Int J Multimed Info Retr
DOI 10.1007/s13735-017-0128-9

REGULAR PAPER

Computational framework for emotional VAD prediction using regularized Extreme Learning Machine

Zied Guendil[1] · **Zied Lachiri**[1] · **Choubeila Maaoui**[2]

Received: 17 March 2017 / Revised: 11 May 2017 / Accepted: 23 May 2017

Abstract With the advancement of Human Computer interaction and affective computing, emotion estimation becomes a very interesting area of research. In literature, the majority of emotion recognition systems presents an insufficiency due to the complexity of processing a huge number of physiological data and analyzing various kind of emotions in one framework. The aim of this paper is to present a rigorous and effective computational framework for humans affect recognition and classification through arousal valence and dominance dimensions. In the proposed algorithm, physiological instances from the multimodal emotion DEAP dataset has been used for the analysis and characterization of emotional pattern. Physiological features were employed to predict VAD levels via Extreme Learning Machine. We adopted a feature-level fusion to exploit the complementary information of some physiological sensors in order to improve the classification performance. The proposed framework was also evaluated in a V–A quadrant by predicting four emotional classes. The obtained results proves the robustness and correctness of our proposed framework compared to other recent studies. We can also confirm the sufficiency of the R-ELM when it was applied for the estimation and recognition of emotional responses.

Keywords Physiological data · Extreme Learning Machine · VA quadrant · DEAP · VAD space

✉ Zied Guendil
zied.guendil@enit.rnu.tn

Zied Lachiri
zied.lachiri@enit.rnu.tn

Choubeila Maaoui
choubeila.maaoui@univ-lorraine.fr

1 LR-SITI Laboratory, University Tunis El Manar, BP 37, Belvdre, 1002 Tunis, Tunisie

2 LCOMS Laboratory, University of Lorraine, ISEA, 7 rue Marconi, 57070 Metz, France

1 Introduction

Over the two last decades, the area of affective computing was growing and received more attention [1,2,4,6]. The capacity of a computer to understand human emotion and perform the appropriate actions is one of the key focus areas of research in Human Computer Interaction (HCI). In fact, the machine should interpret the emotional state of humans and adapt its behavior to them. Picard [1] proposed the first theories to establish a better understanding and recognition of emotional states. Emotions can be estimated from different modalities like facial expressions, gestures, speech, and physiological signals [2].

Several methods were used to induce emotions. The participant in emotional stimuli can be exposed to a set of pictures like (IAPS) [3], listen to music [4] or watch film videos [6–8]. Some emotions like stress and engagement can be sensed when a human is driving [9] or playing games [10]. Physiological measures can be required from multiple or single subject using various biosensors.

Two models of affective states representation were proposed. Ekman [11] presented a discrete model defined through a set of six basic emotions: joy, sadness, fear, anger, disgust, and surprise. A second model proposed by Russell [12] places emotions in the dimensional arousal and valence space. The valence differentiates negative and positive emotions, while the arousal order differentiates emotional states from being calm to being exited. Another dimension called dominance can be used sometimes to know whether the subject is controlling the situation or not.

Many systems were developed for the recognition of human emotion according to discrete or dimensional model by using various physiological modalities and a diversity of induction protocols [13]. But developed systems in general suffer from various drawbacks and their performance depending on factors such as the number of emotions, the number of subjects, type of processing, and also the used machine learning techniques. In our research, we focused on recent studies that investigate arousal and valence dimensions to measure and classify emotions.

One of the crucial steps on pattern recognition is to extract features, which can be used to detect the emotional content of the signal. A variety of statistical, time, and frequency domain features were extracted by many researchers. Kim and Andre [4] have extracted 110 features from four kind of physiological signals such as ECG, EMG, Skin conductance and respiration.

Recently, many researchers have used multiresolution tools like: Fourier transform [14], wavelet transform [15], and HHT [16]. Advanced machine learning techniques like k-nearest neighbor (KNN), regression tree [17], Bayesian networks, support vector machines (SVMs) [18] are used to analyze data with high-dimensional features or having a large volume of samples.

The success of these methods depends on the effectiveness of learning methods. Extreme learning machine is a recent learning algorithm widely used in regression and pattern classification. ELM has many advantages over other classification methods, and promising results were obtained when it was applied in many areas of research, such as biomedical engineering [19], computer vision [20], image processing [21], and system modeling and prediction [22].

ELM is fast and stable in training, easy in implementation, and accurate in modeling and prediction. The goal of this paper is to exploit ELM algorithm for a relevant prediction and classification of arousal, valence, and dominance performed by the fusion of wavelet transform features.

The remainder of this paper is organized as follows: we first give a state of the art on the recent works in Sect. 1. The used dataset is described in Sect. 2. Next, we detail the proposed emotion estimation method than the feature extraction and the classification steps were described in Sect. 4. We give an explicit description of the extreme learning machine technique in Sect. 5. In Sect. 6 we present the results obtained after the application of ELM and the feature-level fusion strategy. Section 7 concludes the paper.

2 Related work

A great interest was given to complex affect recognition in two or three continuous dimensional space [24–26] due to some important challenges like the difficulty in representing and analyzing a large number of emotions in one framework and the lack of validation of the framework through measured signals with complex emotions. Many approaches were developed for humans affect sensing and prediction in a multidimensional space based on analysis of emotional responses induced by multimedia content.

In some studies, emotions are classified according to three dimensions valence, arousal, and dominance [23].The majority of the existing methods over the literature discriminate three or two categories of valence and arousal in order to simplify the problem of classifying the six basic emotions defined by the discrete theory [11].

Another technique can be adopted to reduce the dimensional emotion problem to a four categories problem by separating emotional responses in a quadrant of 2D V–A space.

Music videos have been used in many reported researches to make patients feel different affective states. The aim of all previous works is to improve the recognition accuracy of emotional arousal and valence and to determine the most relevant physiological features which can be used for the characterization of emotional pattern.

