relevance and minimum redundancy optimizing in terms of the Mutual Information Quotient (MIQ) criterion [17]. The number of retained features was determined by minimizing the misclassification error using 10-fold SVM discrimination accuracy between neutral and stress states.

F. Experimental procedure

The experimental procedure used in this study aims to induce stress states to participants employing different types of stressors. These stressors are categorized into 4 different phases (social exposure, emotional recall, mental workload tasks, stressful videos) corresponding to different stress types were determined. The experimental procedure is presented in .Table 2 Each of the participants was seated in front of a computer monitor. The camera was placed on a tripod at the back top of the monitor and at a distance of 90 cm with its field of view covering the participant's face and possible movements during the experiment. At the beginning of the procedure, participants were informed about the whole procedure as well as about the terms of anxiety and stress.

G. Experimental Dataset (SRD'15)

The population of this study were 24 participants (7 women, 17 men) with age 47.3$\pm$9.3 years. The study was approved by the North-West Tuscany ESTAV (Regional Health Service, Agency for the Technical-Administrative Services of Wide Area) Ethical Committee. Data were recorded during the second data acquisition campaign (SRD'15) of a research project aiming at the development of computational platform monitoring cardio-metabolic risk [18].

III. RESULTS

A. Training datasets validation

The SVM model was trained following the procedure described in sections -II.B.II.D The training effectiveness was assessed validating the model after the training phase. A 10-fold cross-validation technique was used for model validation using an SVM classifier. The model's performance was assessed using the measures

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

$$F_1 = \frac{2 \cdot precision \cdot recall}{(precision + recall)}$$

However, the F1 measure was considered more appropriate, as there are great time segments where there is no AU present (neutral face). The results of the validation phase are presented in Table 3 and Table 4. It can be observed that the combined model in most cases of AU outperforms the individual models because it achieves better generalizability.

B. AU intensity extraction

The intensity (scale 0-5) of each of the 15 AU features (AU1, AU2, AU4, AU5, AU6, AU7, AU9, AU10, AU12, AU14, AU15, AU17, AU20, AU23, AU25, AU26) from the FACS which described in section II Awere extracted per frame. Typical timeseries of an AU (AU17) for a neutral and a stressful task (adventure video) is presented in Fig. 2.

The AU present different patterns between neutral and stress states, which in most cases the AU's intensity is significantly higher during stress conditions indicating a more expressive face as it is discussed in section III.C

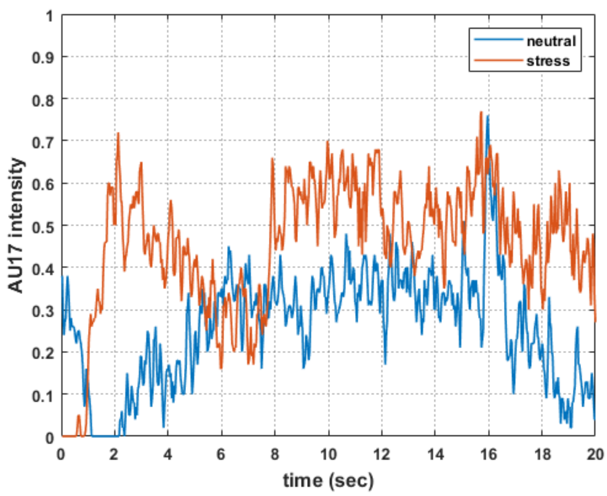

Fig. 2. Plot showing the AU17 intensity temporal evolution over 20 sec of one subject. A neutral state (task 4.1) denoted in blue line and the adventure video (task.4.3) denoted in red line

TABLE III

ACCURACY MEASURE USING 10-FOLD CROSS VALIDATION TECHNIQUE FOR THE UNBC, BOSPHORUS, AND THE COMBINED MODEL (UNBC+BOSPHORUS)

Model	**AU01**	**AU02**	**AU04**	**AU05**	**AU06**	**AU07**	**AU09**	**AU10**	**AU12**	**AU14**	**AU15**	**AU17**	**AU23**	**AU25**	**AU26**
UNBC					0.78	0.87	0.91	0.99	0.78					0.79	0.69
BOSPHORUS	0.74	0.74	0.87	0.75	0.82	0.77	0.95	0.80	0.92	0.88	0.75	0.82	0.79	0.78	0.92
Combined model	0.77	0.75	0.88	0.82	0.62	0.82	0.88	0.86	0.74	0.82	0.84	0.81	0.91	0.82	0.65

TABLE IV

F1 MEASURE USING 10-FOLD CROSS VALIDATION TECHNIQUE FOR THE UNBC, BOSPHORUS, AND THE COMBINED MODEL (UNBC+BOSPHORUS)

Model	**AU01**	**AU02**	**AU04**	**AU05**	**AU06**	**AU07**	**AU09**	**AU10**	**AU12**	**AU14**	**AU15**	**AU17**	**AU23**	**AU25**	**AU26**
UNBC					0.44	0.20	0.15	0.71	0.51					0.21	0.48
BOSPHORUS															
Combined model	0.46	0.42	0.65	0.55	0.32	0.42	0.15	0.51	0.47	0.41	0.33	0.52	0.4	0.28	0.28

C. *Statistical analysis*

The dataset was checked for normality for each AU and each task according to the Kolmogorov-Smirnov (KS) test. In most cases, data samples under consideration follow the normal distribution. An initial statistical evaluation (dependent samples t-test or the corresponding non-parametric Wilcoxon signed rank test respectively) was performed and the results of selected features are presented in Table V.

