Machine Vision and Applications (2020) 31:30
https://doi.org/10.1007/s00138-020-01080-7

ORIGINAL PAPER

Automated facial video-based recognition of depression and anxiety symptom severity: cross-corpus validation

A. Pampouchidou[1] · M. Pediaditis[2] · E. Kazantzaki[2,3] · S. Sfakianakis[2] · I. A. Apostolaki[3] · K. Argyraki[3] · D. Manousos[2] · F. Meriaudeau[1] · K. Marias[2,4] · F. Yang[1] · M. Tsiknakis[2,4] · M. Basta[3] · A. N. Vgontzas[3] · P. Simos[2,3]

Received: 1 September 2019 / Revised: 23 February 2020 / Accepted: 7 April 2020

Abstract

There is a growing interest in computational approaches permitting accurate detection of nonverbal signs of depression and related symptoms (i.e., anxiety and distress) that may serve as minimally intrusive means of monitoring illness progression. The aim of the present work was to develop a methodology for detecting such signs and to evaluate its generalizability and clinical specificity for detecting signs of depression and anxiety. Our approach focused on dynamic descriptors of facial expressions, employing motion history image, combined with appearance-based feature extraction algorithms (local binary patterns, histogram of oriented gradients), and visual geometry group features derived using deep learning networks through transfer learning. The relative performance of various alternative feature description and extraction techniques was first evaluated on a novel dataset comprising patients with a clinical diagnosis of depression ($n = 20$) and healthy volunteers ($n = 45$). Among various schemes involving depression measures as outcomes, best performance was obtained for continuous assessment of depression severity (as opposed to binary classification of patients and healthy volunteers). Comparable performance was achieved on a benchmark dataset, the audio/visual emotion challenge (AVEC'14). Regarding clinical specificity, results indicated that the proposed methodology was more accurate in detecting visual signs associated with self-reported anxiety symptoms. Findings are discussed in relation to clinical and technical limitations and future improvements.

Keywords Affective computing · Depression assessment · Facial image analysis · Image processing · Machine learning

1 Introduction

Depression is the most prevalent mood disorder [1], with a structured clinical interview, according to the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria, being the standard procedure for depression diagnosis [2,3]. Clinical diagnosis of depression may be supported by self-report questionnaires, such as the Beck Depression Inventory (BDI). These instruments are convenient and economical, but carry certain disadvantages, as they do not account for individual trait characteristics, other psychiatric and medical comorbidities (such as anxiety disorders and symptoms) and recent major stressors, and are vulnerable to intentional or unintentional reporting biases [4]. Limited continuous monitoring of persons at risk (e.g., individuals with a history of mental illness or suffering from chronic, debilitating physical diseases) is one of the factors contributing to the high rate of underdiagnosed depressive episodes [5]. In this context, there is a rising interest for technological innovations to enhance accessibility to mental healthcare [6].

A. Pampouchidou and M. Pediaditis have contributed equally to this work.

A. Pampouchidou was funded by the Greek State Scholarship Foundation, under the scholarship instituted in memory of Maria Zaousi.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00138-020-01080-7) contains supplementary material, which is available to authorized users.

✉ A. Pampouchidou
anastasia.pampouchidou@gmail.com

1 ImViA, University of Burgundy, Le Creusot, France

2 Foundation for Research and Technology - Hellas (FORTH), Heraklion, Crete, Greece

3 School of Medicine, Division of Psychiatry, University of Crete (UOC), Heraklion, Crete, Greece

4 Department of Electrical and Computer Engineering, Hellenic Mediterranean University, Heraklion, Greece

Published online: 30 April 2020

Depression is typically manifested through a variety of nonverbal signs, including systematic patterns of facial expression and body posture [7,8]. Objective measures of psychoemotional status derived from facial image analysis could complement self-report instruments and help overcome some of their shortcomings [9], serving as a minimally intrusive tool for monitoring depression symptomatology. Furthermore, the widespread and relatively low-cost accessibility to computer and internet technologies, webcams, and smart phones, render an efficient system for depression assessment viable. However, given the current state-of-the-art, video-based systems for depression assessment are not intended as standalone tools, but mainly as part of decision support systems assisting mental health professionals in remote monitoring of persons at risk. Considering that most of the nonverbal, facial signs of depression are dynamic [4,7,8] nearly all current approaches employ video-based, as opposed to frame-based (static), features.

Very few open datasets are available to support the development of novel technological solutions such as the AVEC'13-'14[1] and the DAIC-WOZ[2] datasets, introduced as part of the Depression Recognition Sub-Challenge (DSC) of the audio visual emotion challenge (AVEC). The AVEC dataset provides full access to audiovisual recordings, annotated by self-reported BDI scores; pre-extracted features and self-reported depression scale scores are available in the DAIC-WOZ dataset. The Pittsburgh[3] dataset has also been made available to the research community in the same manner as the DAIC-WOZ, by sharing pre-extracted features, while the BlackDog dataset which has also been reported in several published reports [10] is kept private.

Cohn et al. [11] was one of the first attempts toward automatic assessment of depression based on facial image analysis. They achieved 79% overall accuracy for binary classification (depressed vs. non-depressed) on the Pittsburgh dataset using Active Appearance Models (AAM) with Support Vector Machines (SVM). Alghowinem et al. [10] achieved 88.3% recall (i.e., sensitivity) for the same binary problem on the BlackDog dataset; they computed a total of 128 statistical features ("functionals") from eyelid and eye-corners distances, also using SVM for the classification. In an earlier attempt toward binary classification of depression symptom severity on the AVEC dataset, our group achieved an F1 score of 87.4% using motion history image variants and Visual Geometry Group (VGG) features [12]. Although previous work was extensive and tested several different algorithms, it focused on solving classification problems at the expense of fully appreciating the continuous nature of depression through regression methods. This gap in knowledge was addressed by the proposed methodology which was further validated on video data, annotated for both depression- and anxiety-related visual features, obtained from patients diagnosed with depression.

Continuous assessment is becoming increasingly popular and a key topic of the AVEC-DSC, where contestants attempted to minimize the error of predicting (healthy) participants' BDI scores. Error is usually quantified as Mean Absolute Error (MAE) and Root-Mean-Square Error (RMSE). Among top performing approaches, Jan et al. [13] achieved MAE/RMSE of 6.68 and 8.01, respectively, with a dynamic history histogram from VGG features using a weighted fusion technique on the Partial Least Square and Linear Regression. Corresponding error rates obtained by de Melo et al. [14] were 6.59 and 8.31; they extracted Convolutional 3D (C3D) network features with 3D Global Average Pooling combined from both the entire face and the eye region, using Linear Regression (for further video-based approaches see also [15–24]).

The AVEC dataset is suitable for assessing the generalizability of novel algorithms and analysis pipelines across cultures as it includes a range of video recording contexts and settings. This is a particularly important enterprise in view of extant research highlighting cultural differences in overt depression manifestations ([25,26]. However, the specificity of extracted visual features in assessing facial manifestations directly related to mood disturbances, as opposed to expressions of distress and anxiety, has not been assessed systematically in previous work. In addition to the very high comorbidity of diagnoses of mood and anxiety disorders [27–30], the two groups of disorders share common nonverbal manifestations.

The present study is addressing four specific aims. Firstly, to describe the development of a novel dataset comprising multimodal recordings from patients diagnosed with Major Depressive Disorder (MDD) and healthy volunteers. Data were obtained on a variety of experimental settings (including emotionally and cognitively neutral conditions and emotionally stimulating social and non-social contexts). Secondly, we aimed at developing a novel algorithmic pipeline for extracting facial expression elements from video recordings. In this context, various feature extraction and validation methods were employed and their relative performance was evaluated for (a) classification of cases according to clinical diagnosis or expert ratings of overt facial manifestations of depression and (b) continuous assessment of depression symptomatology. Thirdly, the generalizability of the best-performing computational approach (i.e., associated with the minimum prediction error for continuous assessment of depression symptomatology) was tested on the AVEC'14 dataset [21]. Fourthly, given the known diversity of psychological and behavioral manifestations of depression, we evaluated the relative performance of this pipeline against

[1] http://avec2013-db.sspnet.eu/.

[2] http://dcapswoz.ict.usc.edu/.

[3] http://www.jeffcohn.net/resources/.

Table 1 Demographic and clinical data

	Control group	Patient group	*p* value
Age (years)	39.96 (7.87)	49.7 (12.68)	0.005
Age range	24-56	24-70	–
Men	17 (37.8%)	3 (15%)	0.068
Women	28 (62.2%)	17 (85%)	
Education (years)	16.69 (5.04)	10.1 (4.77)	0.0001
BDI-II score	6.49 (5.62)	21.8 (14.39)	0.0001
BDI-II>13	7 (15.6%)	14 (70.0%)	0.0001
STAI score	40.27 (9.27)	52.95 (10.13)	0.0001
STAI>39 (at least mild anxiety symptoms)	23 (51.1%)	18 (90.0%)	0.001
STAI>49 (moderate/severe anxiety symptoms)	12 (26.7%)	14 (70.0%)	0.002
Expert rating of depression severity[a]	2.43 (1.25)	4.85 (1.08)	0.001

BDI-II Beck Depression Inventory-II, *STAI* State-Trait Anxiety Inventory
p values are for independent samples t-tests or chi-square tests of proportions. Unless otherwise indicated values are means (SD)
[a]Conducted by psychologists based solely on participants' face videos

self-rated anxiety symptoms, as well as expert ratings of visible facial manifestations of depression. Investigating the co-occurrence of anxiety and depression symptoms is of clinical and research interest, in view of the significant overlap in phenotypic presentations of related disorders [31]. To the best of our knowledge, there is a single other recent report addressing the common co-occurrence of depression- and anxiety-related visual signs [32]. This earlier attempt did not employ persons with clinical diagnoses, a gap in the literature which is addressed in the current study.

2 Data collection

2.1 Participants

The study included two groups of participants: healthy volunteers ($n = 45$) aged 20–65 years without history of mental or neurological disorder, and patients suffering from MDD as diagnosed by their treating psychiatrists at the Psychiatry Outpatient Clinic, University Hospital of Heraklion ($n = 20$). Healthy volunteers were recruited through announcements in social media, flyers posted at public sites, as well as through personal referrals. Patients were informed about the study by their physicians (psychiatrists at the outpatient clinic) during regular appointments and if they consented verbally the details of the study were explained to them both verbally and in writing by a research assistant (psychologist). The study was approved by the Bioethics Committee of the University Hospital of Heraklion (296/7/06-04-2016) and the National Data Protection Authority (Protocol No. ΓΝ/ΕΞ/392-2/21-04-2016).

Demographic information and scores on self-report anxiety and depression instruments can be found in Table 1 for each group. Both groups consisted of individuals of Greek ethnicity, with the exception of 1 patient and 2 healthy participants (immigrants from Albania, Italy, and Germany) who were permanent residents of Greece for over 20 years and competent in reading, speaking and understanding the Greek language. Although an effort was made to match the two groups with respect to demographics, the patient group was older and had completed fewer years of formal education than the control group. The two groups did not differ significantly on the percentage of women, which was higher than the percentage of men in both groups and especially among patients (85%) in accordance with the clinical literature [33].