Ibug research group created DEAP and MAHNOB, two recent physiological databases for human affect quantification and analysis, and several physiological modalities were recorded from different participants when they were watching music videos; they tagged implicitly their feeling according to arousal, and valence dimensions, data file, and annotation details are available for free which create indirectly an open challenge between many researchers who are interested in emotion analysis. It encourages them to test and evaluate their emotion recognition frameworks.

The developed systems differ in the proposed ways of emotion annotation and relabeling, but the common point is the use of two- or three-dimensional arousal, valence, and dominance, and most authors studied 2 or 3 class problem (low, medium, high).

The number of physiological features extracted and the classification techniques differ from an author to another, in some studies the number of samples used is not reported.

Koelstra et al. [24] used 106 features with SVM classifier applied to data from DEAP database. They studied two levels of arousal and valence.

Soleymani et al. [25] studied three levels of arousal and valence. They extracted features from EEG and gaze data signals, and they adopted two fusion strategies, a feature-level fusion and decision feature fusion. In this paper, we

Table 1 Overview of the systems for emotional VA recognition from physiological data

References	Physiological modalities	Machine learning technique	Database	Results	Dimensions
Soleymani et al. [25]	EEG and Eye gaze	SVM	MAHNOB	76% for arousal	3 levels
				68% for valence	
Koelstra et al. [26]	EEG and physiological signals	SVM	DEAP	62% for valence	2 levels
				57% for arousal	
Soleymani et al. [28]	Physiological signals	SVM	MAHNOB	46% for arousal	3 levels
				45% for valence	
Godin and Campagne [27]	Physiological signals	Naive Bayes classifier	DEAP	63% for valence	2 levels
				57% for arousal	
Our method	Physiological signals	R-ELM	DEAP	73.43% for valence	2 levels
				72.65% for arousal	
				69.53% for dominance	
				53% in VA quadrant	4 classes
Chung et al. [34]	EEG	Bayesian classifier	DEAP	66.6% for valence	2 levels
				66.4% for arousal	
				53.4% for valence	3 levels
				51% for arousal	
Zhang et al. [35]	EEG	SVM	DEAP	75.19% for valence	2 levels
				81.74% for arousal	
Zheng Zhu and Lu [36]	EEG	SVM and GELM	DEAP	69.67% in VA quadrant	4 classes

have studied and reviewed the research work on emotion prediction in two-dimensional space.

An explicit state of the art's for some recent surveyed frameworks is presented in Table 1.

3 Elm trends in emotion analysis and novel contributions

Extreme learning machines are certainly gaining in popularity during the last 5 years, and it is considered as a new feat at the existing machine learning models. It has been widely used in various applications to overcome the slow training speed and overfitting problems of the conventional neural network learning algorithms. ELM algorithms were invested in analyzing emotions, especially for estimating the children affective state in intelligent learning application [37], for improving the emotion recognition in speech and glottal signals [38] and also for recognizing emotion in the wild [39]. These recent works based on ELM showed an outstanding efficiency on emotional pattern prediction. The main contribution of our study was to enhance the prediction of emotional VAD levels from physiological features deeply analyzed by using regularized extreme learning machine algorithm. This ultimate goal was approximately reached when we obtained an improved classification results for two levels of VAD and also for the prediction of four classes of emotion in the VA quadrant.

4 Emotional dataset

Recent advances in emotion recognition have motivated many research teams to create their databases like MIT [1], AuBT [4], MAHNOB-HCI [28] and DEAP [26]. In our study, physiological signals from DEAP were used with annotation files and have undergone a specific structuring described in detail in Sect. 5.3.

4.1 DEAP corpus

DEAP is a multimodal database for the analysis of human emotion. The electroencephalogram (EEG) and peripheral physiological signals of 32 participants were recorded when they were watching 40 1-minute-long excerpts of music videos. Participants rated their sensed emotions in terms of arousal, valence, like/dislike, dominance, and familiarity. For 22 of the 32 participants, the frontal face video was also recorded. More details about the DEAP dataset are given in [26].

5 Outline of the proposed approach

The proposed methodology consists of the three major steps: (1) preprocessing of physiological data (2) feature extraction step, and (3) emotion classification stage. The architecture of the VAD recognition framework is illustrated in Fig. 1.

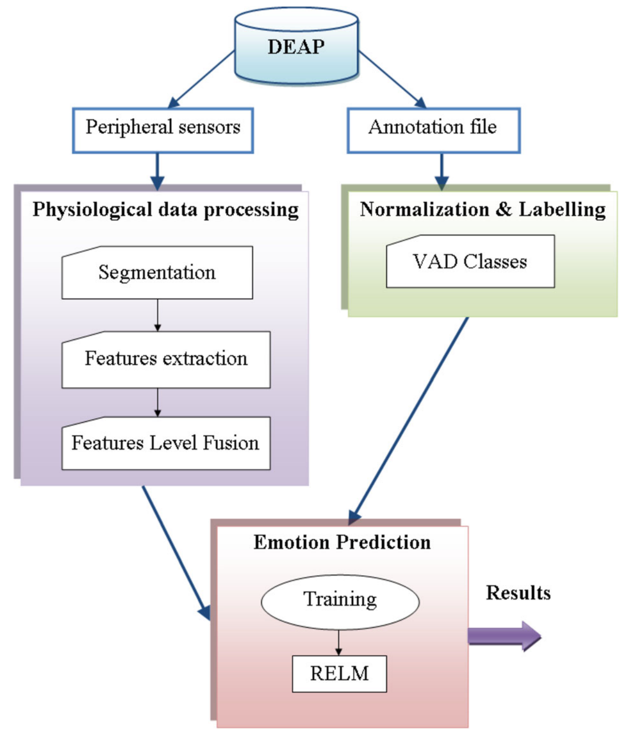


Fig. 1 The proposed framework of emotional VAD estimation

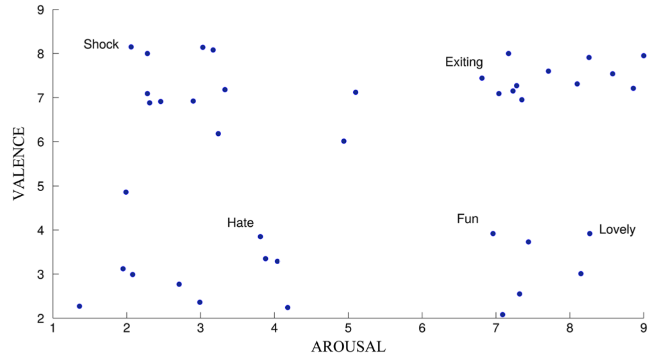


Fig. 2 Distribution of DEAP dataset in the VA space, each point denotes the location of one sentence corresponding to arousal valence scores given by the subject n1

