

Subject independent emotion recognition using EEG and physiological signals – a comparative study

Emotion
recognition –
EEG and
physiological

Manju Priya Arthanarisamy Ramaswamy and Suja Palaniswamy
*Department of Computer Science and Engineering, Amrita School of Computing,
Amrita Vishwa Vidyapeetham, Bengaluru, India*

Received 27 March 2022
Revised 11 June 2022
1 August 2022
Accepted 25 August 2022

Abstract

Purpose – The aim of this study is to investigate subject independent emotion recognition capabilities of EEG and peripheral physiological signals namely: electrooculogram (EOG), electromyography (EMG), electrodermal activity (EDA), temperature, plethysmograph and respiration. The experiments are conducted on both modalities independently and in combination. This study arranges the physiological signals in order based on the prediction accuracy obtained on test data using time and frequency domain features.

Design/methodology/approach – DEAP dataset is used in this experiment. Time and frequency domain features of EEG and physiological signals are extracted, followed by correlation-based feature selection. Classifiers namely – Naïve Bayes, logistic regression, linear discriminant analysis, quadratic discriminant analysis, logit boost and stacking are trained on the selected features. Based on the performance of the classifiers on the test set, the best modality for each dimension of emotion is identified.

Findings – The experimental results with EEG as one modality and all physiological signals as another modality indicate that EEG signals are better at arousal prediction compared to physiological signals by 7.18%, while physiological signals are better at valence prediction compared to EEG signals by 3.51%. The valence prediction accuracy of EOG is superior to zygomaticus electromyography (zEMG) and EDA by 1.75% at the cost of higher number of electrodes. This paper concludes that valence can be measured from the eyes (EOG) while arousal can be measured from the changes in blood volume (plethysmograph). The sorted order of physiological signals based on arousal prediction accuracy is plethysmograph, EOG (hEOG + vEOG), vEOG, hEOG, zEMG, tEMG, temperature, EMG (tEMG + zEMG), respiration, EDA, while based on valence prediction accuracy the sorted order is EOG (hEOG + vEOG), EDA, zEMG, hEOG, respiration, tEMG, vEOG, EMG (tEMG + zEMG), temperature and plethysmograph.

Originality/value – Many of the emotion recognition studies in literature are subject dependent and the limited subject independent emotion recognition studies in the literature report an average of leave one subject out (LOSO) validation result as accuracy. The work reported in this paper sets the baseline for subject independent emotion recognition using DEAP dataset by clearly specifying the subjects used in training and test set. In addition, this work specifies the cut-off score used to classify the scale as low or high in arousal and valence dimensions. Generally, statistical features are used for emotion recognition using physiological signals as a modality, whereas in this work, time and frequency domain features of physiological signals and EEG are used. This paper concludes that valence can be identified from EOG while arousal can be predicted from plethysmograph.

Keywords EEG, EOG, EMG, EDA, Temperature, Plethysmograph, Respiration, Emotion recognition, Arousal, Valence, Multimodal emotion recognition, Subject independent

Paper type Research paper

© Manju Priya Arthanarisamy Ramaswamy and Suja Palaniswamy. Published in *Applied Computing and Informatics*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

The authors of the paper thank Dr. Karthik R.M. for providing his valuable feedback and suggestions on the manuscript.



Applied Computing and
Informatics
Emerald Publishing Limited
e-ISSN: 2210-8327
p-ISSN: 2634-1964
DOI 10.1108/ACI-03-2022-0080

1. Introduction

Subject independent emotion recognition using single or multiple modalities is a burgeoning area of research in affective computing. Emotion recognition (ER) plays a vital role in human computer interaction (HCI) as it tries to make HCI, similar to human–human interaction (HHI) by incorporating ER and emotion expression capabilities in machines. The distinguishing feature between HCI and HHI is the ER and emotion expression capabilities of humans.

Humans recognize others' emotions via facial expression and contextual information in day-to-day life. Emotions serve as evolved communication and hence should evoke behaviors that reveal the subjects' emotional state to others [1]. The emotional state of a person can be inferred from behavior in face, voice, whole-body and observer ratings. James's emotion theory [2] states that emotional response can be measured using peripheral physiological signals. Some of the peripheral physiological signals used in ER are electrodermal activity (EDA), cardiovascular activity and respiration activity. Cannon's emotion theory [3] suggests that emotions are derived from subcortical centers, and this led to the study of emotional responses of central nervous system (CNS) signals using EEG, neuroimaging techniques and electrooculogram (EOG).

Subject dependent unimodal and multimodal ER provides considerable accuracy, while subject independent ER needs improvement. One aspect that hinders baseline of subject independent ER models is the non-availability of subject independent test sets for the publicly available multimodal ER datasets. Many of the subject independent ER studies in the literature provide an average of leave one subject out (LOSO) validation score as final accuracy. In this work, the test subjects used for validation of the model are specified explicitly so that any future work can use these model scores as a baseline.