The generalizability of the developed method was tested on the AVEC'14 dataset, comprising 300 video recordings from 83 participants obtained in the context of two conditions which are very similar to tasks used in the present study, namely reading a neutral text passage and completing a question-and-answer session with the experimenter.

2.2 Psychological measurements

2.2.1 Self-reported symptoms of depression and anxiety

Two self-report questionnaires were administered to assess recent and ongoing depression and anxiety symptomatology, namely the Greek adaptations of the Beck Depression Inventory-II (BDI-II) and the State-Trait Anxiety Scale Form Y (trait anxiety subscale). The BDI-II [34] comprises 21 questions, each scored on 0–3, 0–4, or 0–5 point scales. These questions assess emotional and behavioral signs of depression such as body image, hypochondriasis, difficulty working, sleep loss, appetite loss, thoughts of self-punishment, suicidal ideation, and reduced libido. The higher the score, the more severe the depression symptoms, with scores above

13 points indicating mild to severe depression severity [34]. The STAI-Trait Anxiety Subscale [35] is a self-assessment tool of persistent symptoms (feelings, somatic complaints and behaviors) that are considered as core manifestations of anxiety as a characteristic of the individual. It consists of 20 items rated on a 1–4 point scale, with higher total scores indicating higher levels of anxiety. Scores above 39 points indicate clinically significant anxiety symptomatology [35].

2.2.2 Blinded expert annotation

In addition to clinical diagnosis, BDI-II, and STAI scores, facial expressions and nonverbal cues displayed by each participant over the entire recording session were evaluated independently, based on the recorded video by two psychologists. These ratings aimed at providing a more direct estimate of visual depression manifestations during the experiment. It was surmised that this measure would be more directly comparable to the output of a video processing algorithm than diagnosis, which took place at various times prior to recording, and subjective self-ratings of relevant symptoms. Each rater was blinded with respect to clinical diagnosis and experimental condition, and was instructed to rate each person on an 0-8 point scale of "*depression severity*", with 0 indicating complete absence of visual signs of depression, and 8 indicating the most severe visual signs of depression. Audio playback was turned off during rating, primarily in order to avoid revealing any information regarding the group of the participant, as well as of the task performed. In case of discrepancy between raters equal to or greater than 1.5 points on the scale, a third annotator was employed. The final annotation score registered was the grand average of all available annotations during the entire experimental session. Ratings were registered in real-time while viewing the recorded videos in CARMA,[4] a software for continuous affect rating and media annotation [36]. Absolute inter-rater agreement was adequate as indicated by an intraclass correlation coefficient (ICC, two-way mixed) of 0.85.

2.3 Stimuli and procedures

In designing the study special care was given to: (a) choosing the correct tools to assess depression-related symptoms of participants' everyday life, and (b) ensuring a wide range of experimental conditions to elicit the targeted emotions in the laboratory. The latter is based on the assumption that the quality and intensity of such emotions and their related facial expressions will be altered in the presence of significant depression symptomatology [37]. We adopted techniques that involved "*human-human interaction*" as well as "*human-computer interaction*" [38] in both "*social*" and "*non-social*" context as described below, while the full protocol is also summarized in Table 2.

Table 2 Data collection protocol in the main study

#	Task	Administration method
1	Acquaintance	Orally
2	Read and sign information sheet and consent form	In writing
3	Demographics and clinical history	Webpage
4	Relaxation (baseline)	BioTrace
5	Prolonged /a/ utterance	Orally
6	Positioning in front of the camera	–
7	Recall and description of positive experience (173 ± 111 s)	Orally
8	"Joy" clip (69 ± 2 s)	Webpage
9	Emotion ratings for "Joy" clip	Webpage
10	Relaxation (breathing exercise)	BioTrace
11	Reading aloud a neutral text (173 ± 29 s)	Webpage/orally
12	Recall and description of negative experience (139 ± 153 s)	Orally
13	"Sad" clip (310 ± 100 s)	Webpage
14	Emotion ratings for "Sad" clip	Webpage
15	Complete STAI	Webpage
16	Complete BDI-II	Webpage
17	Prolonged /a/ utterance	Orally

Conditions that contributed to the data analyzed in the present work are shown in bold (average ± SD duration in parentheses)

2.3.1 Non-social emotion elicitation setting

In the context of a pilot study involving $n = 10$ healthy volunteers aged 26–39 years (5 men) we evaluated thirteen video clips that were considered as suitable for eliciting each one of four of Ekman's six basic emotions (4 clips for joy, 3 clips for disgust, 3 clips for sadness, 3 clips for fear) [39]. After reviewing participant ratings (see Fig. S1a) the two excerpts associated with more unequivocal emotion elicitation were a scene from the movie "*Terms of Endearment*" for sadness and a scene from a popular Greek comedy series ("*Para Pente*") for joy. Very similar average ratings of emotional states were registered by the main study participants immediately after viewing each clip (Fig. S1b).

2.3.2 Social emotion-elicitation setting

In the context of a semi-structured interview with a research assistant (psychologist) participants were asked to describe a positive personal experience in detail and were probed to relive this experience as vividly as possible. In a sim-

[4] https://github.com/jmgirard/CARMA/.

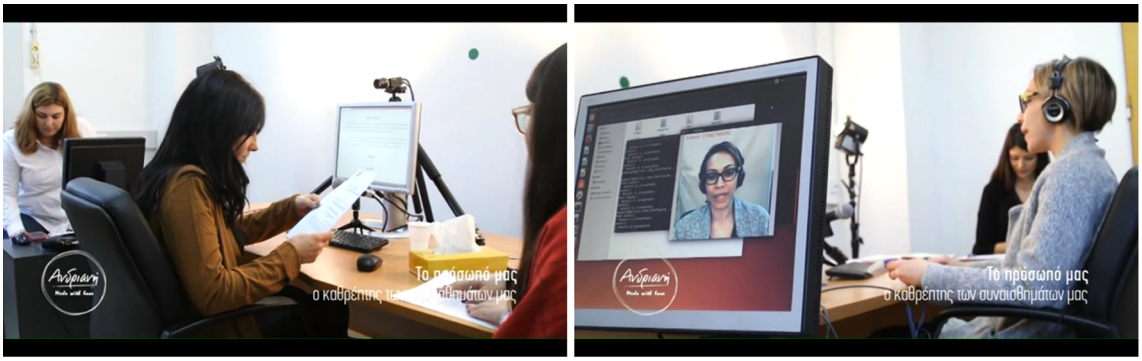


Fig. 1 Data collection experimental setup

ilar manner, prior to viewing the sadness clip participants were asked to describe a negative personal experience, which involved sadness or distress. A neutral baseline for the positive/negative experience description comprised reading aloud a 260-word narrative text describing a country excursion.

2.3.3 Experimental procedure

As shown in Fig. 1,[5] the participant was seated in front of a PC monitor where stimuli, rating scales and questionnaires were presented and where his/her responses were registered by a camera and a biosignal recording device. Two researchers conducted the experiment, one operated the stimulus delivery and recording devices and the second, a psychologist interacted with the participant. The psychologist obtained consent, provided instructions and additional explanations if needed, performed guided relaxation, and conducted the semi-structured interviews. Although a formal clinical evaluation of healthy volunteers was not part of the experimental protocol, the study psychologist was instructed to monitor verbal and nonverbal signs that may be indicative of an undiagnosed mood or anxiety disorder, and administer additional probing questions.

The order of conditions was fixed to ensure minimal cross-task "*emotional contamination*". Conditions intended to produce positive emotions were presented first, followed by a paced relaxation session before the neutral condition was administered. This was followed by conditions designed to induce negative emotions. Questionnaires were administered at the end of the sequence to minimize fatigue effects on the video recordings. On two predetermined occasions, the participant was guided through relaxation exercises, involving controlled breathing and brief mindfulness techniques to ensure that emotional states and stress levels returned to baseline levels. This took place prior to the description of positive experience/joy clip and, again, prior to the description of negative experience/sadness clip. Specifically, in the beginning of the protocol (c.f. Table 2 task #4) the participants were instructed by the research assistant on how to breathe in order to relax, while their heart rate and peripheral Blood Volume Pulse (BVP) were monitored through photoplethysmography using a NeXus-10 device (Mind Media, Netherlands). Participants were offered a second guided relaxation session immediately following the "*Joy*" video clip (Task #8 in Table 2) to help them resume baseline levels of emotional and psychophysiological states (i.e., as recorded at the beginning of the experimental session). This was ensured by monitoring BVP on the Nexus-10 device during the breathing exercise. Facial video data used in the present study originated from steps 7–8 and 11–13 of the study protocol, shown in bold. The total duration of the experiment ranged from 60–90 min.

A Point Grey Grasshopper®3 camera was employed to record high-resolution video at a high frame-rate. Camera settings were set to permit future assessment of the impact of recording quality on the algorithm efficiency. Benchmark tests showed that 80 frames per second (fps) and a resolution of 1920 × 1920 pixels was the maximum configuration that the available PC could support. Indirect lighting was applied to the participant's face to ensure uniform facial illumination and minimize shadows.

3 Video processing algorithm

Although three types of signals were recorded from all participants (i.e., visual, audio, and physiological), the present study focused on the video recordings, according to the algorithmic pipeline proposed in [4]. The first step involved extraction of the facial region in each frame and alignment of the successive facial images using OpenFace 2.0,[6] which is a widely used open source tool that offers reliable facial

[5] Figures of the setup are screen-shots from the interview given to a local TV channel (complete video available at: https://youtu.be/lH6Lo4S9KQ0).

[6] https://github.com/TadasBaltrusaitis/OpenFace/.