Increased AU intensities can be observed along all stressful tasks meaning that the face tends to be more "expressive", i.e. manifesting more intense AU during stress conditions. Only during emotional recall phase, there were not significant widespread differences.

D. *AU involvement in stress conditions*

The AU most implicated in stress conditions were also investigated. Towards this end, the mRMR and random forest (RF) were employed. The top-ranked features were inserted iteratively in the feature subset evaluating each candidate subset's performance in terms of 10-fold SVM classification accuracy used as the objective function. The results are presented in Table VI.

This procedure revealed that for the AU stress detection problem under investigation, a subset of 5 or 6 most relevant features may differentiate effectively the two states. It should be noted that there is a consistent selection of relevant features along the 3 algorithms used.

E. *Application of the trained model on stress detection*

Then, the model was tested in the experimental SRD'15 dataset. The videos were the input of the system and were grouped according to the label of the task in 2 groups (no stress, stress states). All the AU under investigation were extracted from the trained combined model that was used as the initial features matrix. The most relevant features were assessed using traditional Machine Learning classification schemes in terms of their ability to discriminate between the two classes (no stress, stress) for all experimental phases. A 10-fold cross-validation scheme on was used utilizing the classifiers k nearest neighbours (KNN), Generalized Linear Model (GLM), Naïve Bayes (NVB), Linear Discriminant Analysis (LDA), Support Vector Machines (SVM). The results are shown in Table VII. The results of Table VII indicate that the SVM outperforms all other classification schemes with a classification accuracy of 74.6%. It should be noted that the results would be accounted fair, considering they are subject-independent because of the normalization that performed in the face alignment and the removal of the rigid information. Besides, except for the interview task, the other tasks do not include intense facial expressions.

IV. DISCUSSION

This paper presents a system for facial Action Units (AU) detection that was applied for stress recognition. This system was trained on two publicly available datasets, the UNBC and BOSPHORUS, fusing them in a combined model. It was shown that in most cases the combined model present better performance in comparison to the individual performance of each model. There is the belief that the inclusion of more datasets in the model would probably enhance its generalizability that should be checked. The most efficient AU tracking is observed for the AU09 (Nose wrinkle), AU10 (Upper lip raiser), AU04 (Brow lowerer), AU23 (Lip tightener), AU14 (Dimpler).

Regarding the application of the model on stress dataset (SRD'15), the AU that implicated in stress conditions were in vestigated, leading to the notion that the ones that are modulated and are able to discriminate the two emotional states are the AU17 (Chin raiser), AU25 (Lips part), AU1 (Inner brow raiser), AU7 (Lid tightener), AU26 (Jaw drop). An interesting conclusion is that stressful tasks lead to significant increased AU intensities, i.e. a more "expressive" face. This is in accordance with the increased head motility during stress conditions [19] and reduced motility according to the levels of depression severity [20]. It is notable that the most efficient classification accuracies are observed, as expected, on experiment phases where the participant was asked to be more communicative such as interview and Stroop Colour Word task.

TABLE V

SUMMARY OF AU STATISTICS ALONG EXPERIMENTAL TASKS PRESENTING SIGNIFICANT DIFFERENCES DURING STRESS CONDITIONS

AU	Interview		Recall anxious event		Recall stressful event		IAPS		Stroop CWT		Adventure video		Psychological pressure video	
	p	**diff**	***p***	**diff**	***p***	**diff**	***p***	**diff**	***p***	**diff**	***p***	**diff**	***p***	**diff**
AU01	0.000	↑					0.004	↑	0.000	↑	0.017	↑	0.006	↑
AU02	0.000	↑					0.016	↑	0.000	↑			0.035	↑
AU04	0.029	↓	0.009	↑			0.000	↑			0.001	↑	0.001	↑
AU05	0.000	↑			0.015	↑	0.005	↑	0.006	↑				
AU06	0.000	↑							0.001	↑				
AU07	0.000	↑	0.02	↑					0.003	↑				
AU09	0.000	↑							0.000	↑	0.005	↑		
AU10	0.000	↑							0.001	↑				
AU12	0.000	↑							0.006	↑				
AU14					0.012	↑					0.041	↑		
AU15	0.000	↑							0.019	↑				
AU17	0.000	↑					0.007	↑	0.000	↑	0.01	↑	0.016	↑
AU20	0.000	↑					0.026	↑	0.000	↑				
AU23	0.000	↑					0.003	↑	0.000	↑				
AU25	0.000	↑					0.011	↑	0.000	↑				
AU26	0.000	↑					0.017	↑	0.000	↑			0.049	↑
AU45	0.001	↑											0.023	↓