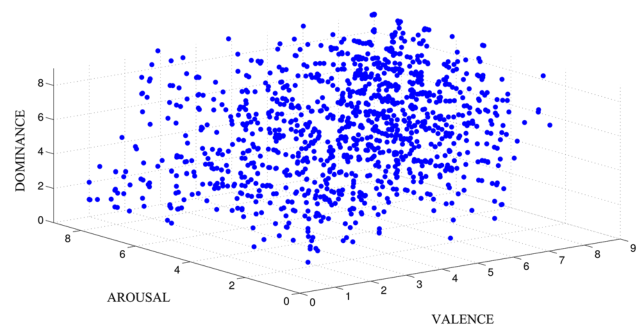


Fig. 3 Emotion projection in VAD space, each point denotes the position of one tagged physiological instance in the new 3D emotional space

5.1 Dimensional arousal and valence space

Dimensional emotion representation [28] is based on some psychological theories. The emotion plane representation can be viewed as a continuous 2D space where each point corresponds to a separate emotional state. The two capital dimensions of this plane are valence (V) and arousal (A). Valence component is more difficult, and it can be determined by the right questions and questionnaires. For DEAP database, valence scale varies from 1 (unpleasant) to 9 (pleasant). Arousal, on the other hand, represents the intensity of the affective state, and it is ordered also from 1 (passive, calm) to 9 (active). Each emotional state can be considered as a linear combination of these two dimensions, and some emotions induced in DEAP and rated by one participant are given in Fig. 2.

5.2 Emotion representation in the VAD space

The projection of emotion in the 3D continuous space (valence, arousal and dominance) in short (VAD space)is a new way [5] to characterize and describe humans affect. For the used DEAP database, a total number of 1280 instances (40 trials for each 32 subjects) containing various emotions categories like (fun, happy, joy, cheerful, melancholy, depressing, terrible, exiting, love, sad, mellow, shock, hate) are represented in 3D space as shown in Fig. 3. Each emotion was placed in the 3D space based on the VAD values provided by each participant.

5.3 Preprocessing step

5.3.1 Data segmentation

We proceed to the segmentation of data derived from eight physiological sensors, and in order to optimize and to reduce the number of the used samples, EEG signals were not used because they were recorded via 32 sensors, and physiological signals of each participant were trimmed to a fixed length of 8064. The total number of instances obtained for the eight biosensors after segmentation was 10,240.

5.3.2 Normalization of annotations and label building

The second treatment is the normalization of the annotation scores given by all participants. In fact, each induced emotion was relabeled based on the median value of valence, arousal, and dominance scales. We considered an annotation as a high label if the given score is over 4.5. Otherwise, it is a low label which generates two classes for arousal: valence and dominance.

5.4 Extraction of physiological features

Physiological activity is considered as an important component of an emotion. In the next step, we have to extract emotional features effectively from physiological signals. Through the literature, authors proposed a large list of potential features to characterize emotion [4]. The wavelet transform is suitable for multiresolution analysis, where the signal can be presented and analyzed at different frequencies and timescales. We have used Daubechies wavelet transform coefficients for feature extraction [30] because its shape looks like the waveforms of the analyzed signals. It locates the frequency variation of each signal which generally reflects physiological changes associated with emotion sensing. A set of features as variance and blink rate of EOG signal and others statistical features were carefully chosen for each physiological modality in order to be extracted from detail wavelet coefficients. Empirical results allowed us to conclude that the signal decomposition up to five levels is sufficient to capture the information from physiological signals. The detailed list of the extracted features used for the building of a pertinent physiological descriptor is given in Table 2.

6 Extreme learning machine

Through the literature, the usage of different classification algorithms is increasing. Extreme learning machine (ELM) is a new learning technique proposed by Huang et al. [31] for single hidden-layer feedforward neural networks (SLFNs), and it is a simple and fast method that can create a model from high-dimensional data sets. ELM algorithm classification chooses randomly the hidden node and analytically determines the output weights of SLFNs. Basically, an ELM implementation trains an SLFN in two main stages [33].

Table 2 Extracted features from various physiological modalities

Physiological sensors	Feature-based wavelet
EOG	Horizontal variance, vertical variance blink rate
GSR	Variance, zero crossing, min ratio
TEMP	Standard derivation, mean min ratio, max ratio
RESP	Standard derivation, pulsation max Ratio range (greatest breath)
EMG	Variance, Wilson amplitude
PPG	Standard derivation, (HRV) heart rate variability, RMSSD, HRV median HRV mean