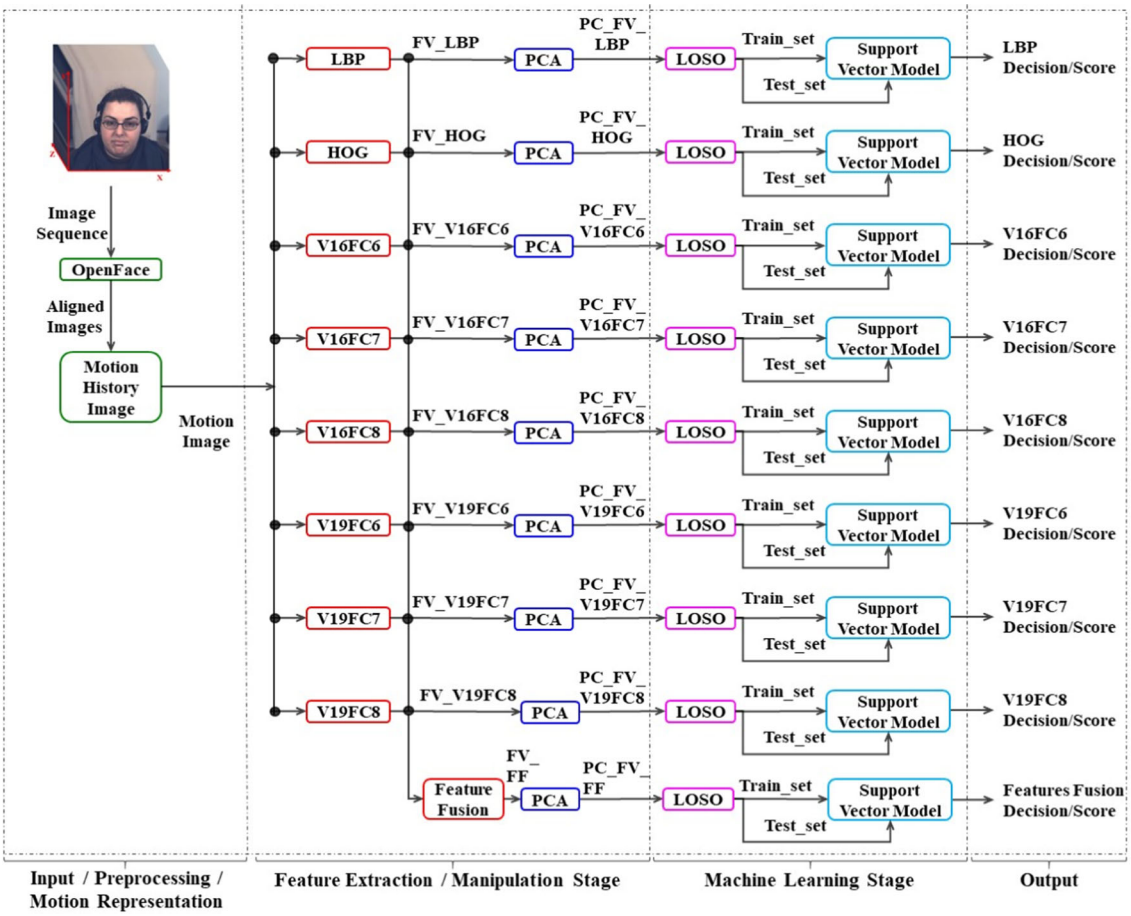


Fig. 2 Pipeline of the proposed methodology. Pipeline of the proposed methodology. *LBP* local binary patterns, *HOG* histogram of oriented gradients, *VxFCy* VGG-x fully connected layer y, *FV_* feature vector, *PCA* principal component analysis, *PC_* principal components, *LOSO* leave one subject out

landmark detection, head pose estimation, facial action unit recognition, and eye-gaze estimation [40]. Facial landmark detection offers the basis for face alignment and is performed using a Convolutional Experts Constrained Local Model (CE-CLM) [41]. It consists of a Convolutional Experts Network that computes a response map which helps to accurately localize individual landmarks by evaluating the landmark alignment probability at individual pixel locations. An expert layer votes on alignment probability. A Point Distribution Model (PDM) captures landmark shape variations and is used to regularize the shape. The next steps in our pipeline involved feature extraction based on dynamic facial signs, as described below. The complete pipeline of the proposed methodology is illustrated by Fig. 2.

3.1 Motion history image

Signs of depression are dynamic by nature, emanating from motion related patterns (e.g., motor retardation), information which is not conveyed by a static image. Therefore, there is the need to employ an algorithm which considers movement patterns over an image sequence (video versus image processing). Hereby, dynamic facial signs were represented through the motion history image (MHI), and feature extraction was based on this derived image. The latter is a gray scale image, where white pixels correspond to the most recent movements in the video, intermediate gray scale values correspond to less recent movements, and black pixels to the absence of movement. MHIs efficiently encode dynamic behavior over a longer period of time into a static image. Moreover, they capture motion flow, as well as the actual moving parts/regions in the video, while being sensitive to the direction of motion, a characteristic which is essential in facial expression analysis. Finally, they register the history of temporal changes in each pixel, thus keeping spatial relationships intact (e.g., eye positions) [42]. Therefore, they have been commonly successfully employed in motion analysis [43], human action recognition [44] as well as in facial action recognition from videos [45].

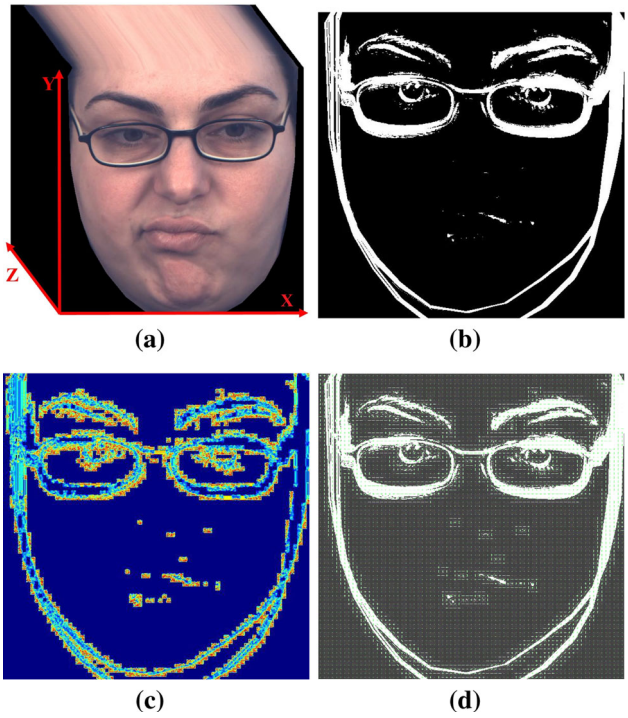


Fig. 3 Motion history image (**b**) as extracted from a sequence of aligned face images (**a**). Appearance-based features for LBP and HOG are shown in (**c**, **d**), respectively

The MHI H, with a resolution equal to that of the aligned faces, was computed based on an update function $\Psi(x, y)$ as follows:

$$H_i(x, y) = \begin{cases} 0 & i = 1 \\ i \cdot s & \Psi_i(x, y) = 1 \\ H_{(i-1)}(x, y) & \text{otherwise} \end{cases} \quad (1)$$

where $s = 255/N$, N the total number of video frames, (x, y) the position of the corresponding pixel, and i the frame number. $\Psi_i(x, y)$ represents the presence of movement, derived from the comparison of consecutive frames, using a threshold ξ:

$$\Psi_i(x, y) = \begin{cases} 1 & D_i(x, y) \geq \xi \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where $D_i(x, y)$ is defined as a difference distance:

$$D_i(x, y) = \left| I_i(x, y) - I_{(i-1)}(x, y) \right| \quad (3)$$

$I_i(x, y)$ is the pixel intensity value in (x, y) at the ith frame. The final MHI is the $H_N(x, y)$.

Feature extraction, explained in some detail at the following subsection, was performed on the MHI pseudo-images (rather than the original video frames). An example of the resulting visualization of the appearance-based features is illustrated in Fig. 3.

3.2 Feature extraction

The next step involved constructing meaningful appearance based descriptors to exploit intensity and texture based attributes. Two appearance-based descriptors were employed, Local Binary Patterns (LBP) and Histogram of Oriented Gradients (HOG), as implemented in MATLAB. LBP [46] entails dividing the image into partially overlapping cells. Each pixel of the cell is compared to its neighbors to produce a binary value (pattern). The resulting descriptor is a histogram which represents the occurrence of different patterns. LBP combines aspects of statistical and structural texture representation. As shown in [47], it can be treated as a special case of a multi-dimensional co-occurrence statistic, a common statistical texture measure. Considering the structural representation, LPBs efficiently encode texture primitives such as spots, flat areas, edges, edge ends, curves etc. [47]. For the proposed work the rotation invariant LBP was selected, which for two sets of {radius, neighborhood}, produces two feature vectors of size 1×59 for $\{1, 8\}$ and size 1×243 for $\{2, 16\}$. The two sets of parameters function supplementing one another, with $\{1, 8\}$ corresponding to micro-movements, while $\{2, 16\}$ to movements of larger scale. HOG [48] entails counting gradient orientations in a dense grid. Each image is divided into uniform and non-overlapping cells, the weighted histogram of binned gradient orientations for each cell is computed, and subsequently combined to form the final feature vector. The output corresponds to the concatenated individual histograms resulting in a single spatial HOG histogram. HOG descriptors are very popular in object detection tasks, and can be utilized in encoding facial regions [49], a property we are exploiting with the aim to add regional information to the features.

Additional features were obtained using the pre-trained Convolutional Neural Network (CNN) architecture developed by the Visual Geometry Group (VGG), University of Oxford, and previously tested by the winning 2014 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [50]. Relying on multiple convolutional layers, VGG is suitable for describing both textural and appearance-related features and was selected among other pre-trained networks providing a good trade-off between high depth/complexity versus performance while bearing in mind that this work does not target image classification tasks such the ILSVRC. We used two CNNs, VGG-16 consisting of 13 convolutional and 3 fully connected layers, and VGG-19 with 16 convolutional and 3 fully connected layers. The MHI image, with pixel values ranging from 0-255, was normalized by subtracting the mean pixel value. The input to VGG (a fixed-size 224×224 MHI image) passed through a stack of convolutional layers, with very small filters who had a receptive field of size 3×3 to capture the notion of left/right, up/down, and center. The convolution stride was fixed to 1 pixel and the spatial padding

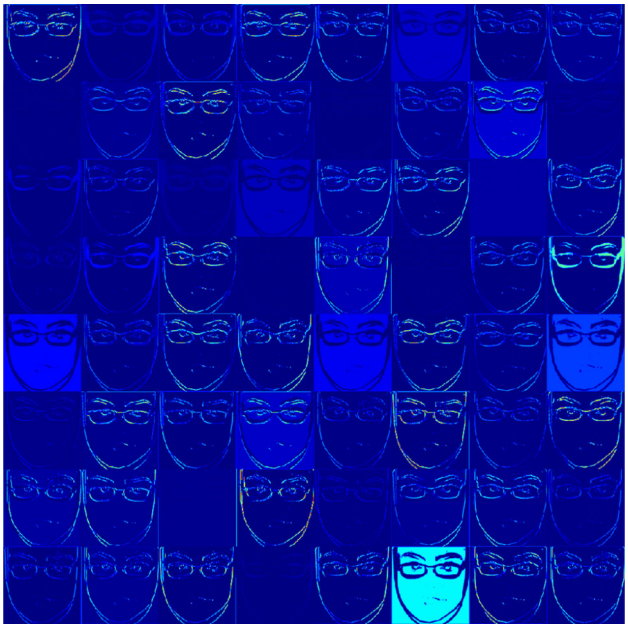


Fig. 4 Visualization of Pool ReLU1_1 activations. Each image represents the different channels of the specific layer

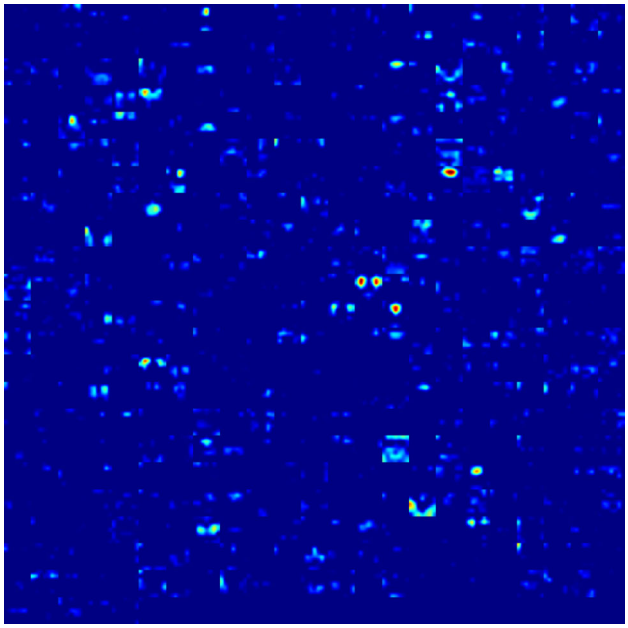


Fig. 5 Visualization of Pool 5 activations. As depicted in the illustration the different images are increased in number yet smaller in resolution, thus extracting different type of features than the previous layers (e.g., than in Fig. 4)

of a convolutional layer ensured that the input spatial resolution was preserved after convolution (i.e., padding was 1 pixel for 3×3 convolutional layers). Certain convolutional layers were followed by spatial pooling in five max-pooling layers. Max-pooling was performed, while a stack of convolutional layers (which had a different depth in different architectures) was followed by three fully connected layers. The final layer was the soft-max layer. The configuration of the fully connected layers was the same in all networks. All hidden layers used the Rectified Linear Unit (ReLU) activation function [51]. Both VGG-16 and VGG-19 were employed in the MATLAB model which was trained on a subset of the ImageNet database (from ILSVRC). Visualization of different activations is illustrated in Fig. 4 and Fig. 5.