Subject independent ER capabilities of time and frequency domain features of EEG and peripheral physiological signals, namely EOG, EMG, EDA, temperature, plethysmograph and respiration both independently and in combination on the DEAP dataset in arousal and valence dimensions using classifiers - Naïve Bayes, logistic regression, linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), logit boost and stacking are explored in this research. Through this research work, it is found that, from an ergonomic perspective, valence can be measured from the eyes while arousal can be measured from changes in blood volume. The model scores of this work can be used as a baseline for future work as this work reports the results on a truly subject independent test set.

2. Related works

The recent advances in multimodal ER are in the areas of feature extraction, feature selection, modeling and fusion strategies. Multimodal ER involves three important aspects: extracting shared representations from multiple modalities, removing redundant features and learning key features from each modality. To address all three aspects, multimodal deep belief network (MDBN) was investigated [4]. Recent studies in the literature used global average pooling [5], deep belief network (DBN) [6] and multi-hypergraph neural network [7] to investigate the aspect of correlation among features in multimodal ER. The optimal combination of features play a significant role in multimodal ER and was studied using multi-kernel learning approach [8] and deep learning based hierarchical feature fusion approach [9]. Recent studies have explored the significance of features in ER [10].

The body of work in literature has explored the feature extraction ability of deep learning networks for end-to-end ER architectures and its performance was determined by the strength of the input signals [11]. Deep learning architectures like ensemble convolution neural network (ECNN) [5], DBN [6], inception ResNet v2 [12], spiking neural networks (SNN) [13], autoencoder [14], hierarchy modular neural network (HMNN) [15], MDBN [16], transfer learning [17], transformer-based architecture using CNN [17] and high resolution network (HRNet) [18] were explored for ER.

Decision level fusion versus feature level fusion is a long-standing contention in the field of multimodal ER. The decision level fusion improves the accuracy by 5% in comparison to unimodal accuracy [19], whereas the feature level fusion provides ER accuracy comparable to decision level fusion with less computation time [10]. Some of the literature [10, 12, 19] reported LOSO validation score as final accuracy which is a limitation in subject independent ER.

The work reported in this paper sets the baseline for subject independent ER using DEAP dataset by clearly specifying the subjects used in training and test set. In addition, this work specifies the cut-off score used to classify the scale as low or high in arousal and valence dimensions. Generally, statistical features are used for ER using physiological signals as a modality, whereas in this work, time and frequency domain features of physiological signals and EEG are used. The experiment is conducted on both modalities independently and in combination. This work arranges the physiological signals in order based on the prediction accuracy obtained on test data using time and frequency domain features.

3. Materials and methods

DEAP dataset is used to compare the prediction ability of time and frequency domain features of EEG and physiological signals over a similar set of classifiers and to sort the physiological signals. In this experiment, two ensemble classifiers – logit boost and stacking and two statistical classifiers – Naïve Bayes and QDA are used. All four classifiers are used independently and in combination of EEG and physiological signals. The feature selection and training of classifiers are performed using Weka software [20]. The proposed methods for arousal and valence prediction in multimodal and unimodal environments are shown in Figure 1.

3.1 DEAP dataset description

DEAP [21] dataset has EEG and peripheral physiological signal recordings of 32 participants (16 for each gender). The signals were recorded when the participants watched music video of length one minute. Each participant watched a subset of 40 music videos and rated the valence, arousal, dominance and liking of each video. For each trial, 32 channels of EEG signals and 12 channels of peripheral signals were recorded using Biosemi active two system at 512 Hz.

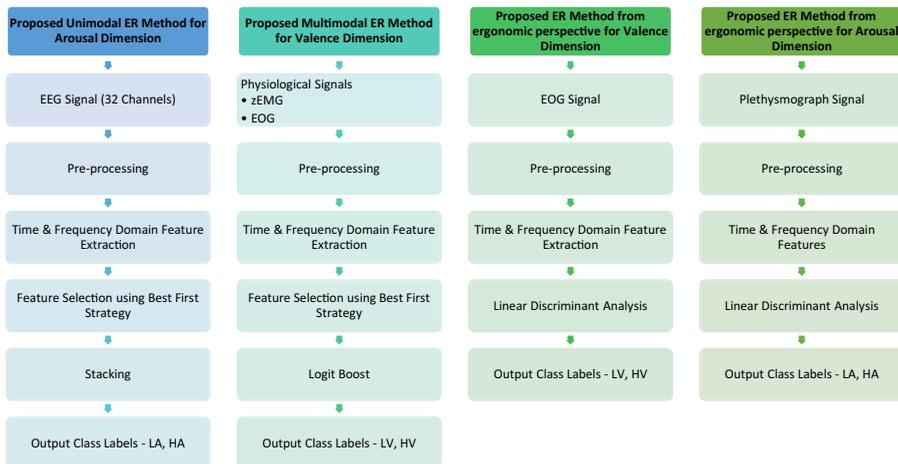


Figure 1. Proposed multimodal and unimodal emotion recognition methods for arousal dimension and valence dimension

3.2 Evaluation measures

The evaluation metrics used to compare different models in this experiment are accuracy and F1-score. Additional metrics, namely ROC area and kappa statistic are reported for the proposed methods.

3.2.1 Accuracy. Accuracy is the measure of correctly classified instances. The accuracy percentage ranges from 0 to 100, where 100 is the best possible accuracy and is shown in [equation \(1\)](#).