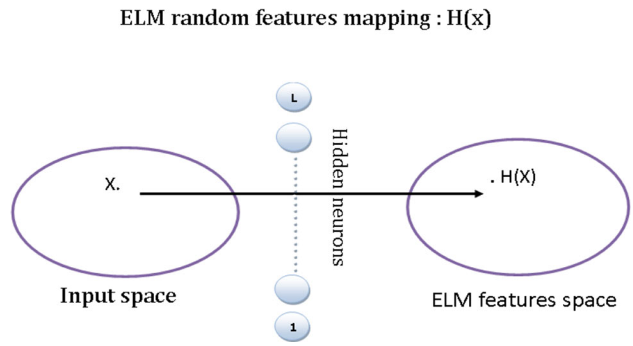


Fig. 4 ELM features mapping and feature space

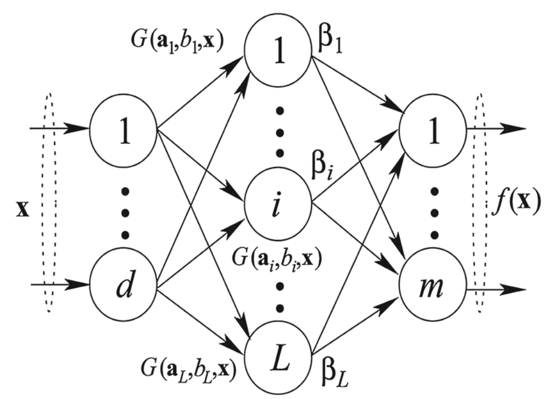


Fig. 5 Structure of ELM neural network

– Random feature mapping,
– Linear parameters solving,

In the first stage, ELM arbitrarily initializes the hidden layer to map the input data into a feature space (called ELM feature space) (Fig. 4) using some nonlinear mapping functions.

The standard ELM architecture is illustrated in (Fig. 5).

Generally, the output function of ELM for SLFNs with L hidden neuron is (1):

$$f_L(x) = \sum_{i=1}^{L} \beta_i h_i(x) = h(x)\,\beta \tag{1}$$

$\beta = [\beta_1 \cdots \beta_L]^T$ is the output weight vector connecting the hidden layer L nodes.

$h_i(x) = G(a_i, b_i, x)$ with $a_i \in R^d, b_i \in R$ is the real application and $G(a, x, b)$ the nonlinear function with hidden node parameters (a, b) which is randomly generated.

In the second stage, the training of the ELM consists of solving the linear system [31] formed by these nonlinearly transformed outputs of the hidden layer, and their corresponding target values by the calculation of the hidden-layer output randomized matrix H are defined as (2):

$$H = \begin{bmatrix} h(x_1) \\ \\ \\ h(x_N) \end{bmatrix} = \begin{bmatrix} h_1(x_1) & \ldots & h_L(x_1) \\ & \vdots & \\ h_1(x_N) & \ldots & h_L(x_N) \end{bmatrix} \quad (2)$$

and T is the training data target matrix:

$$T = \begin{bmatrix} t_1^T \\ \\ \\ t_N^T \end{bmatrix} = \begin{bmatrix} t_{11} & \ldots & t_{1m} \\ & \vdots & \\ t_{N1} & \ldots & t_{Nm} \end{bmatrix} \quad (3)$$

Thereafter we calculate:

The hidden-layer output connection weights β by solving the least squares problem $\beta = H^{*}T$ (4) where H^{*} is the generalized inverse matrix of the matrix H. The nonlinear mapping functions in ELM can be any nonlinear piecewise continuous functions. The regularized extreme learning machine algorithm used in this study is summarized in Table 3.

7 Prediction and classification of emotion via VAD dimensions

This part is devoted to the validation and the test of the proposed framework, two cases of emotion classification will be presented in details. First of all, we evaluated our system for the recognition of two levels of arousal, valence, and dominance, and then we tried to discriminate emotion in a normalized VA quadrant. We also showed the performance of extreme learning machine and the efficiency of level feature fusion applied to combine information derived from various sensors.

7.1 Experimental results

7.1.1 *Emotional arousal valence and dominance classification using ELM*

In this section, we have treated the prediction of emotional VAD as 1D regression problem. Two levels of arousal valence and dominance were defined. We used the entire segmented data from DEAP to evaluate our system. ELM classifier was tested with sigmoid mapping function (4):

$$G(a_i, b_i, x) = \frac{1}{1 + \exp^{(-(ax+b))}} \quad (4)$$

To build the physiological descriptors, we used the selected feature- based wavelet mentioned in Table 2. The number of evaluated samples was 1280 per sensor. All features used for the training of the R-ELM have been normalized to the range of [−1,1]. There are two hyper-parameters for ELM: the number of hidden nodes L and the regularization factor C.

Recent studies which applied ELM for multiclass classification [21] evaluated on various biomedical and big data sets showed that generally the C parameter can be chosen from the range: $\left[0, \ldots, 2^{25}\right]$.

First, we tested the ELM classifier with a random number of hidden neuron $L \in [100, \ldots, 1000]$. In the implementation of ELM, it is found that the generalization performance of ELM is not sensitive to the dimensionality of the feature space [32].

In our study, we separate two classes of emotion using a set of 20 physiological features, the number of multi output nodes ($m > 1$). For multiclass cases, the predicted class label of a given testing sample is the index number of the output node which has the highest output value for the given testing samples. Let $f_j(x)$ denote the output function of the jth output node. $f(x) = [f1(x), \ldots fm(x)]\, T$, the predicted class label of sample x is label $(x) = \text{argmax} f_j(x)$ with $j \in \{1, \ldots, m\}$ [21].

Table 3 The used extreme learning machine implementation

Algorithm
Input $\mathrm{N} = \left\{(x_i, t_i) \backslash x_i \in R^d, t_i \in R^m, i = 1, 2, \ldots.., \mathrm{N}\right\}$, the activation function g, the number of hidden nodes K and the regularization parameter C
Output : output weight matrix β .
Step 1: Randomly assign input weights $\mathrm{w_j}$, and the biases $\mathrm{b_j}$ with $j = 1, .., K$
Step 2: Calculate the hidden-layer output matrix H.
Calculate the output weight matrix β according to :
$\beta = \left(HH^T + \frac{I}{C}\right)^{-1} HT^T$ (5).
where C is a constant and I is the identity matrix.