3.3 Dimensionality reduction

Principal Component Analysis (PCA) was employed next on all descriptors for dimensionality reduction. PCA is one of the most popular methods for this purpose, and is based on the linear transformation of the original feature vector, into a set of principal components, resulting in uncorrelated data in the new space. For a dataset of N samples with M features PCA identifies an $M \times M$ coefficient matrix (component loadings) that maps each data vector from the original space to a new space of M principal components. Solutions with 5, 10, 20, 40, 45, and 50 principal components were tested separately. The number of principle components to be extracted was chosen empirically based on relative performance.

3.4 Classification and regression

The efficiency of the extracted features as indices of participant emotional characteristics and state was assessed using Support Vector Machines (SVM) to address binary classification and Support Vector Regression (SVR) for outcome score prediction. The chosen methods are well-established in the machine learning literature [52]. In the present work we use Linear SVMs due to their simplicity, interpretability, training efficiency, and the relatively low number of features used (principal components identified by PCA). The main purpose of SVM is to find an optimal hyperplane that discriminates the samples of the two classes in the feature space. The best hyperplane in a given SVM model corresponds to the one with the largest margin ("distance") between the two classes as defined by the selected support vectors i.e., the data points that are closer to the separating hyperplane. SVR is the extension of SVM for handling regression tasks. The performance of the SVM models was assessed with the following complementary measures: Cohen's Kappa [53], F1 score, accuracy, precision, and recall were used for the classification tasks, whereas Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) were used for the regression models. Model accuracy was calculated from the following confusion matrix:

$$C = \begin{bmatrix} \text{TP} & \text{FN} \\ \text{FP} & \text{TN} \end{bmatrix} \tag{4}$$

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

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{5}$$

where precision is given by:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

and recall by:

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

The F_1 score, is given by:

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{8}$$

Cohen's Kappa statistic [53], a chance and skew robust metric also based on the confusion matrix, is given by:

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \tag{9}$$

where p_0 is the proportion of accurately predicted decisions given by the accuracy formula as defined in (5), and p_e the proportion of expected chance agreement, given by:

$$p_e = \frac{M_a + M_b}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \tag{10}$$

where M_a and M_b are defined as follows:

$$M_a = \frac{(\text{TP} + \text{FN}) * (\text{TP} + \text{FP})}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \tag{11}$$

$$M_b = \frac{(\text{TN} + \text{FP}) * (\text{TN} + \text{FN})}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \tag{12}$$

RMSE and MAE are given by the following:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - a_i)^2} \tag{13}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|p_i - a_i| \tag{14}$$

where n is the number of samples, p the predicted value, and a the actual value.

3.5 Cross-validation

The evaluation of the proposed algorithm entailed Leave-One-Subject-Out (LOSO) cross-validation. This process was repeated as many times as the number of participants in the dataset (rather than samples); each time all samples of the given participant were held out from the training set and then used for testing. Cross-validation was performed separately in gender-independent and gender-dependent modes [10]. In the gender-dependent mode both training and testing of the model was conducted with features derived from either male or female participants.

Given the high computational requirements of preprocessing and feature extraction, tests were performed offline on a PC (Intel®Core™i7-4720HQ CPU @ 2.60 GHz, 8 GB RAM, 64-bit Windows 10 Home©Microsoft Corporation, 512 GB SSD). PCA, SVM/SVR and cross-validation were executed on a remote virtual machine with 8 virtual CPUs, 256 GB RAM, and 100 GB hard disk, running Ubuntu 16.04 LTS on a physical server with an Intel Xeon E5-2690 v3 2.6GHz processor.

3.6 Analytic strategy

The second aim (development and testing of an algorithmic pipeline) was pursued by comparing model performance against (a) group labels and (b) continuous assessment of depression symptomatology. During preliminary evaluation of the classification model the performance for high versus low BDI-II scores revealed very low performance (see Table S1), which further encouraged the already intended tests on continuous assessment, supported also by literature [4].

The third aim of the study involved applying the best performing model, developed in the context of the 2nd aim to perform continuous assessment of depression symptomatology, on the AVEC'14 dataset in order to evaluate the generalizability of the proposed methodology.

To pursue the fourth aim of the study we evaluated the relative performance of the proposed pipeline against self-rated depression symptoms, anxiety symptoms, and expert ratings of overt facial manifestations of depression (continuous assessment models).

4 Results

4.1 Psychological measures

As expected, self-reported depression (BDI-II) and anxiety (STAI) scores, as well as the percentage of persons scoring in the clinically significant range on each scale were significantly higher in the patient group. Moreover, the aver-

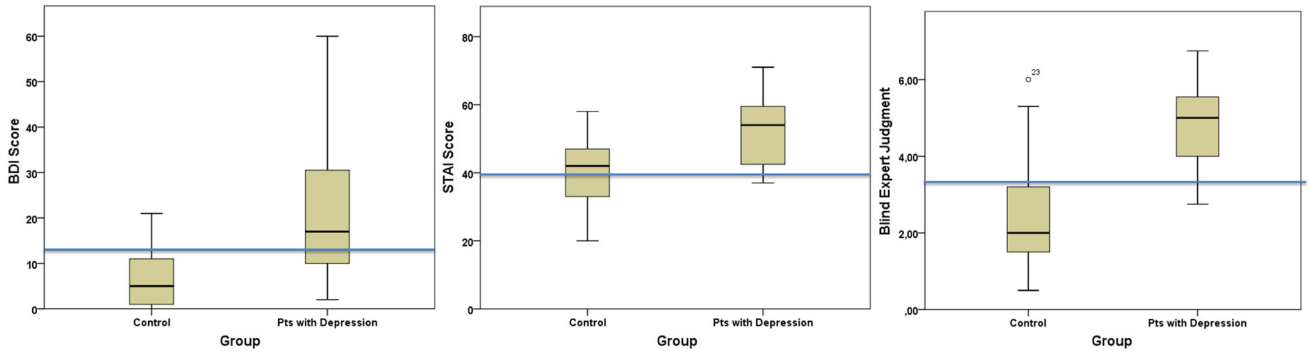


Fig. 6 Dispersion of BDI-II (left-hand panel), STAI (middle panel), and blinded expert judgment values (right-hand panel) per group displaying median (dark horizontal line), interquartile range (boxes), and range of values. The 13/14 point cutoff on BDI-II, 39/40 point cutoff on STAI, and 3.4/3.5 point cutoff on expert judgment are indicated by a blue horizontal line

age BDI-II and STAI scores obtained by the control group were well within the range of previous reports from normative samples of Greek speaking young adults [34,54]. There was, however, some evidence that the present sample was somewhat biased toward reporting frequent/more severe anxiety-related symptoms: (a) the percentage of participants in the control group reporting at least mild anxiety symptoms (51.1%) was considerably higher than the corresponding percentage reporting depression symptoms (15.6%), (b) patients reported moderate/severe anxiety symptomatology as frequently as mood disturbances (70.0%). It should be noted, however, that based on the semi-structured interview with the study psychologist, none of the participants in the control group met criteria for a mood or anxiety disorder.

Self-reported values of depression and anxiety symptoms (STAI and BDI-II scores, respectively) were strongly cross-correlated, as expected (r = .768). The degree of association between self-report measures and expert judgments of depression signs based on visual cues alone was in the moderate range, albeit considerably weaker, given the diverse nature of the two sets of measures. The slightly higher association between STAI and expert ratings ($r = 0.527$), as compared to the correlation between expert ratings and BDI-II ($r = 0.424$), may simply reflect the tendency of human raters to focus on manifest signs of psychological distress rather than depression per se.

The dispersion of BDI-II, STAI, and expert ratings of depression is shown in Fig. 6, revealing that the best separation between the two groups was achieved by expert ratings of depression relying solely on facial expressions. This impression was confirmed by Receiver Operating Characteristic Curve (ROC) analyses, revealing sensitivity/specificity estimates of 70/85% for BDI-II (at the recommended cut-off of 13/14 points), 75/50% for STAI (at the recommended cutoff of 39/40 points), and 85/90% for expert ratings (at the optimal cutoff of 3.4/3.5 points on the 0-8 point scale).

4.2 Model development and testing for depression assessment

The relative performance of various feature extraction and cross-validation schemes was first assessed on the current dataset against BDI-II scores (i.e., the only common continuous variable with the AVEC'14 dataset). Three participants did not complete one of the tasks, each for different reasons and a different task, bringing the total number of recordings to 322 (out of a possible of 325: 65 participants × 5 tasks). The proposed methodology was tested across all tasks, and for each condition and gender mode separately.

Performance of the categorical assessment models for high versus low BDI-II scores was very low when compared to previous tests [12], as indicated by F1 scores < 56.4% (see Table S1). Classification models against clinical diagnosis and expert judgment-based groupings performed comparably (F1 = 61.5% and 57.9%, respectively). Accordingly, the 3rd and 4th specific aims of the study were addressed via continuous assessment models.

Among continuous assessment models, the best performance was achieved by VGG-19 features derived from the passage reading condition in gender-independent mode (RMSE/MAE = 10.54/7.86; see Table 3). In this run, there were only two false negative cases (the model underestimated self-reported depression severity for two participants who scored > 55 points on BDI-II). In gender-specific mode our method relying on HOG features produced very similar results.