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + False\ Positive + True\ Negative + False\ Negative} \quad (1)$$

3.2.2 F1-score. F1-score is the harmonic mean of precision and recall. The range of F1-score varies from 0 to 1, where 0 is the worst possible score and 1 is the best possible score and is shown in [equation \(2\)](#). F1-score gives better measure of incorrectly classified instances.

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (2)$$

3.2.3 ROC area. The area under the ROC curve measures the ability of the binary classifier to distinguish between classes. The value ranges from 0 to 1, where 1 implies the classifier is able to perfectly distinguish between the classes.

3.2.4 Cohen's kappa. Cohen's kappa values range from -1 to 1 , where 1 implies the model is good. A kappa value of 0 indicates that the model is as good as a chance classifier.

3.3 Training and test dataset split

The dataset is split into subject independent training and test sets in the ratio of 70:30. The data of 22 participants is used as training set, while the remaining data of 10 participants is used as test set. Subjects – s02, s04, s05, s09, s15, s20, s23, s28, s29 and s30 are used in the test set while the rest of the subjects are used in the training set.

3.4 Labeling strategy

As the objective is to train classifiers using supervised learning, the continuous scale ratings of valence and arousal are converted into labels by splitting the continuous scale. The scale range $[0, 5.0]$ is considered as low, while $(5.0, 9.0]$ is considered as high. The scale value of 5.0 is chosen as the split point as the mean value of the ratings lie approximately around 5.0 . The labeling strategy and distribution of labels across the training and test sets are shown in [Table 1](#).

3.5 Pre-processing

The DEAP dataset provides the pre-processed data and is explained in this sub-section. The EEG signals were down-sampled to 128 Hz and EOG artefacts were removed. Bandpass filter with frequency range of 4.0 to 45.0 Hz was applied. The EEG data were averaged to common

Table 1.
Labeling strategy and distribution of labels across training and test set

Attribute	Scale range	Label	Training set [Count (%)]	Test set [Count (%)]
Valence	≥ 0 and ≤ 5	LV	398 (45.23%)	174 (43.50%)
Valence	> 5 and ≤ 9	HV	482 (54.77%)	226 (56.50%)
Arousal	≥ 0 and ≤ 5	LA	348 (39.55%)	195 (48.75%)
Arousal	> 5 and ≤ 9	HA	542 (60.45%)	205 (51.25%)

reference and pre-trial baseline was removed. The physiological signals were down-sampled to 128 Hz and the pre-trial baseline was removed.

3.6 Feature extraction

Time domain and frequency domain features were used to find the electrode position for top-30 features and it was found that frequency-based power spectrum density provided better accuracy [22]. In contrast, another study found that power spectral density did not perform well [23]. As emotions vary with time, Hjorth features were widely used in ER as these features are useful in monitoring time varying EEG signals [24]. Hence, this work extracted both time domain features – Hjorth activity, Hjorth complexity and frequency domain feature – power spectral density (PSD) and used feature selection process to select the best performing features [25]. Hjorth activity and Hjorth complexity features are computed for the entire time range of the signal. A total of 280 features are computed as shown in Table 2.

Horizontal EOG, vertical EOG, zygomaticus major EMG and trapezius EMG are computed by subtracting the corresponding values between two channels and is shown in Table 2. In this work, EEG is considered as one modality, and all other signals are grouped under physiological signals. For the feature extraction process, $y(t)$ is considered as the signal and $\frac{dy(t)}{dt}$ as the first derivative of the signal.

3.6.1 Hjorth activity. Hjorth activity [24] parameter is the total power of the signal. It is the surface of the power spectrum in the frequency domain and is shown in equation (3).

$$Activity = Variance(y(t)) \quad (3)$$

3.6.2 Hjorth complexity. Hjorth complexity [24] is a dimensionless parameter defined as the ratio of mobility of the first derivative of the signal to the mobility of the signal, as shown in equation (4). The mobility is defined as the square root of the ratio of the variance of the first derivative of the signal to the variance of the signal, as shown in equation (5). The mobility of the signal represents the frequency variance of the power spectrum and can be illustrated as the standard deviation of the power spectrum along the frequency axis. The Hjorth complexity gives an estimate of the bandwidth of the signal and indicates the shape similarity of the signal to a pure sine wave.

$$Complexity = \frac{Mobility\left(\frac{dy(t)}{dt}\right)}{Mobility(y(t))} \quad (4)$$

where mobility is defined as in equation (5).