Table 4 VAD recognition results obtained from different physiological modalities using R-ELM classifier

Physiological sensors	Valence accuracy (%)	Arousal accuracy (%)	Dominance accuracy (%)
EOG	70.31	66.40	65.18
EMG	64.84	66.40	67.40
RESP	72.65	67.18	66.66
PPG	64.06	63.28	63.70
TEMP	66.40	69.53	63.70
GSR	67.96	67.18	65.18

Table 4 shows the accuracy rates of emotional valence, arousal, and dominance classification obtained from all peripheral sensors using R-ELM classifier. All experiments here are performed with tenfold cross-validation.

For arousal recognition, the most relevant sensors are GSR and TEMP, and they achieved 68.70 and 69.53%, respectively.

For valence, we can clearly observe that EOG and RESP are the most affective modalities, and they achieved a high recognition rate of 72.65 and 70.31% compared to the rest of the sensors.

For dominance, we found that EMG, GSR, and RESP are the most effective modalities, and they achieved an acceptable recognition rate of 67.40, 65.18, and 66.66%, respectively, compared to the rest of the sensors. Figures 6 and 7 shows the parameters sensitivity analysis (L, C) for R-ELM on the recognition of emotional arousal, valence from RESP sensor, and demonstrated that increasing the value of L affects the recognition rate.

The best R-ELM performance was obtained when the number of hidden node L tends to 1000.

7.1.2 Fusion of physiological sensors

Classification in different modalities can be fused at both feature level and decision level. Feature-level fusion is obtained by concatenating all the features from multiple cues to form

Fig. 6 Parameters sensitivity analysis (L, C) of R-ELM using RESP features for arousal levels recognition

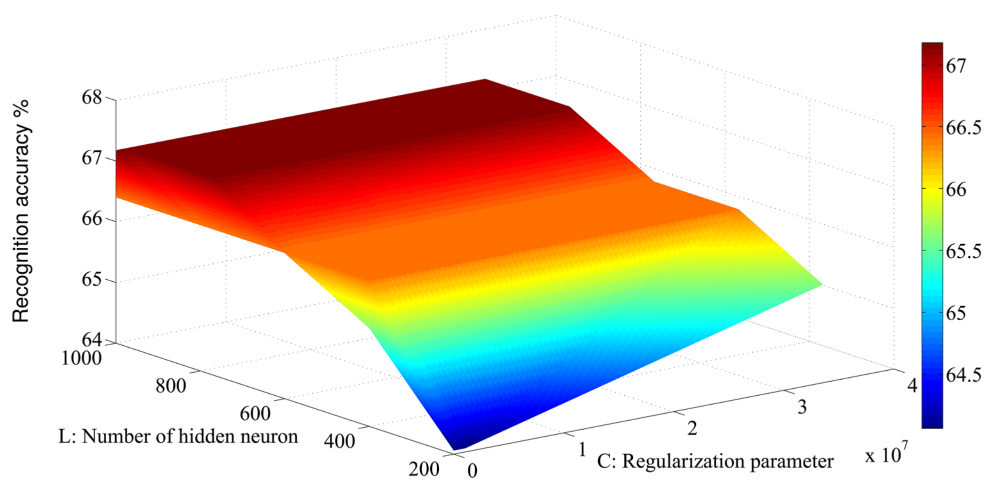
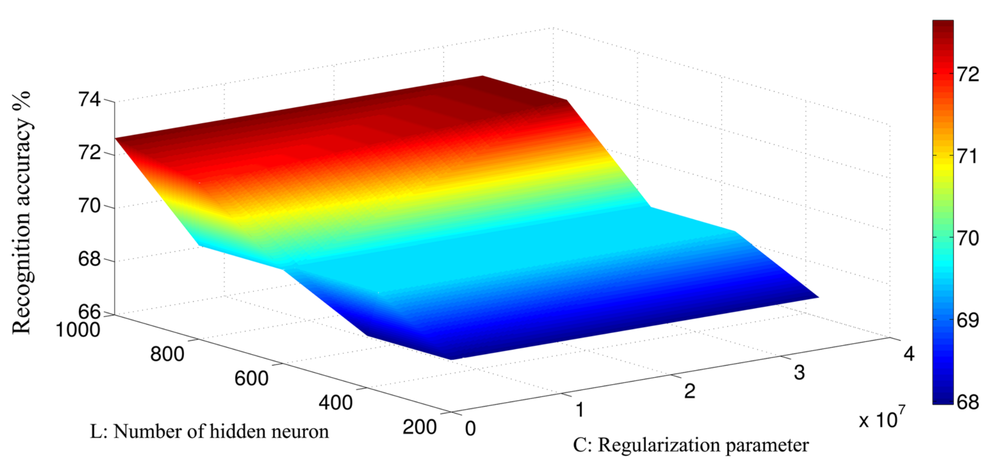


Fig. 7 Parameters sensitivity analysis (L C) of R-ELM using RESP features for valence levels recognition