The same feature extraction and cross-validation scheme was applied to the AVEC'14 data from the comparable condition ("*Northwind*": passage reading for continuous assessment of BDI-II scores). Results indicated similar performance (RMSE=10.74, MAE=8.91 on the Development set, and RMSE=11.45, MAE=9.92 on the Test set; see Table 4). In relation to the results of previously reported approaches using the AVEC dataset the proposed method-

Table 3 Best-performing continuous assessment schemes in gender-based and gender-independent modes on the current dataset

Label	Gender mode	Condition	Feature	PCs	RMSE	MAE	Normalized	
							RMSE	MAE
BDI-II	Based	Read	HOG	20	10.59	7.46	16.81	11.84
BDI-II	Independent	Read	V19FC7	40	10.54	7.86	16.73	12.48
STAI	Based	Read	HOG	20	10.53	8.56	13.16	10.71
STAI	**Independent**	**Read**	**HOG**	**20**	**9.94**	**7.88**	**12.42**	**9.85**
Expert judgment	Based	Negative	HOG	20	1.61	1.37	20.17	17.07
Expert judgment	Independent	Joy	V19FC7	40	1.47	1.21	18.39	15.06

Read passage reading condition, *Negative* recall of negative experience, *PCs* number of extracted components, *RMSE* root mean square error, *MAE* mean absolute error, *HOG* histogram of oriented gradients, *V19FC7* fully connected layer 7 from VGG-19. Columns RMSE and MAE are in original units, while the normalized RMSE and MAE were rescaled from 0 to 100 to provide with a direct comparison across label

Table 4 Best-performing continuous assessment schemes against BDI-II scores in gender-based and gender-independent modes in the current and AVEC'14 datasets

Test partition	Condition	Gender mode	PCs	Feature	RMSE	MAE
Current dataset						
LOSO	Read	Based	20	HOG	10.59	7.46
LOSO	**Read**	**Independent**	**40**	**V19FC7**	**10.54**	**7.86**
AVEC'14 dataset						
Test	Freeform	Based	10	HOG	10.63	8.58
Test	Northwind	Based	10	HOG	11.45	9.92
Test	**Freeform**	**Independent**	**20**	**HOG**	**10.15**	**8.48**
Test	Northwind	Independent	40	HOG	10.95	9.22
Development	Freeform	Based	5	HOG	9.20	7.81
Development	Northwind	Based	40	HOG	10.74	8.91
Development	**Freeform**	**Independent**	**45**	**HOG**	**9.15**	**7.83**
Development	Northwind	Independent	40	HOG	10.97	9.33
LOSO	Both	Based	90	HOG	10.96	8.89
LOSO	Both	Independent	100	HOG	10.89	8.87

Read passage reading condition, *Northwind* passage reading task, *Freeform* oral question and answer session, *PCs* number of extracted principal components, *RMSE* root-mean-square error, *MAE* mean absolute error, *HOG* histogram of oriented gradients, *V19FC7* fully connected layer 7 from VGG-19, *LOSO* leave one subject out cross validation on the entire dataset

ology performed comparably to state-of-the-art methods as shown in Fig.7.

Although the best performing scheme for depression on the current dataset (Normalized RMSE = 16.73) was achieved with the VGG-19 features, as becomes apparent from both Tables 3 and 4, HOG features also proved to outperform the rest of the features in various setups.

4.3 Assessment of model specificity for depression versus anxiety

Model performance was evaluated for continuous assessment of BDI-II, STAI, and expert rating scores in both gender-dependent and gender-independent modes. The best-performing scheme in terms on both RMSE and MAE among those listed in Table 3 was the gender-independent model conducted on video recordings from the passage reading condition. Relative model performance can be evaluated by examining log-transformed differences between actual and predicted STAI scores using the following formula:

$$\log_2 \frac{A_{\text{norm}}}{P_{\text{norm}}} \tag{15}$$

where A_{norm} is the normalized observed value and P_{norm} the normalized predicted value. Figure 8 also includes a Bland-Altman plot [55] that helps to compare the agreement between the observed and predicted STAI score. It was created using the following formula:

$$M = \alpha - \beta, \quad A = \frac{\alpha + \beta}{2} \tag{16}$$

where α is the predicted and β the observed value. Figure 8b plots the difference between observed and predicted values

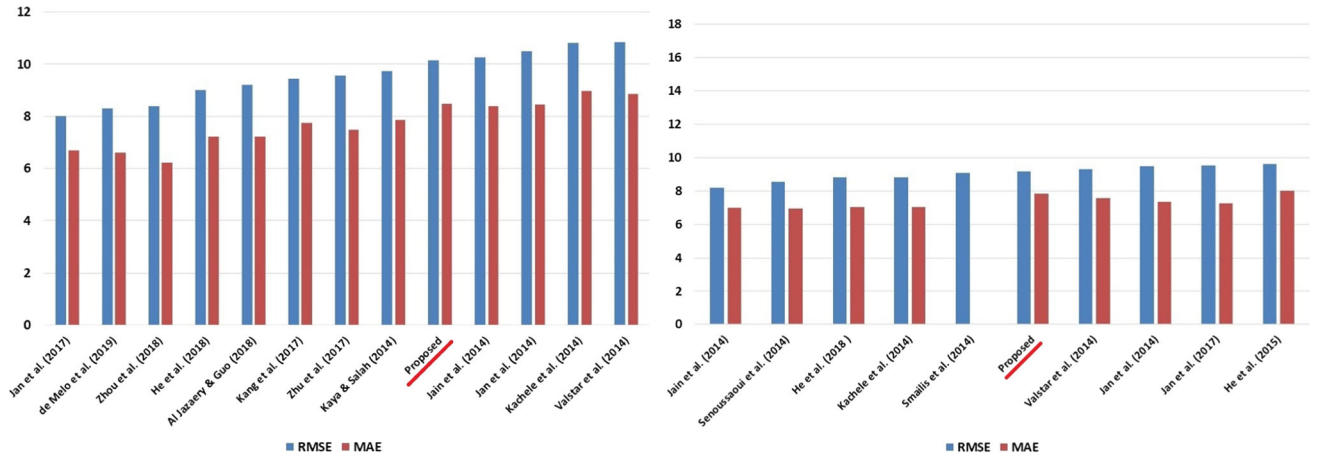
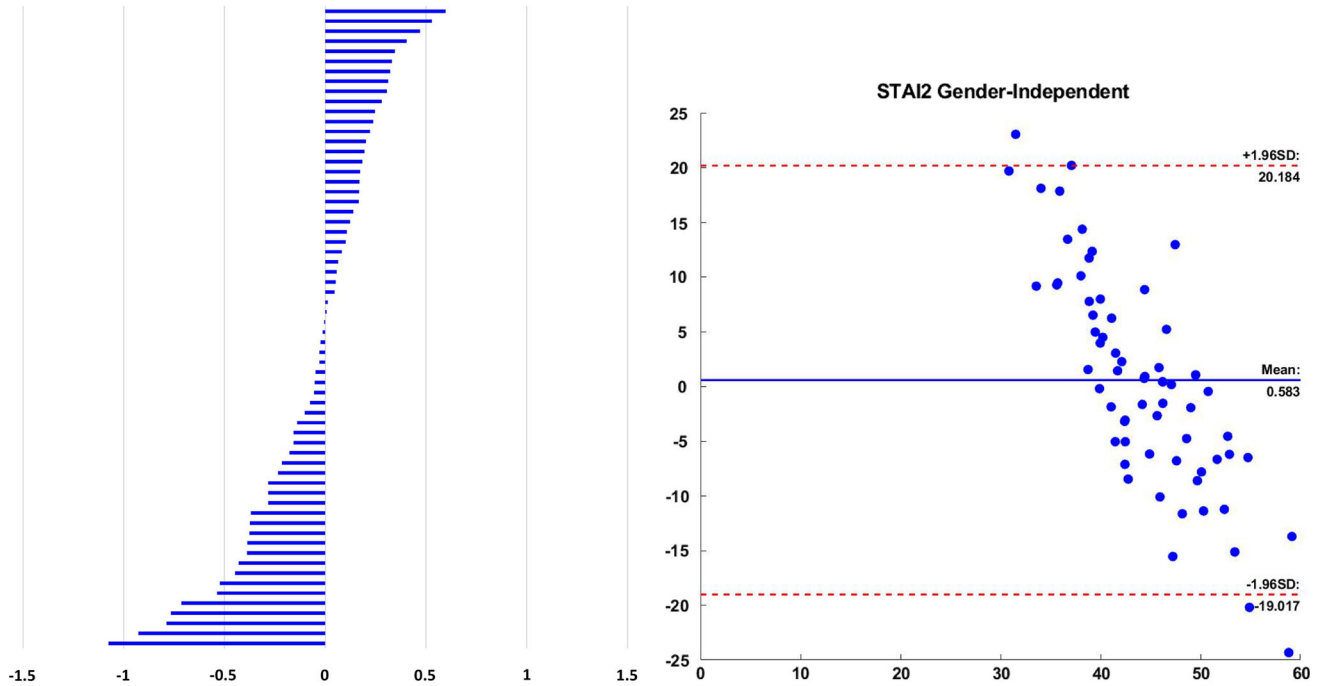


Fig. 7 Comparison of our approach (Proposed) with previously reported results on the AVEC'14 Test (left panel) and Development (right panel) datasets relying solely on facial, video-based features. Results from the proposed continuous assessment algorithm were obtained with HOG features derived from the Northwind condition in gender-based mode



(a) Log(Actual_norm/ Predicted_norm)

Fig. 8 Results of continuous assessment: prediction of STAI scores based on HOG features derived from the passage reading condition in gender-independent mode. Upper panel: log transformed differences between actual and predicted STAI scores for each participant. Lower panel: Bland–Altman plot displaying the distribution of observed-predicted value differences (y-axis) over the range of STAI scores (x-axis)

(M; plotted on the y axis) as a function of their average (A; plotted on the x axis). For future clinical applications of automated assessment methods, false negative results (i.e., failure to detect significantly high scores on a psychopathological trait) are critical. In our results, such events correspond to cases where our models significantly underestimated self-reported depression/anxiety or expert-judgment ratings as indicated by scores $>$ 1.96 SDs from the sample mean.

With respect to STAI scores (Fig. 8), both best-performing models underestimated self-reported anxiety severity in two participants scoring $>$ 50 points in the scale. Both models relied on HOG features derived from the passage read-

Table 5 Comparison of the AVEC'14 with our dataset on several parameters

	AVEC'14	Current dataset
Number of recordings	300	322
Number of participants	83	65
Age [M (SD)]	31.5 (12.3)	42.4 (11.92)
Men	32.67%	30.77%
BDI-II [M (SD)]	15.06 (11.9)	11.2 (11.73)
Participants	Volunteers	Volunteers and patients
Country	Germany	Greece
Protocol	Non-social	Interpersonal and non-social
Setup	Independent	Controlled
Illumination	Non-controlled	Controlled indirect lighting
Image resolution (pixels)	640 × 480	1920 × 1920
Facial image size (pixels)	112 × 112	600 × 600
Frame rate (fps)	30	80

ing condition. With respect to expert judgment values, the best prediction was achieved marginally by HOG features derived from the Negative Experience Recall condition in gender-based mode (data not shown). In this run, the model significantly underestimated expert ratings of depression severity in two participants.

In auxiliary tests we run SVRs trained on data from all but one condition and tested on the remaining condition. Overall, RMSE values were slightly higher across various features than for the best-performing (passage reading) condition. Moreover, classification models against participants groupings based on STAI and expert rating scores performed poorly as well (see Table S1).

5 Discussion

5.1 Algorithm generalization

The proposed algorithmic pipeline produced comparable results in predicting BDI-II scores in two diverse datasets in terms of sample characteristics (clinical and cultural background) and video image quality, as outlined in Table 5. The robustness of the proposed algorithm is further supported by the fact that the AVEC dataset consisted of pre-extracted dynamic facial landmarks whereas the present dataset consisted of raw video recordings.