Modality	Channel name	Features	Number of features
EEG	FP1, AF3, F3, F7, FC5, FC1, C3, T7, CP5, CP1, P3, P7, PO3, O1, Oz, Pz, FP2, AF4, Fz, F4, F8, FC6, FC2, Cz, C4, T8, CP6, CP2, P4, P8, PO4, O2	Hjorth Activity, Hjorth Complexity, Alpha Band Welch PSD, Beta Band Welch PSD, Gamma Band Welch PSD, Theta Band Welch PSD, Delta Band Welch PSD	$32 \times 7 = 224$
EOG	EXG1 – EXG2, EXG3 – EXG4		$2 \times 7 = 14$
EMG – Zygomaticus Major, Trapezius	EXG5 – EXG6, EXG7 – EXG8		$2 \times 7 = 14$
EDA	GSR1		$1 \times 7 = 7$
Respiration	Resp		$1 \times 7 = 7$
Plethysmograph	Plet		$1 \times 7 = 7$
Temperature	Temp		$1 \times 7 = 7$

Table 2.
Features for EEG
channels and
physiological channels

$$Mobility = \sqrt{\frac{\text{Variance}\left(\frac{dy(t)}{dt}\right)}{\text{Variance}(y(t))}} \quad (5)$$

3.6.3 Power spectral density. PSD refers to the spectral energy distribution of the signal per unit time [26] and is computed separately for alpha (8 – 12 Hz), beta (12 – 30 Hz), gamma (30 – 45 Hz), theta (4 – 8 Hz) and delta band (0 – 4 Hz) of each channel using the Welch method.

3.7 Feature selection

In this work, feature selection is done using best first search strategy which is a correlation-based feature subset selection [27] where the correlation between the feature and the output class, and the correlation among the features is computed. Feature selection is done such that the subset of features are highly correlated with the class while intercorrelation among the selected features is low. In this work, best first strategy is carried out with an initial empty feature list followed by iteratively including and excluding all possible single attributes. In best first search strategy, single features that have high correlation with the class are added to the search space. If the added feature does not contribute to the improvement of accuracy, then the algorithm backtracks to the last best subset in the feature space and continues the search. In order to avoid exploring the entire feature space a stopping criterion is used. In this work, the search procedure is terminated if there is no improvement for the last five iterations.

The features selected for each of EEG, physiological, and combined modalities are shown in Table 3. From the features selected, it is observed that only frequency-based PSD features are selected for physiological signals while time based Hjorth features are selected for T7, P7, Fz, FP1, FC6 electrodes of EEG signal. The position of T7, P7, Fz, FP1 and FC6 electrode is associated with superior temporal gyrus, lateral occipital cortex, superior frontal gyrus, frontal pole and precentral gyrus, respectively [28]. From the feature selection process, it is observed that the time domain features of electrodes associated with gyrus and frontal pole

Modality	Features selected
EEG	Arousal – LA, HA GAMMA_F3_Welch_PSD, T7_HjorthActivity, GAMMA_CP5_Welch_PSD, ALPHA_P3_Welch_PSD, P7_HjorthComplexity, Fz_HjorthComplexity, ALPHA_FC6_Welch_PSD, GAMMA_CP6_Welch_PSD, GAMMA_P4_Welch_PSD Valence – LV, HV FP1_HjorthActivity, FC6_HjorthComplexity, ALPHA_FC2_Welch_PSD
Physiological	GAMMA_hEOG_Welch_PSD, GAMMA_zEMG_Welch_PSD
EEG + Physiological	GAMMA_F3_Welch_PSD, T7_HjorthActivity, GAMMA_CP5_Welch_PSD, ALPHA_P3_Welch_PSD, P7_HjorthComplexity, Fz_HjorthComplexity, ALPHA_FC6_Welch_PSD, GAMMA_CP6_Welch_PSD, GAMMA_P4_Welch_PSD, GAMMA_zEMG_Welch_PSD FC6_HjorthComplexity, GAMMA_zEMG_Welch_PSD

Table 3.
Features selected from
each of modalities

brain regions are selected [28]. This is in accordance with literature which states that the gyrus [29] and the frontal pole [30] have a role in emotion regulation.

3.8 Classifiers

In this work, the features from different modalities are fed to the supervised machine learning algorithms, namely, Naïve Bayes [31], logistic regression [32], LDA [33], QDA [33], logit boost [34], and stacking [35] and the accuracy determined is compared. These classifiers are explained briefly in supplementary material at <https://github.com/armanjupriya-er/er-comparison-supplementary>.

4. Results and discussion

The accuracy and F1-score of the experiment are shown in Table 4. The graphic illustration is available as Figure S1 at: <https://github.com/armanjupriya-er/er-comparison-supplementary>. The results obtained using EEG and physiological signals as independent modalities indicate that EEG signals are better at arousal prediction compared to physiological signals by 7.18%, while physiological signals are better at valence prediction compared to EEG signals by 3.51%. Combining EEG and physiological modalities, the arousal prediction is better than physiological signal modality by 2.39% and is inferior to EEG modality by 4.46%, while the valence prediction of the combined modality is better than EEG modality by 3.07% and is inferior to physiological modality by 0.42%. From the prediction accuracy in arousal and valence dimension, it is observed that EEG as a single modality and physiological signal as a single modality performs better than combining EEG with physiological signals.

A one-way ANOVA test was conducted in order to validate whether there is any significant difference in prediction ability between the EEG and physiological modalities using the same set of features. One-way ANOVA for arousal accuracy ($F(1,6) = 7.05$, $p = 0.0378$) shows that there is a significant difference in accuracy levels reported by EEG and physiological signals, while the difference in F1-Score ($F(1,6) = 5.07$, $p = 0.0653$) is not significant at 5% level of significance. One-way ANOVA for valence accuracy ($F(1,6) = 0.08$, $p = 0.7874$) and F1-score ($F(1,6) = 0.25$, $p = 0.6372$) shows that there is no significant difference in accuracy and F1-score between EEG and physiological modalities at 5% level of significance.