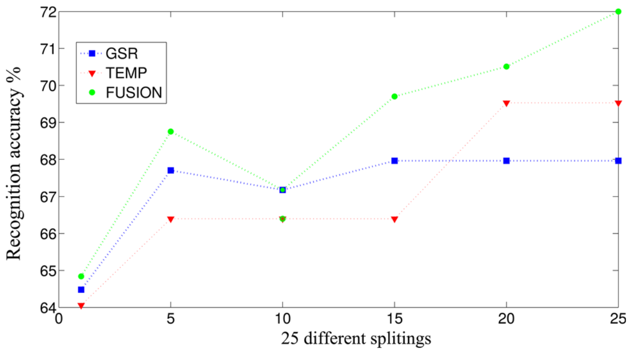


Fig. 8 Multimodal arousal classification results obtained using R-ELM

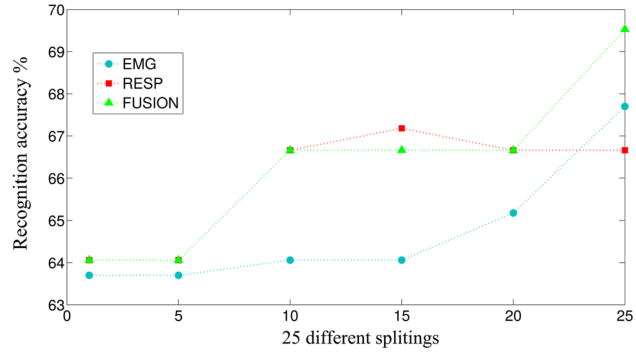


Fig. 10 Multimodal dominance classification results obtained using R-ELM

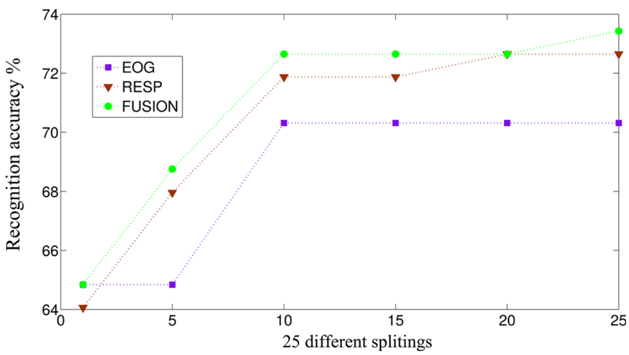


Fig. 9 Multimodal valence classification results obtained using R-ELM

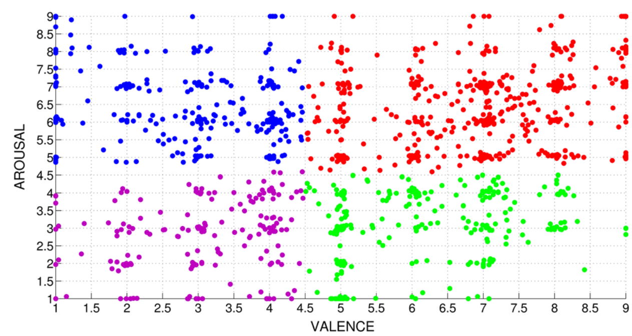


Fig. 11 Distribution of all tagged physiological instances in the normalized VA quadrant

one feature vector which is then fed into a machine learning technique. Generally, the feature-level fusion technique is used to improve the single modality results. In our case, the most perfectionist sensors previously determined have to be fused. In order to enhance the global rates of emotional VAD prediction, a simple concatenation of features extracted from different physiological modalities was applied. We concatenate directly the selected features extracted from EOG and RESP to form a single representation for each level of valence, without applying a feature reduction method. The recognition accuracy has been improved to achieve a rate of 73.43% (Fig. 8).

This procedure is repeated 25 times to give a better estimation of VAD levels, Figs. 8, 9, and 10 show the obtained recognition results.

For arousal, a set of features extracted from GSR and TEMP were fused, and the best recognition accuracy obtained was 72%.

The prediction of valence was also improved to 73.43% by combining EOG and GSR features.

For a better estimation and prediction of dominance levels, we fused physiological features extracted from EMG and RESP which are mainly the most informative sensors, and the best recognition accuracy obtained was 69.53%

7.1.3 Affect classification in VA quadrant

In order to predict four classes of emotions, we segment the four quadrants of the valence arousal (VA) space according to the ratings LALV, HALV, LAHV, and HAHV denoting low arousal/low valence, high arousal/low valence, low arousal/high valence, and high arousal/high valence, respectively. The two axes of arousal and valence were divided, and the median value of the annotation scale is considered as the landmark of the arousal valence plane, the rating distribution of DEAP in four areas of the (VA plane) is shown in Fig. 11.

In this section, we tried to recognize emotion in the VA quadrant using the new annotation of four classes. The same conditions were applied in the training of our system. As in the first experiment, the same number of hidden neurons was used, and the regularization parameter C was fixed to 2^{20}.

Figure 12 shows the accuracy rates obtained using all physiological signals, and EOG, GSR and RESP are the most informative modalities for the classification of emotion in VA quadrant, they provided better performance than the rest of signals. The higher recognition accuracy achieved from EOG signal was 52.34%, for the RESP signal, and the best accuracy rate was 46.87%. We obtained 45.31% using GSR features.

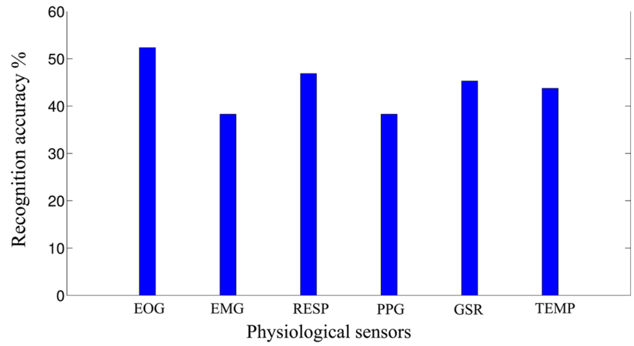


Fig. 12 Emotion classification results in VA quadrant using R-ELM classifier

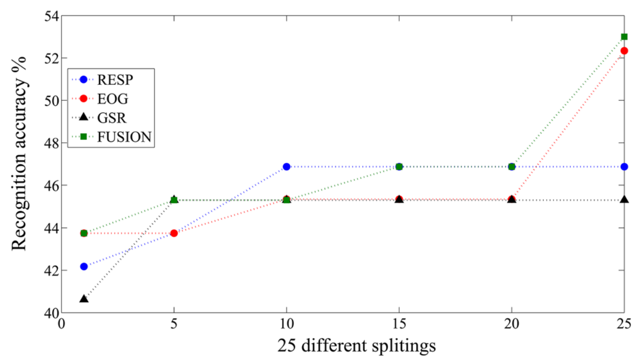


Fig. 13 Multimodal emotion classification results in VA quadrant using R-ELM

The classification rates are low. In order to increase the mean recognition accuracy obtained when we used a single physiological modality, we applied the FLF (feature-level fusion) technique. We concatenate directly features extracted from the physiological signal that archived the best classification rates. A one feature vector was formed and used for the new trained model. As can be seen from the results shown in Fig. 13, we denote an improvement after applying the FLF. The best recognition accuracy obtained was 53% obtained from the fusion of informations derived from three biosensors EOG, GSR, and RESP.