Besides the direct comparison of results across the two datasets, it is difficult to establish common grounds between the present and previous work due to important methodological differences. For instance, a recent report of depression severity prediction using the Pittsburgh dataset relied on serial video recordings [56] whereas a single measurement was available in the current study. In other reports deep learning was employed for both algorithm training and tuning comparing [13,16] whereas in the current work this technique was used solely for feature extraction.

5.2 Algorithm specificity

Given that the proposed pipeline produced considerably better prediction of STAI as compared to either BDI-II scores or expert ratings of depression in the current dataset, one may surmise that it is primarily sensitive to facial features more closely associated with self-reported anxiety (STAI scores). This pattern was present across experimental conditions, thus likely reflected relatively characteristic, dynamic facial signs, although best performance was achieved with video recordings obtained during an emotionally and cognitively non-challenging condition (i.e., reading a neutral passage). It should be noted, however, that best results (prediction of STAI or BDI-II scores) were obtained with video recordings from the neutral, passage-reading task in gender-independent mode.

As shown in the Bland-Altman plot of Fig. 8, the most notable failure involved underestimation of STAI scores, given the higher associated clinical risk. Among the four patients in this category, two suffered from severe depression and were treated with high doses of anti-depressants, which may have affected the dynamics of facial expressions. Another patient spoke Greek as a second language and experienced some difficulty in reading the text, while the fourth patient also experienced some difficulty in reading due to reduced visual acuity.

Clinical diagnosis of depression did not emerge as a robust outcome variable in binary classification schemes. In part this finding may be attributed to the considerable overlap between the two study groups on self-reported depression symptomatology (BDI-II scores) and expert-rated facial signs of depression as shown in Fig. 6. The fact that the patients were

not treatment-naïve may in part account for this overlap by allowing cases who had responded adequately to treatment and experienced significant remission of depression symptoms to be included in the clinical group. Moreover, none of the participants in the control group met criteria for a mood or anxiety disorder (based on the clinical interview conducted as part of the experiment); yet they reported elevated symptoms of trait anxiety on STAI. As a result the overlap between the two clinical groups on STAI scores was even more extensive than the overlap on BDI-II and experts ratings of depression. Given that the proposed algorithm appears to be primarily sensitive to overt signs of anxiety symptoms this extensive overlap may have contributed to the poor classification results of the proposed algorithm against clinical diagnosis of depression.

5.3 Algorithm development and performance

A notable finding of the present study concerns the relatively poor binary classification performance of our algorithm with respect to all four outcome measures (including clinically acceptable cut-off on STAI scores). For clinical purposes, however, methods such as the one developed here are not intended as standalone diagnostic systems; they are evaluated as decision support tools providing tentative recommendations for further clinical exploration. In addition, the relatively poor classification performance may reflect the continuous nature of depression symptomatology. This is likely reflected in a corresponding continuous variation of facial motion dynamic patterns across patients rendering continuous assessment more appropriate. This problem is compounded by the potential moderating role of clinical factors that may affect the intensity and quality of facial expression of depression symptoms (e.g., negative mood, apathy, helplessness). Such factors include the use of Selective Serotonin Reuptake Inhibitors (SSRIs) medications which may enhance apathy and therefore reduce emotional expressivity, as well as person-specific characteristics (illness-related cognitions and personality characteristics).

Another notable finding was that the deep learning approach for feature extraction outperformed the other approaches on most tests (data partitions and experimental conditions). Given that only the generic VGG was employed in the present work, it is highly probable that further training and tuning of the network would significantly improve prediction of depression severity. The highly competitive performance of deep learning has been noted in several research areas.

The fact that deep learning does not rely on pre-specified rules, in a manner similar to human cognition, may account for its superiority. The performance of pre-trained deep neuronal networks with a relatively small dataset, such as the one tested in the present work, generates great promises regarding their capacity to provide clinically meaningful results for depression assessment if trained with sufficiently large training datasets.

The experience gained through the experimental tests leads to the conclusion that performance metrics need to be jointly evaluated. The majority of previous reports relied on a single metric (i.e., accuracy) which does not reliably show the capacity of a model, especially in cases of highly unbalanced datasets. By default Cohen's Kappa encompasses most pertinent information and it is not surprising that the majority of complementary metrics generally agree to Kappa values across studies. Moreover, the F1-score does not necessarily reflect accurate recognition across all classes (i.e., a high F1-score may be associated with good recognition in one class and poor recognition in another) and does not consider the level of chance. Feinstein and Cicchetti [57] however present a more moderate view regarding Kappa. Their conclusions suggest that Kappa may be overly conservative, as it appears to have serious interpretative problems in the presence of high skew just like accuracy but in the opposite direction. This fact provides an explanation for the very low Kappa values for the categorical assessment models.

5.4 Data related issues

A note is in place regarding the small but notable superiority of the gender-independent models. This finding, however, may be an artifact of the small sample size which was reduced to marginal levels in the gender-dependent modes, especially when such models were applied to the (relatively few) male participants. Furthermore, although the ratio of depressed to non-depressed participants in the present study does not reflect the population prevalence of depression disorders-based ratios, where the percentage of persons with diagnosis of depression individuals are about 17.5% for women and 14.6% for men [58], still it is better suited for a machine learning problem where the ratio of the different classes cannot be small as it highly impacts the training of the model.

Although not a primary aim of the present study, it was assumed that, in spite of their significant computational cost, high video acquisition specifications (i.e., illumination, image resolution, frame rate) would significantly improve algorithm performance. However, this assumption was not supported by experimental results: binary classification results on the high resolution videos from our dataset performed slightly worse than on the AVEC'14 dataset, while for the continuous assessment they performed at the same level. Therefore, the specifications of the video recordings do not have a great impact on the performance.

5.5 Limitations and future plans

In view of the nature of the problem pursued and the specific requirements of feature extraction and cross-validation methods used in the present study, perhaps the most important limitation concerns overall sample size. The size of the patient group was especially small given the significant variability on clinical characteristics present. These limitations reduced the evaluative power of the proposed algorithm in gender-based mode. Datasets from larger patient samples are paramount in order to properly test the generalizability of the algorithm presented in the current work. This daunting task is particularly important in order to eventually take into account the multitude of factors that may affect overt signs of emotional states and symptoms encompassed in depression, including disease type, severity, and duration, and the impact of treatment (pharmacological and psychotherapeutic). A complementary experimental setup would involve multiple serial measurements (video, biosignals, self- and clinician ratings of psychoemotional state and symptoms) from a relatively smaller group of patients. This approach has produced very promising results in a recent study involving multiple serial video recordings of clinician interviews with patients with depression. The goal of this approach would be to identify features that show systematic variability over time in relation to clinical outcomes.

During the recall and description of positive and negative experience, although the participants were instructed to look at a fixed green mark on the wall, sometimes the participants showed a tendency in changing their head pose to face the interviewer. Although this was not a frequent occurrence its potential impact on algorithm performance was not formally evaluated.

Moreover, in view of the poor performance of the algorithm against expert judgment of depression, it would be helpful to evaluate the proposed algorithm against expert ratings of anxiety/distress levels derived from visual cues. This test is crucial to address the notion that the proposed algorithm is indeed more sensitive to facial signs of anxiety/distress. It should further be noted that emotional and behavioral symptoms of depression in the present study were assessed using a single self-report scale (BDI-II). Clinician-administered scales, such as the Hamilton Depression Rating Scale (HAM-D) employed in the Pittsburgh University dataset [56], may be more sensitive to clinically significant signs and symptoms of depression.

Extraction of additional visual and non-visual features (auditory from spontaneous speech and continuous phonation [59,60], biosignals) from the existing dataset and systematic evaluation of fusion schemes is forthcoming. Multimodal approaches entail further challenges including the selection of adequate fusion schemes [4,61], such as feature-level fusion (mere concatenation of individual features) and decision-level fusion (combining predictions derived independently from each modality through AND and OR operands, or through weighing). The Posterior Probability Classification Model [61] and stacking [62] represent alternative fusion techniques.

6 Conclusions

The proposed work provided a number of insights, while identifying many questions open to further investigation. A novel dataset was presented, along with the data collection methods, which went beyond the state-of-the-art in terms of specifications, protocol, annotation, and participants. The proposed methodology was evaluated on two diverse datasets achieving a competitive to the state-of-the-art performance for predicting individual BDI-II scores. In addition to predicting BDI-II scores, the performance was proven consistent also in terms of predicting the anxiety levels based on STAI. The current results support recent arguments in favor of regression methods as more suitable to the nature of targeted clinical outcomes (e.g., depression severity).

Acknowledgements Funding was provided by State Scholarships Foundation (Grant No. Legacy fund in the memory of Maria Zaousi).

References

1. World Health Organization (WHO) (2017). http://www.who.int/mental_health/management/depression/en/
2. First, M.B.: Structured Clinical Interview for DSM-IV-TR Axis I Disorders: Patient Edition. Biometrics Research Department, Columbia University, New York (2005)
3. Chmielewski, M., Clark, L.A., Bagby, R.M., Watson, D.: Method matters: understanding diagnostic reliability in DSM-IV and DSM-5. J. Abnorm. Psychol. **124**(3), 764 (2015)
4. Pampouchidou, A., Simos, P.G., Marias, K., Meriaudeau, F., Yang, F., Pediaditis, M., Tsiknakis, M.: Automatic assessment of depression based on visual cues: a systematic review. IEEE Trans. Affect. Comput. **10**(4), 445–470 (2019)
5. Falagas, M., Vardakas, K., Vergidis, P.: Under-diagnosis of common chronic diseases: prevalence and impact on human health. Int. J. Clin. Pract. **61**(9), 1569–1579 (2007)
6. Comer, J.S.: Introduction to the special series: applying new technologies to extend the scope and accessibility of mental health care. Cogn. Behav. Pract. **22**(3), 253–257 (2015)
7. Ellgring, H.: Non-verbal Communication in Depression. Cambridge University Press, New York (2007)
8. Waxer, P.H.: Therapist training in nonverbal communication. I: nonverbal cues for depression. J. Clin. Psychol. **30**(2), 215 (1974)
9. Girard, J.M., Cohn, J.F.: Automated audiovisual depression analysis. Curr. Opin. Psychol. **4**, 75–79 (2015)
10. Alghowinem, S., Goecke, R., Wagner, M., Epps, J., Hyett, M., Parker, G., Breakspear, M.: Multimodal depression detection: fusion analysis of paralinguistic, head pose and eye gaze behaviors. IEEE Trans. Affect. Comput. **9**(4), 478–490 (2018)
11. Cohn, J.F., Kruez, T.S., Matthews, I., Yang, Y., Nguyen, M.H., Padilla, M.T., Zhou, F., De La Torre, F.: Detecting depression from