↑/↓ significant increase/decrease during stress conditions
ns: non-significant difference

Results indicate that during stress conditions there are specific AU that are differentiated from the normal state. In fact there some facial areas that appear motility leading to what is considered micro-expressions which is sometimes the result of nervousness and irritability. There were some clear patterns in AU during stress A good research question is how the AU implicated most in stress conditions are combined in order to form complex facial expressions and how they are manifested in the human face that would be the subject of future research. The dataset that was investigated in this study had framerate of 30 and surely a better temporal resolution would reveal more cues coming from micro-expressions that in this resolution remain hidden. In general, facial expression as represented through facial AU are a promising approach to face analysis related to the affective state estimation.

TABLE VI

RELEVANT FEATURES IMPLICATED IN STRESS CONDITIONS USING MRMR, RF AND FISCHER RATIO ALGORITHMS AND THEIR CORRESPONDING OBJECTIVE FUNCTION ACCURACY

Algorithm	Relevant features	Classification Accuracy (10-fold)
mRMR	AU17, AU25, AU01, AU07, AU26	0.73
RF	AU01, AU17, AU20, AU25, AU26	0.71
Fischer ratio	AU25, AU17, AU01, AU23, AU26, AU02, AU05	0.67

TABLE VII

CLASSIFICATION RESULTS OF THE 10-FOLD CROSS VALIDATION SVM METHOD ON THE COMBINED TRAINED MODEL (UNBC+BOSPHORUS) APPLIED ON THE EXPERIMENTAL DATASET (SRD'15)

Classifier	Classification Accuracy (%)	Sensitivity (%)	Specificity (%)
KNN	68.2	92.3	61.5
GLM	69.7	88.6	37.5
Naïve Bayes	57.7	72.7	40.0
LDA	58.0	83.3	40.0
SVM	**74.6**	83.3	50.0

REFERENCES

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

[2] B. Martinez, M. F. Valstar, B. Jiang, and M. Pantic, "Automatic analysis of facial actions: A survey," *IEEE Transactions on Affective Computing*, 2017.

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

[4] F. Bevilacqua, H. Engström, and 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*, vol. 2018, 2018.

[5] C. Daudelin-Peltier, H. Forget, C. Blais, A. Deschênes, and D. Fiset, "The effect of acute social stress on the recognition of facial expression of emotions," *Scientific Reports*, vol. 7, no. 1, p. 1036, 2017.

[6] P. Ekman and W. Friesen, "Facial action coding system (FACS): manual," 1978.

[7] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[8] I. Matthews and S. Baker, "Active appearance models revisited," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 135–164, 2004.

[9] D. Watson, *Contouring: a guide to the analysis and display of spatial data*. Elsevier, 2013, vol. 10.

[10] T. Baltrušaitis, M. Mahmoud, and 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.

[11] T. F. Cootes and C. J. Taylor, "Active shape models—'smart snakes'," in *BMVC92*. Springer, 1992, pp. 266–275.

[12] P. Carcagnì, M. Del Coco, M. Leo, and C. Distante, "Facial expression recognition and histograms of oriented gradients: a comprehensive study," *SpringerPlus*, vol. 4, no. 1, p. 645, 2015.

[13] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.

[14] P. Lucey, J. F. Cohn, K. M. Prkachin, P. E. Solomon, and 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.

[15] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3d face analysis," in *European Workshop on Biometrics and Identity Management*. Springer, 2008, pp. 47–56.

[16] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Journal of bioinformatics and computational biology*, vol. 3, no. 02, pp. 185–205, 2005.

[17] G. Gulgezen, Z. Cataltepe, and L. Yu, "Stable and accurate feature selection," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2009, pp. 455–468.

[18] Y. Andreu, F. Chiarugi, S. Colantonio, G. Giannakakis, D. Giorgi, P. Henriquez, E. Kazantzaki, D. Manousos, K. Marias, B. J. Matuszewski, M. A. Pascali, M. Pediaditis, G. Raccichini, and M. Tsiknakis, "Wize mirror - a smart, multisensory cardio-metabolic risk monitoring system," *Computer Vision and Image Understanding*, vol. 148, pp. 3–22, 2016.

[19] G. Giannakakis, D. Manousos, V. Chaniotakis, and 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.

[20] K. Anis, H. Zakia, D. Mohamed, and C. Jeffrey, "Detecting depression severity by interpretable representations of motion dynamics," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG2018)*, 2018, pp. 739–745.