8 Discussion

The objective of this work is to procure a novel framework for a better estimation and prediction of emotional VAD classes using the original input features extracted from peripheral physiological sensors. The entire DEAP database was used to evaluate our proposed algorithm. In our approach, we used a combined modality fusion applied to physiological signals of the 32 participant in DEAP. We started by comparing our work to Koelstra et al. [26] that have developed the DEAP data set. They obtained an average accuracy of 62.0 and 57.6% for valence and arousal (2 classes), respectively. Another study based on physiological signals in DEAP is that of Godin et al. [27] who tried to find the most relevant physiological sensors for a better recognition of valence and arousal (2 classes), and their optimized selected features space was given to a naive Bayesian classifier, and the obtained accuracy rate was 63% for valence and 59% for arousal.

8.1 Comparison with the state-of-the-art

The classification rates obtained by our system 73.43% for valence and 72.65% for arousal (2 classes) are greater than those obtained by the both cited methods, and it confirms the effectiveness of the proposed multimodal framework-based ELM. The direct comparison of our work to some developed frameworks based on EEG features is relatively difficult because of the nature of EEG signal which differs from the used physiological signals, and it contains more frequency bands and variations which makes its analysis more informative for the capture of emotional responses. Previous studies using DEAP are shown in Table 5 and compared to our approach.

EEG signal in DEAP was recorded via 32 channels; in fact, in our study, we used data derived from only 8 channels. But the rates obtained by our method 73.43 and 72.65% of two levels for (valence and arousal) are still better than those obtained by Chung et al. [34] that used the Bayes classifier with only 10 participant from DEAP dataset. Their accuracy for valence and arousal classification was 66.6 and 66.4% for two classes and 53.4 and 51.0% for three classes. Zhang et al. [35] described an ontological system based on EEG, and their model reached an average recognition ratio of 75.19% on valence and 81.74% on arousal in eight participants. Although their accuracy was relatively high, the categories of each dimension are only two, and these results were achieved with a subset of the original dataset. The obtained rate of 52.34% for the prediction of 4 V–A classes is also better than the one obtained by Soleymani et al. [28] when they studied three levels of arousal and valence using physiological signals from MAHNOB, and still comparable to Zheng et al. [36] method-based EEG features that achieves an average accuracy of 69.67% on the same data set for four classes in VA quadrant with PSD features from theta frequency band for all 32 participants using GELM classifier.

Table 5 Accuracy comparison with the state of the art's for DEAP database

Author (s)	Year	Approach	Classification accuracy
Koelstra et al. [26]	2012	Power spectral features	57.6% (V) 62.0% (A) for 2 levels
Chung et al. [34]	2012	Bayesian classifier with 10 selected participants	66.6% (V) 66.4% (A) for 2 levels 53.4% (V) 51% (A) for 3 levels
Zhang et al. [35]	2013	Power spectral features and SVM with 8 selected participants	75.19% (V) 81.74% (A) for 2 levels
Godin and Campagne [26]	2015	Statistical Power spectral features Fisher scores and Bayesian classifier with all 32 participants	63% (V) 57% (A) for 2 levels
Zheng Zhu and Lu [36]	2016	PSD features and GELM classifier with all 32 participants	69.67% in VA quadrant
Our approach	2017	Feature-based wavelet and R-ELM classifier with all 32 participants	73.43% (V) 72.65% (A) 69.3% (D) for 2 levels 53% in VA quadrant

9 Conclusion

In this paper, we have evaluated the performance of a computational framework developed for a better estimation and recognition of emotional responses. DEAP dataset was structured, and physiological signals of the 32 participants and their annotation files were segmented and normalized; a feature extraction step was applied, and extreme learning machine was adopted as classification algorithm. For the classification of two arousal and valence levels, significant results are obtained from the two biosensors EOG and REPS (70 and 72%) for valence and (68 and 69%) for arousal from GSR and TEMP. A feature-level fusion approach was applied to improve these single modality results. The best recognition accuracy attained after fusion was 73.45% for valence and 72.65% for arousal. In the case of four classes (HAHV, LALV, LAHV, and HALV) the best average classification accuracy was 53%. From the experimental results, it can be observed that ELM is an efficient machine learning method, and it may improve the prediction of emotional VAD classes. The comparative classification accuracy achieved shows the superior performance of our employed machine learning method in comparison with some existing approaches that used other classification algorithms such as SVM and naive Bayes classifier. We have shown that generally the performance of an emotion recognition framework can be improved in most cases using multimodal fusion technique like FLF, and it depends evidently on the efficiency of the used machine learning technique. It is advisable to well structure a given physiological dataset to ensure a rigorous analysis of humans affect. Moreover, we can conclude that valence classes can be predicted better than arousal. In the future work, it will be very interesting to include other modalities like facial features which can be extracted from face video frames in order to form a visual descriptor and to exploit the complementary relation between physiological human responses and facials signs for a better prediction and estimation of emotional VAD.

Acknowledgements The authors want to thank all iBUG group members and the DEAP dataset administrator for the sharing of all database files for free.