facial actions and vocal prosody. In: 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, pp. 1–7. IEEE (2009)
12. Pampouchidou, A., Pediaditis, M., Maridaki, A., Awais, M., Vazakopoulou, C.M., Sfakianakis, S., Tsiknakis, M., Simos, P., Marias, K., Yang, F., Meriaudeau, F.: Quantitative comparison of motion history image variants for video-based depression assessment. EURASIP J. Image Video Process. **2017**(1), 64 (2017)
13. Jan, A., Meng, H., Gaus, Y.F.B.A., Zhang, F.: Artificial intelligent system for automatic depression level analysis through visual and vocal expressions. IEEE Trans. Cogn. Dev. Syst. **10**(3), 668–680 (2018)
14. de Melo, W.C., Granger, E., Hadid, A.: Combining global and local convolutional 3D networks for detecting depression from facial expressions. In: 2019 14th IEEE International Conference on Automatic Face Gesture Recognition (FG 2019), pp. 1–8 (2019)
15. Kang, Y., Jiang, X., Yin, Y., Shang, Y., Zhou, X.: Deep transformation learning for depression diagnosis from facial images. In: Zhou, J., Wang, Y., Sun, Z., Xu, Y., Shen, L., Feng, J., Shan, S., Qiao, Y., Guo, Z., Yu, S. (eds.) Biometric Recognition, pp. 13–22. Springer, Cham (2017)
16. Zhu, Y., Shang, Y., Shao, Z., Guo, G.: Automated depression diagnosis based on deep networks to encode facial appearance and dynamics. IEEE Trans. Affect. Comput. **9**(4), 578–584 (2018)
17. Kaya, H., Salah, A.A.: Eyes whisper depression: a CCA based multimodal approach. In: International Conference on Multimedia, pp. 961–964. ACM (2014)
18. Jain, V., Crowley, J.L., Dey, A.K., Lux, A.: Depression estimation using audiovisual features and fisher vector encoding. In: 4th ACM International Workshop on Audio/Visual Emotion Challenge (AVEC '14), pp. 87–91. ACM (2014)
19. Jan, A., Meng, H., Gaus, Y.F.A., Zhang, F., Turabzadeh, S.: Automatic depression scale prediction using facial expression dynamics and regression. In: 4th ACM International Workshop on Audio/Visual Emotion Challenge (AVEC '14), pp. 73–80. ACM (2014)
20. Kächele, M., Glodek, M., Zharkov, D., Meudt, S., Schwenker, F.: Fusion of audio-visual features using hierarchical classifier systems for the recognition of affective states and the state of depression. In: 3rd International Conference on Pattern Recognition Applications and Methods, pp. 671–678. SciTePress (2014)
21. Valstar, M., Schuller, B., Smith, K., Almaev, T., Eyben, F., Krajewski, J., Cowie, R., Pantic, M.: AVEC 2014: 3D dimensional affect and depression recognition challenge. In: 4th ACM International Workshop on Audio/Visual Emotion Challenge (AVEC '14), pp. 3–10. ACM (2014)
22. Jazaery, M.A., Guo, G.: Video-based depression level analysis by encoding deep spatiotemporal features. IEEE Trans. Affect. Comput. (2018). https://doi.org/10.1109/TAFFC.2018.2870884
23. He, L., Jiang, D., Sahli, H.: Automatic depression analysis using dynamic facial appearance descriptor and dirichlet process fisher encoding. IEEE Trans. Multimed. **21**(6), 1476–1486 (2019)
24. Zhou, X., Jin, K., Shang, Y., Guo, G.: Visually interpretable representation learning for depression recognition from facial images. IEEE Trans. Affect. Comput. (2018). https://doi.org/10.1109/TAFFC.2018.2828819
25. Chentsova-Dutton, Y.E., Tsai, J.L., Gotlib, I.H.: Further evidence for the cultural norm hypothesis: positive emotion in depressed and Control European American and Asian American Women. Cult. Divers. Ethn. Minor. Psychol. **16**(2), 284 (2010)
26. Girard, J.M., McDuff, D.: Historical heterogeneity predicts smiling: evidence from large-scale observational analyses. In: 2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017), pp. 719–726 (2017)
27. Malhi, G.S., Parker, G.B., Gladstone, G., Wilhelm, K., Mitchell, P.B.: Recognizing the anxious face of depression. J. Nerv. Ment. Dis. **190**(6), 366–373 (2002)
28. Katz, M.M., Wetzler, S., Cloitre, M., Swann, A., Secunda, S., Mendels, J., Robins, E.: Expressive characteristics of anxiety in depressed men and women. J. Affect. Disord. **28**(4), 267–277 (1993)
29. Kotov, R., Krueger, R.F., Watson, D., Achenbach, T.M., Althoff, R.R., Bagby, R.M., Brown, T.A., Carpenter, W.T., Caspi, A., Clark, L.A., et al.: The hierarchical taxonomy of psychopathology (HiTOP): a dimensional alternative to rraditional nosologies. J. Abnorm. Psychol. **126**(4), 454 (2017)
30. Kaufman, J., Charney, D.: Comorbidity of mood and anxiety disorders. Depress. Anxiety **12**(S1), 69–76 (2000)
31. Cohn, J.F., Cummins, N., Epps, J., Goecke, R., Joshi, J., Scherer, S.: Multimodal Assessment of Depression from Behavioral Signals, pp. 375–417. Association for Computing Machinery, Morgan & Claypool, New York, Williston (2018)
32. Jaiswal, S., Song, S., Valstar, M.: Automatic prediction of depression and anxiety from behaviour and personality attributes. In: 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), pp. 1–7 (2019)
33. Bogner, H.R., Gallo, J.J.: Are higher rates of depression in women accounted for by differential symptom reporting? Soc. Psych. Psych. Epidemiol. **39**(2), 126–132 (2004)
34. Giannakou, M., Roussi, P., Kosmides, M., Kiosseoglou, G., Adamopoulou, A., Garyfallos, G.: Adaptation of the beck depression inventory-II to greek population. Hell. J. Psychol. **10**(2), 120–146 (2013)
35. Fountoulakis, K.N., Papadopoulou, M., Kleanthous, S., Papadopoulou, A., Bizeli, V., Nimatoudis, I., Iacovides, A., Kaprinis, G.S.: Reliability and psychometric properties of the greek translation of the state-trait anxiety inventory form Y: preliminary data. Ann. Gen. Psych. **5**(1), 2 (2006)
36. Girard, J.M.: CARMA: software for continuous affect rating and media annotation. J. Open Res. Softw. **2**(1), e5 (2014)
37. Davies, H., Wolz, I., Leppanen, J., Fernandez-Aranda, F., Schmidt, U., Tchanturia, K.: Facial expression to emotional stimuli in non-psychotic disorders: a systematic review and meta-analysis. Neurosci. Biobehav. Rev. **64**, 252–271 (2016)
38. Kächele, M., Schels, M., Schwenker, F.: The influence of annotation, corpus design, and evaluation on the outcome of automatic classification of human emotions. Front. ICT **3**, 27 (2016)
39. Ekman, P.: Are there basic emotions? Psychol. Rev. **99**, 550–553 (1992)
40. Baltrušaitis, T., Zadeh, A., Lim, Y.C., Morency, L.: Openface 2.0: facial behavior analysis toolkit. In: 2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018), pp. 59–66 (2018)
41. Zadeh, A., Lim, Y.C., Baltrusaitis, T., Morency, L.: Convolutional experts constrained local model for 3D facial landmark detection. In: 2017 IEEE International Conference on Computer Vision Workshop (ICCVW), pp. 2519–2528 (2017)
42. Ahad, M.A.R.: Motion History Images for Action Recognition and Understanding. Springer, New York (2012)
43. Ahad, M.A.R., Tan, J.K., Kim, H., Ishikawa, S.: Motion history image: its variants and applications. Mach. Vis. Appl. **23**(2), 255–281 (2012)
44. Bobick, A., Davis, J.: Real-time recognition of activity using temporal templates. In: Proceedings Third IEEE Workshop on Applications of Computer Vision. WACV'96, pp. 39–42 (1996)
45. Valstar, M., Pantic, M., Patras, I.: Motion history for facial action detection in video. In: 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583), vol. 1, pp. 635–640 (2004)

46. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
47. Mäenpää, T., Pietikäinen, M.: Texture analysis with local binary patterns. In: Chen, C.H., Pau, L.F., Wang, P.S.P. (eds.) Handbook of Pattern Recognition and Computer Vision, pp. 197–216. World Scientific, Singapore (2005)
48. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, pp. 886–893 (2005)
49. Chen, J., Chen, Z., Chi, Z., Fu, H.: Facial expression recognition based on facial components detection and HOG features. In: International workshops on electrical and computer engineering subfields, pp. 884–888 (2014)
50. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-scale Image Recognition (2014). arXiv preprint arXiv:1409.1556
51. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems, Curran Associates Inc., USA, NIPS'12, pp. 1097–1105 (2012)
52. Cristianini, N., Shawe-Taylor, J., et al.: An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press, Cambridge (2000)
53. Cohen, J.: Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. Psychol. Bull. **70**(4), 213 (1968)
54. Karekla, M., Michaelides, M.P.: Validation and invariance testing of the greek adaptation of the acceptance and action questionnaire-II across clinical vs. nonclinical samples and sexes. J. Context. Behav. Sci. **6**(1), 119–124 (2017)
55. Bland, J.M., Altman, D.G.: Statistical methods for assessing agreement between two methods of clinical measurement. Int. J. Nurs. Stud. **47**(8), 931–936 (2010)
56. Dibeklioğlu, H., Hammal, Z., Cohn, J.F.: Dynamic multimodal measurement of depression severity using deep autoencoding. IEEE J. Biomed. Health Inf. **22**(2), 525–536 (2018)
57. Feinstein, A.R., Cicchetti, D.V.: High agreement but low kappa: I. The problems of two paradoxes. J. Clin. Epidemiol. **43**(6), 543–549 (1990)
58. Stylianidis, S., Pantelidou, S., Chondros, P., Roelandt, J., Barbato, A.: Prevalence of mental disorders in a Greek island. Psychiatriki **25**(1), 19–26 (2014)
59. Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., Quatieri, T.F.: A review of depression and suicide risk assessment using speech analysis. Speech Commun. **71**, 10–49 (2015)
60. Simantiraki, O., Charonyktakis, P., Pampouchidou, A., Tsiknakis, M., Cooke, M.: Glottal source features for automatic speech-based depression assessment. In: INTERSPEECH, pp. 2700–2704 (2017)
61. Pampouchidou, A., Simantiraki, O., Fazlollahi, A., Pediaditis, M., Manousos, D., Roniotis, A., Giannakakis, G., Meriaudeau, F., Simos, P., Marias, K., Yang, F., Tsiknakis, M.: Depression assessment by fusing high and low level features from audio, video, and text. In: Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge, ACM, New York, NY, USA, AVEC '16, pp. 27–34 (2016)
62. Sfakianakis, S., Bei, E.S., Zervakis, M.: Stacking of network based classifiers with application in breast cancer classification. In: XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016, pp. 1079–1084. Springer (2016)
63. Gross, J.J., Levenson, R.W.: Emotion elicitation using films. Cognit. Emot. **9**(1), 87–108 (1995)

Dr A. Pampouchidou PhD, is a researcher in the area of Affective Computing. She carried out her post-graduate studies at the University of Burgundy, ImVia Laboratory, obtaining her PhD degree in November 2018 and her masters degree in June 2011, both in the fields of Image Processing and Machine Learning. She received her bachelors degree in Applied Informatics and Multimedia, from the former Technological Educational Institute of Crete, in 2004 (the institute has evolved into the Hellenic Mediterranean University). Further, she has worked as external collaborator at the BMI and CBML research laboratories, carrying out various research activities, and has served as a reviewer for several peer-reviewed conferences and journals (i.e. IEEE Transactions of Affective Computing, Biomedical Signal Processing and Control, Journal of Biomedical and Health Informatics). Her research interests include Facial Expression Recognition, Facial Image Analysis, Affective Computing, Nonverbal Communication, Image Processing, and Pattern Recognition.