Feature selection shows that EEG electrodes F3, T7, CP5, P3, P7, Fz, FC6, CP6, P4 are used in arousal prediction, while signals from FP1 and FC6 are used in valence prediction. Based

Modality	Classifier	Arousal – LA, HA		Valence – LV, HV	
		Accuracy (%)	F1 score	Accuracy (%)	F1 score
EEG	Logit Boost	53.75	0.536	55.00	0.504
	Naïve Bayes	51.75	0.468	56.75	0.513
	QDA	51.75	0.470	57.00	0.524
	Stacking	56.00	0.560	56.75	0.538
Physiological	Logit Boost	49.50	0.415	59.00	0.592
	Naïve Bayes	49.00	0.395	55.00	0.405
	QDA	48.75	0.394	55.00	0.401
	Stacking	52.25	0.508	57.75	0.576
EEG + Physiological	Logit Boost	46.75	0.451	58.75	0.588
	Naïve Bayes	49.50	0.446	56.50	0.566
	QDA	50.25	0.464	55.25	0.402
	Stacking	53.50	0.534	56.50	0.566

Table 4.
Accuracy and F1-score
for EEG, physiological
and
EEG + physiological
modality

on the observation, FC6 electrode is common for arousal and valence prediction; therefore prediction ability of the FC6 electrode is studied and shown in Table 5. The experimental results suggest that the ability of the FC6 electrode to predict the valence is 59.00%, which is equal to the best prediction accuracy, obtained using all of the physiological signals. The prediction accuracy of the FC6 electrode with respect to arousal is 52.25%, which is at par with the prediction accuracy of the physiological signals. The FC6 electrode position corresponds to the primary motor cortex area in the brain, which is associated with the function of controlling different muscle groups [36]. This concludes that the prediction accuracy of the FC6 electrode in the valence dimension comes from muscle activity.

On further analysis of the selected features list, it is observed that zEMG plays a significant role in the prediction of both arousal and valence. Classifiers were trained on the physiological signals: EOG, EMG, EDA, temperature, plethysmograph and respiration using the features listed in Table 2 to study their prediction ability. To sort the physiological signals based on the prediction accuracy, same set of features are fed to the classifiers listed earlier. The best prediction accuracy obtained and the corresponding classifier for each of the physiological signals is shown in Table 5. The graphic illustration is available as Figure S2 at: <https://github.com/armanjupriya-er/er-comparison-supplementary>.

Study of the prediction capability of time and frequency domain features of EOG, EMG, EDA, temperature, plethysmograph and respiration indicates that the plethysmograph shows an arousal prediction accuracy of 55.50%, which is inferior to the EEG modality by 0.89%, while the EOG shows valence prediction accuracy of 60.00%, which is better than the combination of all physiological signals by 1.69%. Features of EDA and zEMG each resulted in valence prediction accuracy of 58.25%. The sorted order of physiological signals based on arousal prediction accuracy is as follows: plethysmograph, EOG (hEOG + vEOG), vEOG, hEOG, zEMG, tEMG, temperature, EMG (tEMG + zEMG), respiration, EDA, whereas based on valence prediction accuracy the sorted order is EOG (hEOG + vEOG), EDA, zEMG, hEOG, respiration, tEMG, vEOG, EMG (tEMG + zEMG), temperature, plethysmograph. The valence prediction accuracy of EOG is superior to zEMG and EDA by 1.75% at the cost of higher number of electrodes (EOG requires four electrodes, whereas zEMG and GSR each require two electrodes). The results indicate that the valence prediction accuracy comes from muscle activity. Another notable observation is that, in a high dimensional feature space, ensemble classifiers (logit boost, stacking) perform better (Table 4), and in a low dimensional feature space, statistical models (logistic regression, LDA, QDA) perform better (Table 5) which is in line with literature [37].

The results of the experiment in comparison with the state-of-the-art (SOTA) is presented as supplementary Table S1 at <https://github.com/armanjupriya-er/er-comparison->

Modality	Model	Arousal		Valence		
		Accuracy	F1-score	Model	Accuracy	F1score
EMG (tEMG + zEMG)	Logistic	52.00	0.414	Logistic	56.75	0.502
EOG (hEOG + vEOG)	Naïve Bayes	53.75	0.513	LDA	60.00	0.601
EDA	Logit Boost	51.25	0.427	Logistic	58.25	0.446
hEOG	QDA	53.00	0.460	Logit Boost	57.25	0.573
Plethysmograph	LDA	55.50	0.460	Logistic	55.00	0.417
Respiration	Logit Boost	51.75	0.454	LDA	57.25	0.469
tEMG	Naïve Bayes	52.50	0.397	Logit Boost	57.25	0.537
Temperature	Logit Boost	52.25	0.385	LDA	56.75	0.426
vEOG	QDA	53.75	0.528	Logistic	57.25	0.536
zEMG	Logistic	52.50	0.404	Logistic	58.25	0.522
FC6 EEG Electrode	Stacking	52.25	0.475	QDA	59.00	0.500

Table 5. Prediction accuracy of physiological signals and FC6 EEG electrode

supplementary. It is observed that only 36.4% of the studies [10, 12, 19, 38] in literature have used subject independent ER; while all other studies in literature are on subject specific ER. All the subject independent studies have used LOSO validation, and the reported results are average of the LOSO results across the subjects. The current experiment is subject independent, and it varies from all the previous studies as it splits the subjects into 70:30 ratio for training and testing respectively. The results reported in Tables 4, 5, 6 and S1 are from the test set.