References

1. Picard RW, Vyzas E, Healey J (2001) Machine emotional intelligence: analysis of affective physiological state. IEEE Trans Pattern Anal Mach Intell 23:1175–1191

2. Zeng Z, Pantic M, Roisman GI, Huang TS (2009) Survey of affect recognition methods audio, visual, and spontaneous expressions. IEEE Trans Pattern Anal Mach Intell 31:39–58
3. International Affective Picture System, IAPS (1997)
4. Kim J, Andre E (2008) Emotion recognition based on physiological changes in music listening. IEEE Trans Pattern Anal Mach Intell 30:2067–2083
5. Schuller B (2011) Recognizing affect from linguistic information in 3D continuous space. IEEE Trans Affect Comput 2(4):192–205
6. Soleymani M, Chanel G, Kierkels JJM, Pun T (2009) Affective characterization of movie scenes based on content analysis and physiological changes. Int J Semant Comput 30(2):235–254
7. Rottenberg J, Ray RD, Gross JJ (1976) Emotion elicitation using films series in affective science. Oxford University, Oxford, pp 9–28
8. Mauss IB, Robinson MD (2009) Measures of emotion: a review. Cognit Emot 23(2):209–237
9. Healey JM, Picard RW (2005) Detecting stress during real-world driving tasks using physiological sensors. IEEE Trans Intell Transp Syst 6(2):156–166
10. Chanel G, Rbetez C, Betancourt M, Pun T (2008) Boredom, engagement and anxiety as indicators for adaptation to difficulty in games. The 12th international conference on Entertainment and media in the ubiquitous era, Finland, pp 13–17
11. Ekman P (1982) Emotions in the human faces. 2nd ed studies in emotion and social interaction. Cambridge University Press, Cambridge
12. Russell JA (1982) A circumplex model of affect. J Personal Soc Psychol 39:1161–1178
13. Jerritta S, Murugappan M, Nagarajan R, Kahrunizam W (2011) Signals based human emotion recognition: a review. 7th international colloquium on signal processing and its application
14. Rani P, Liu C, Sarkar N, Vanman E (2006) An empirical study of machine learning techniques for affect recognition in Human–robot interaction. Pattern Anal Appl 9:58–69
15. Long Z, Liu G, Dai X (2010) Extracting emotional features from ECG by using wavelet transform, International conference on biomedical engineering and computer sciences (ICBECS), Wuhan, pp 1–4
16. Cong Z, Chetouani M (2010) Hilbert–Huang transform based physiological signals analysis for emotion recognition, International symposium on in signal processing and information technology (ISSPIT), pp 1–7
17. Schlkopf B, Smola AJ (2001) Learning with Kernels: support vector machines, regularization, optimization, and beyond (adaptive computation and machine learning). 1st edition
18. Cortes C, Vapnik V (1995) Support vector networks. Mach Learn 20(3):273–297
19. Shi L, Lu BL (2013) EEG-based vigilance estimation using extreme learning machines. Neurocomputing 102:135–143
20. He B, Xu D, Nian R, van Heeswijk M, Yu Q, Miche Y, Amaury L (2014) Fast face recognition via sparse coding and extreme learning machine. Cognit Comput 6:264–277
21. Bazi Y, Alajlan N, Melgani F, AlHichri H, Malek S, Yager RR (2014) Differential evolution extreme learning machine for the classification of hyperspectral images. IEEE Geosci Remote Sens Lett 6:1066–1070
22. Huang G-B, Zhou H, Ding X, Zhang R (2012) Extreme learning machine for regression and multiclass classification. IEEE Trans Syst Man Cybern 42:513–529
23. Chanel G, Ansari-Asl K, Pun T (2007) Valence–arousal evaluation using physiological signals in an emotion recall paradigm systems. ISIC. IEEE international conference on man and cybernetics, pp 2662–2667
24. Koelstra S, Yazdani A, Soleymani M, Muhl C, Lee JS, Nijholt A, Pun T, Ebrahimi T, Patras I (2010) Single trial classification of EEG and peripheral physiological signals for recognition of emotions induced by music videos. In Brain informatics, ser. Lecture Notes in Computer Science 6334(9):89–100
25. Soleymani M, Pantic M, Pun T (2012) Multimodal emotion recognition in response to videos. IEEE Trans Affect Comput 3:211–223
26. Koelstra S, Muhl C, Soleymani M, Lee JS, Yazdani A, Ebrahimi T, Pun T, Nijholt A, Patras I (2012) DEAP: a database for emotion analysis using physiological signals. Trans Affect Comput 3(1):18–31
27. Godin CH, Compagne A (2015) Selection of the best physiological features for classifying emotion. The 3th International conference on physiological computing systems
28. Soleymani M, Lichtinauer J, Pantic M, Pun T (2012) A multimodal database for affect recognition and implicit tagging. IEEE Trans Affect Comput 3(1):42–55
29. Russell JA, Mehrabian A (1977) Evidence for a three-factor theory of emotions. J Res Personal 11(3):273–294
30. Verma GK, Tiwary US (2014) Multimodal fusion frame-work: A multiresolution approach for classification and recognition from physiological signals. Neuroimage 102:162–172
31. Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1–3):489–501
32. Huanga G, Huang GB, Songa S, Youa K (2014) Trends in extreme learning machines: a review. Neural Netw 61:32–48
33. Huang GB, Wang DH, Lan Y (2011) Extreme learning machines: a survey. Int J Mach Learn Cybern 2(2):107–122
34. Chung SY, Yoon HJ (2012) Affective classification using Bayesian classifier and supervised learning. IEEE 12th international conference on control, automation and systems (ICCAS)
35. Zhang X, Hu B, Chen J, Moore P (2013) Ontology-based context modeling for emotion recognition in an intelligent web. World Wide Web 16(4):497–513
36. Zheng W, Zhu JY, Lu BL (2016) Identifying stable patterns over time for emotion recognition from EEG. In Human–Computer Interaction. arXiv:1601.02197
37. Filix A, Daniela H, Puica M (2015) Neural network approaches for children's emotion recognition in intelligent learning applications. 7th international conference on education and new learning technologies (1) doi:10.13140/RG.2.1.3413.2969
38. Muthusamy H, Polat K, Yaacob S (2015) Improved emotion recognition using Gaussian mixture model and extreme learning machine in speech and glottal signals. Math Probl Eng. doi:10.1155/2015/394083
39. Kaya H, Salah AA (2016) Combining modality-specific extreme learning machines for emotion recognition in the wild. J Multimodal User Interfaces 2(10):139–149