M. Pediaditis received the engineering degree in Electrical Engineering—Biomedical Engineering from Graz University of Technology (TUG), Austria, in 2008. He received his PhD in 2014 from the Institute of Health Care Engineering, TUG, in collaboration with the Computational BioMedicine Laboratory, FORTH, where he worked on facial expression analysis in epilepsy. Subsequently, he continued to work on national and EU-funded research projects. He has deep-seated experience in numerical electromagnetic field dosimetry, image and signal processing, as well as human motion analysis for medical applications using computer vision. From 2017 to 2019, he worked as a postdoctoral researcher at IBM Research Zurich with a focus on deep learning techniques for patient monitoring. His current research at FORTH focuses on using recurrent neuronal networks for video-based patient monitoring with emphasis on real-time alarming, patient health status modelling and anomaly detection.

E. Kazantzaki is a PhD candidate at the University of Crete, Medical School, Department of psychiatric and behavioral sciences, and has joined CBML since 2012. She has an MSc in Brain and Mind in the field of neuroscience from the Medical School of University of Crete and a BA in Psychology from the University of Crete. Eleni has received PhD fellowships from Leventis Foundation, Hellenic Foundation for Research and Innovation and Foundation for Research and Technology. She has participated in psycho-emotional and cognitive research in various international and national R&D projects including PMedicine, Semeoticons and iManageCancer, MyPal. Her activities focus on patient empowerment, personal health records, psycho-cognitive, personality and emotional assessment in patients with chronic and life-threatening diseases and on early signs of serious mental disorders such as schizophrenia and bipolar disorder.

Dr. S. Sfakianakis received his BSc and MSc diplomas from the Department of Informatics and Telecommunications of the University of Athens and his PhD from the School of Electrical and Computer Engineering of the Technical University of Crete. Since 2000, he has been with the Computational Biomedicine Laboratory of FORTH-ICS, working on integrating systems in the field of biomedicine, semantic interoperability, and building tools and services for intelligent data analysis. He has participated as work package leader in numerous

European research projects, focusing on the development of innovative ICT solutions to support large-scale transcription research on cancer, the discovery of biological cancer markers, transcription medicine, and the design and implementation of computational infrastructures for the integration of data and services. His research interests include biomedical computing, web-based and cloud-based software architectures, and data mining and analysis with modern computing tools. He has published more than 60 articles in international journals and conference proceedings related to his areas of expertise.

I. A. Apostolaki joined the Master of Science in Clinical Psychology at University of Nicosia in September 2018. She received her Bachelor of Arts in the field of Psychology from the European University Cyprus (2017). Mrs. Apostolaki volunteered as a psychologist in the Psychiatric Clinic of the University General Hospital of Heraklion in Crete. During that period, she also worked for the project facial image analysis for emotion recognition (2016–2018).

K. Argyraki received her bachelor degrees in Psychology from Staffordshire University in 2017. Following her graduation, she started her practical training at the Psychiatric Clinic of the University Hospital of Heraklion in 2018. Her training involved the coordination of patient groups at the clinic, as well as attending patients' sessions with their therapists. Further, she contributed to a research project carried out by the University of Crete in terms of annotating audiovisual recordings for signs of depression.

D. Manousos is working as technical staff at Computational Biomedicine Laboratory of Institute of Computer Science of FORTH from 2007 until today. He received his basic degree as Software Engineer from the Informatics Engineering department of TEI Crete on 2005. He has since worked in a variety of research projects as a research and development engineer. Among the projects involved, he has designed and implemented telemedicine applications such as electronic/personal health records (PHR, EHR), developing high-efficiency image and video processing algorithms. He has also been involved in 3D modeling and programming of online games. He has experience in a variety of programming domains such as full stack developer using SpringBoot, Angular, Hibernate, ASP.NET MVC, SQL (MySQL, Postgres, SQL Server), UI (JavaScript, HTML5, CSS3), mobile application developer using Android, MATLAB & C++ for image and video processing, Blender3D for game modeling as well as in Arduino microcontroller programming.

F. Meriaudeau was born on March 18, 1971. He received the masters degree in physics at Dijon University, France, and as an engineering degree (FIRST) in material sciences in 1994. He also obtained a Ph.D. in image processing at the same University in 1997. He was a postdoc for a year at The Oak Ridge National Laboratory. He is currently Professeur des Universites at the University of Burgundy. He was the director of the Institute Health and Analytics (2017/2018) at the Universiti Teknologi PETRONAS Malaysia and was the Director of the Le2i (UMR CNRS) France, which has more than 200 staff members, from 2011 to 2016. His research interests were focused on image processing for non-conventional imaging systems (UV, IR, polarization) and more recently on medical/biomedical imaging. He coordinated an Erasmus Mundus Master in the field of Computer Vision and Robotics from 2006 to 2010 and was the Vice-president for International Affairs for the University of Burgundy from 2010 to 2012. He has authored and co-authored more than 150 international publications and holds three patents.

K. Marias PhD, is an Associate Professor in Medical Image Processing at the Department of Electrical & Computer Engineering at the Hellenic Mediterranean University in Greece and is the head and founder of the Computational Biomedicine Laboratory at FORTH-ICS. ?e served as Principal Researcher at the Institute of Computer Science (ICS-FORTH) from 2006 to 2017. During 2000–2002, he worked as a Researcher at the University of Oxford and from 2003 to 2006 as Associated Researcher at FORTHICS. He received his PhD in Medical Image Analysis and Medical Physics from UCL Royal Free & University College Medical School working jointly with the University of Oxford, UK. He has coordinated two EC projects on cancer modeling (ContraCancrum and TUMOR projects) and has actively participated in several other EC-funded projects developing ICT technology for personalized medicine. He is co-author of more than 200 papers in international journals, books and conference proceedings focusing on medical image processing and analysis, biomedical informatics, image-based modeling and radiomics/deep learning medical imaging applications.

F. Yang received the BS degree in electrical engineering from the University of Lanzhou, China, in 1982 and the MS (computer science) and PhD degrees (image processing) from the University of Burgundy, France, in 1994 and 1998, respectively. She is currently a full professor at University of Burgundy, France. From 1996, Fan YANG is an active member of the team "Sensors & Hardware Architecture for Real-time Image Processing" in the Le2i laboratory. Her research interests include pattern recognition, neural network, multispectral imaging, parallelism and real-time implementation, and more specifically, biometric image processing: algorithms and architectures. From the beginning of her carrier, she published 40 papers in international peer-reviewed journals, three books and chapters and authored more than 80 conference papers. She has cosupervised or supervised 20 PhD students since 2000. She was a reviewer of significant scientific journals (IEEE Trans. on Neural Networks, on Circuits and Systems for Video Technology, on Circuits and Systems, Pattern Recognition, Pattern Recognition letters) and international conferences.

Prof M. Tsiknakis received his masters (1983) and PhD (1989) degrees from the University of Bradford, UK. He performed his post-doctoral training at the University of Bradford (1990–1991) and the Institute of Computer Science at FORTH (1992). Subsequently, he was elected as a Principal Researcher at the Computational Biomedicine Laboratory (CBML) of FORTH/ICS. He was an eEurope/eHealth award winner in 2003 and the recipient of FORTH's award for the most innovative applied research in 2004. Since 2012, he is a Professor of Biomedical Informatics and eHealth at the Department of Electrical and Computer Engineering at the Hellenic Mediterranean University and a visiting Professor at the Computational Biomedicine Laboratory of FORTH/ICS. He is an Associate Editor of IEEE Journal of Biomedical and Health Informatics and Editorial Board of the European Biomedical Informatics Journal. He is the author of over 300 publications. His main areas of expertise include approaches for semantic health data integration and interoperability of health information systems; affective computing and its application in developing smart eHealth solutions; service platforms for pervasive eHealth and mHealth services.

Dr M. Basta was born in Heraklion/Crete/Greece. She graduated at the Medical School of Crete in 1997. She completed her residency in Psychiatry in the Department of Psychiatry, University Hospital of Crete, in 2006. During her residency, she was trained in Sleep

Medicine for 6 months in the Sleep Research and Treatment Center, Pennsylvania State University, Hershey, PA. She acquired her PhD Diploma from the Medical School of Crete in 2003 in Sleep Medicine. Between 2006 and 2007, she did a 13-month postdoctoral fellowship in Sleep Medicine in the Sleep Research and Treatment Center, Pennsylvania State University, Hershey, PA, USA. From 2007 to 2012, she was an Attending Psychiatrist in Venizeleio General Hospital, Heraklion, Crete, Greece, in the Mental Health Center of Heraklion, Crete, Greece, and in the Department of Psychiatry, University Hospital, Heraklion, Crete, Greece. During 2013–2017, she was an Assistant Professor of Psychiatry, University of Crete, Greece, School of Medicine (tenure since January 2017), and during 2018–2020 an Associate Professor of Psychiatry, University of Crete, Greece, School of Medicine (tenure). She has published over 50 original articles/review articles, most of which are on field of Sleep Medicine (citations: 1976, h-index:21). Dr Basta is currently involved in several research grants as a co-investigator/primary investigator and is a member of the European Sleep Research Society (ESRS) and the American Association of Sleep Medicine (AASM) and since 2015 is qualified with the title of "European Somnologist, expert in Sleep Medicine".

Dr A. N. Vgontzas is a Professor of Psychiatry at the School of Medicine of the University of Crete. He has over 25 years of experience at the sector of physiology and pathophysiology of sleep and has over 220 publications original papers at international scientific journals, with over 17,000 citations (h-index = 65) and significant contribution to the field of Sleep Disorders Medicine. His research interests focus on Neuroendocrinology and Neuroimmunology of Sleep and its disorders. During the last decades, his focus has been on phenotyping and treatment of insomnia. Also, in the last 5 years, he has completed two major funded studies as principal investigator or co-investigator at the field of geriatric neuropsychiatry and the study of biological, genetic and psychosocial factors, which influence the cognitive functionality at middle-aged or elderly healthy volunteers and/or patients with mild cognitive impairment or dementia.

P. Simos received his PhD degree in Experimental Psychology-Biopsychology (1995) from Southern Illinois University. He served as Assistant and Associate Professor at the Departments of Neurosurgery, University of Texas-Houston Medical School, and Psychology, University of Crete, Greece. He is currently Professor of Developmental Neuropsychology at the School of Medicine, University of Crete. His research has been supported by several federal and national grants and focuses on neuropsychological and brain imaging studies of reading and memory using magnetoencephalography, MRI and fMRI with children and adults. Ongoing studies explore psychoeducational, emotional, and neurophysiological profiles associated with specific reading disability, ADHD and neurodegenerative disorders. He has also developed and adapted in Greek several psychometric instruments for cognitive and linguistic abilities across the lifespan.