The performance of regularized deep fusion of kernel machines (RDFKMs) on EEG, EMG, EDA and respiratory rate [38] and pretrained inception ResNet v2 on facial expression, EEG and GSR modalities [12] were explored in recent literature. Similarly, recent research investigated the performance of statistical features on combination of multiple modalities [10, 19]. Unlike the experiment carried out in this work, all the above-mentioned recent research reported the average LOSO validation score as the final accuracy. Also, some of the recent works did not publish the cut-off score used to distinguish low and high values in the arousal and valence dimensions [10, 38] whereas, two other recent works mentioned the cut-off score as 4.5 [19] and cut-off score range as [1.0,3.0] (for low) and [7.0, 9.0] (for high) [12]. This is in contrast to the experiment carried out in this work which uses scale ranges of [0, 5.0] and (5.0, 9.0) as low and high, respectively.

The accuracy obtained by proposed unimodal valence recognition using EOG and multimodal valence recognition using zEMG and EOG is better than the accuracy obtained in literature [10, 12] by 5.44% and 11.52% respectively, but less than the accuracy obtained in literature [19, 38] by 16.87% and 6.97%, respectively. The accuracy obtained by proposed unimodal arousal recognition using EEG or plethysmograph is better than the accuracy obtained in Ref. [12] by 4.08% and is less compared to all other methods. This research work is not compared with the subject dependent ER studies listed in Table S1, as this experiment is about subject independent ER. The low accuracy reported in this experiment can be partly attributed to the dataset used to report the test accuracy. This experiment specifically uses a separate test set while all other subject independent ER work [10, 12, 19, 38] listed in Table S1

Metric	Proposed Unimodal-1	Proposed Unimodal-2	Proposed Multimodal-1	Proposed Unimodal-3	Proposed Unimodal-4	Proposed Unimodal-5
Modality	EEG	Plethysmograph	zEMG, EOG	EOG	EDA	zEMG
Class Labels	LA, HA	LA, HA	LV, HV	LV, HV	LV, HV	LV, HV
Method	Stacking	LDA	Logit Boost	LDA	Logistic	Logistic
Accuracy (%)	56.00	55.50	59.00	60.00	58.25	58.25
Precision	0.561	0.644	0.596	0.614	0.760	0.575
Recall	0.560	0.555	0.590	0.600	0.583	0.583
F1-Score	0.560	0.460	0.592	0.601	0.446	0.522
Kappa	0.121	0.091	0.177	0.207	0.045	0.081
Statistic						
ROC Area	0.554	0.582	0.608	0.638	0.586	0.546
TPR	0.560	0.555	0.590	0.600	0.583	0.583
FPR	0.439	0.466	0.411	0.387	0.542	0.508
TP Count (0 as 0)	114	25	102	114	7	33
TN Count (1 as 1)	110	197	134	126	226	200
FN Count (0 as 1)	81	170	72	60	167	141
FP Count (1 as 0)	95	8	92	100	0	26

Table 6.
Evaluation metrics for
the proposed models

reports an average of LOSO accuracy. Also, in this experiment, the same set of features is used across different modalities. More research is needed to determine whether modality specific features improve the prediction accuracy. Table 6 shows additional evaluation metrics for the proposed methods, including accuracy, ROC area, kappa statistic, precision, recall, true positive rate (TPR), false positive rate (FPR), F1-score, true positive count (TP count), false positive count (FN count), true negative count (TN count) and false negative count (FN count). According to the kappa statistic, F1-score and accuracy, EEG is better suited for arousal prediction, whereas EOG is better suited for valence prediction. The ROC area reported for arousal prediction by EEG modality is less compared to the plethysmograph modality by 0.028. From an ergonomic perspective, obtaining a plethysmograph signal is easier compared to obtaining EEG signals.

5. Conclusion

The experimental results of this work suggest that arousal dimension prediction ability is high for EEG signals, while valence dimension prediction ability is high for the combination of EOG and zEMG signals. In addition, valence can be measured from the eyes (EOG) while arousal can be measured from the changes in blood volume (plethysmograph). Also, muscle activity plays a significant role in valence prediction.

Further research is required to examine whether the prediction ability of the EEG signal is resulting from brain regions associated with muscle activity or not. Whether modality specific features improve the prediction accuracy or not is yet to be explored. The experiment needs to be repeated on other existing or new datasets to identify the best modality for each emotion dimension. To determine the effect of stimulus on eye muscle, further study of eye movements while expressing emotions can be performed.

References

1. Darwin C. The expression of the emotions in man and animals. Chicago: The University of Chicago Press; 1965.
2. Lange CG, James W (Eds). The emotions. Williams & Wilkins Co; 1922. 1.
3. Cannon WB. The James–Lange theory of emotions: a critical examination and an alternative theory. *The Am J Psychol.* 1927; 39: 106-24.
4. Wang Z, Zhou X, Wang W, Liang C. Emotion recognition using multimodal deep learning in multiple psychophysiological signals and video. *Int J Machine Learn Cybernetics.* 2020; 11: 923-34. doi: [10.1007/s13042-019-01056-8](https://doi.org/10.1007/s13042-019-01056-8).
5. Huang H, Hu Z, Wang W, Wu M. Multimodal emotion recognition based on ensemble convolutional neural network. *IEEE Access.* 2020; 8: 3265-71. doi: [10.1109/access.2019.2962085](https://doi.org/10.1109/access.2019.2962085).
6. Hassan MM, Alam MGR, Uddin MZ, Huda S, Almogren A, Fortino G. Human emotion recognition using deep belief network architecture. *Inf Fusion.* 2019; 51: 10-18. doi: [10.1016/j.inffus.2018.10.009](https://doi.org/10.1016/j.inffus.2018.10.009).
7. Zhu J, Wei Y, Feng Y, Zhao X, Gao Y. Physiological signals-based emotion recognition via high-order correlation learning. *ACM Trans Multimedia Comput Commun Appl.* 2020; 15: 1-18. doi: [10.1145/3332374](https://doi.org/10.1145/3332374).
8. Poria S, Peng H, Hussain A, Howard N, Cambria E. Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis. *Neurocomputing.* 2017; 261: 217-30. doi: [10.1016/j.neucom.2016.09.117](https://doi.org/10.1016/j.neucom.2016.09.117).
9. Yin Z, Zhao M, Wang Y, Yang J, Zhang J. Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. *Computer Methods Programs Biomed.* 2017; 140: 93-110. doi: [10.1016/j.cmpb.2016.12.005](https://doi.org/10.1016/j.cmpb.2016.12.005).

10. Bota P, Wang C, Fred A, Silva H. Emotion assessment using feature fusion and decision fusion classification based on physiological data: are we there yet?. *Sensors (Switzerland)*. 2020; 20: 4723. doi: [10.3390/s20174723](https://doi.org/10.3390/s20174723).
11. Dzieżyc M, Gjoreski M, Kaziienko P, Saganowski S, Gams M. Can we ditch feature engineering? End-to end deep learning for affect recognition from physiological sensor data. *Sensors (Switzerland)*. 2020; 20: 1-21. doi: [10.3390/s20226535](https://doi.org/10.3390/s20226535).
12. Cimtay Y, Ekmekcioglu E, Caglar-Ozhan S. Cross-subject multimodal emotion recognition based on hybrid fusion. *IEEE Access*. 2020; 8: 168865-78. doi: [10.1109/access.2020.3023871](https://doi.org/10.1109/access.2020.3023871).
13. Tan C, Ceballos G, Kasabov N, Subramaniam NP. FusionSense: emotion classification using feature fusion of multimodal data and deep learning in a brain-inspired spiking neural network. *Sensors (Switzerland)*. 2020; 20: 5328. doi: [10.3390/s20185328](https://doi.org/10.3390/s20185328).
14. Fu J, Mao Q, Tu J, Zhan Y. Multimodal shared features learning for emotion recognition by enhanced sparse local discriminative canonical correlation analysis. *Multimedia Syst*. 2017; 25: 451-61. doi: [10.1007/s00530-017-0547-8](https://doi.org/10.1007/s00530-017-0547-8).
15. Li W, Chu M, Qiao J. Design of a hierarchy modular neural network and its application in multimodal emotion recognition. *Soft Comput*. 2019; 23: 11817-28. doi: [10.1007/s00500-018-03735-0](https://doi.org/10.1007/s00500-018-03735-0).
16. Wang Z, Zhou X, Wang W, Liang C. Emotion recognition using multimodal deep learning in multiple psychophysiological signals and video. *Int J Machine Learn Cybernetics*. 2020; 11: 923-34. doi: [10.1007/s13042-019-01056-8](https://doi.org/10.1007/s13042-019-01056-8).
17. Kaya H, Gürpınar F, Salah AA. Video-based emotion recognition in the wild using deep transfer learning and score fusion. *Image Vis Comput*. 2017; 65: 66-75. doi: [10.1016/j.imavis.2017.01.012](https://doi.org/10.1016/j.imavis.2017.01.012).
18. Tzirakis P, Chen J, Zafeiriou S, Schuller B. End-to-end multimodal affect recognition in real-world environments. *Inf Fusion*. 2021; 68: 46-53. doi: [10.1016/j.inffus.2020.10.011](https://doi.org/10.1016/j.inffus.2020.10.011).
19. Ayata D, Yaslan Y, Kamasak ME. Emotion recognition from multimodal physiological signals for emotion aware healthcare systems. *J Med Biol Eng*. 2020; 40: 149-57. doi: [10.1007/s40846-019-00505-7](https://doi.org/10.1007/s40846-019-00505-7).
20. Frank E, Hall MA, Witten IH. *The WEKA workbench. Online appendix for data mining: practical machine learning tools and techniques*. 4th ed. Morgan Kaufmann; 2016.
21. Koelstra S, Muhl C, Soleymani M, Lee J-S, Yazdani A, Ebrahimi T, Pun T, Nijholt A, Patras I, DEAP . A database for emotion analysis using physiological signals. *IEEE Trans Affective Comput*. 2012; 3: 18-31. doi: [10.1109/t-affc.2011.15](https://doi.org/10.1109/t-affc.2011.15).
22. Wang XW, Nie D, Lu BL. EEG-based emotion recognition using frequency domain features and support vector machines. In: Lu BL, Zhang L, Kwok J (Eds.). *Neural information processing. ICONIP 2011. Lecture notes in computer science*, Vol. 7062. Berlin, Heidelberg: Springer; 2011. doi: [10.1007/978-3-642-24955-6_87](https://doi.org/10.1007/978-3-642-24955-6_87).
23. Jenke R, Peer A, Buss M. Feature extraction and selection for emotion recognition from EEG. *IEEE Trans Affective Comput*. 2014; 5(3): 327-39. doi: [10.1109/TAFFC.2014.2339834](https://doi.org/10.1109/TAFFC.2014.2339834).
24. Hjorth B. The physical significance of time domain descriptors in EEG analysis. *Electroencephalography Clin Neurophysiol*. 1973; 34(3): 321-5. ISSN 0013-4694 doi: [10.1016/0013-4694\(73\)90260-5](https://doi.org/10.1016/0013-4694(73)90260-5).
25. Hjorth B. EEG analysis based on time domain properties. *Electroencephalography Clin Neurophysiol*. 1970; 29(3): 306-10. doi: [10.1016/0013-4694\(70\)90143-4](https://doi.org/10.1016/0013-4694(70)90143-4).
26. Welch P. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans Audio Electroacoustics*. 1967; 15(2): 70-3. doi: [10.1109/TAU.1967.1161901](https://doi.org/10.1109/TAU.1967.1161901).
27. Hall MA. *Correlation-based feature subset selection for machine learning*. Hamilton, New Zealand: University of Waikato; 1998.
28. Scrivener CL, Reader AT. Variability of EEG electrode positions and their underlying brain regions: visualizing gel artifacts from a simultaneous EEG-fMRI dataset. *Brain Behav*. 2022; 12(2): e2476. doi: [10.1002/brb3.2476](https://doi.org/10.1002/brb3.2476).

-
29. Portugal LCL, Alves RDCS, Orlando FJ, Sanchez TA, Mocaiber I, Volchan E, Smith Erthal F, Antunes David I, Kim J, Oliveira L, Padmala S, Chen G, Pessoa L, Garcia Pereira M. Interactions between emotion and action in the brain. *NeuroImage*. 2020; 214: 116728. ISSN 1053 - 8119 doi: [10.1016/j.neuroimage.2020.116728](https://doi.org/10.1016/j.neuroimage.2020.116728).
 30. Koch SBJ, Mars RB, Toni I, Roelofs K. Emotional control, reappraised. *Neurosci Biobehavioral Rev*. 2018; 95: 528-34. ISSN 0149-7634 doi: [10.1016/j.neubiorev.2018.11.003](https://doi.org/10.1016/j.neubiorev.2018.11.003).
 31. John GH. Pat langley: estimating continuous distributions in bayesian classifiers. In: Eleventh Conference on Uncertainty in Artificial Intelligence, San Mateo; 1995. p. 338-45.
 32. le Cessie S, van Houwelingen JC. Ridge estimators in logistic regression. *Appl Stat*. 1992; 41(1): 191-201.
 33. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning : with applications in R. New York: Springer; 2013.
 34. Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting. *The Annals of Statistics*. 2000; 28: 337-407. doi: [10.1214/aos/1016218223](https://doi.org/10.1214/aos/1016218223).
 35. Wolpert D. Stacked generalization. *Neural Networks*. 1992; 5: 241-59. doi: [10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1).
 36. Rimbert S, Al-Chwa R, Zaepffel M, Bougrain L. Electroencephalographic modulations during an open- or closed-eyes motor task. *Peer J*. 2018; 6: e4492. doi: [10.7717/peerj.4492](https://doi.org/10.7717/peerj.4492).
 37. Pappu V, Pardalos PM. High-dimensional data classification. In: Aleskerov F, Goldengorin B, Pardalos P, (Eds.). *Clusters, orders, and trees: methods and applications*. Springer optimization and its applications, Vol. 92. New York, NY: Springer; 2014. doi: [10.1007/978-1-4939-0742-7_8](https://doi.org/10.1007/978-1-4939-0742-7_8).
 38. Zhang X, *et al*. Emotion recognition from multimodal physiological signals using a regularized deep fusion of kernel machine. *IEEE Trans Cybernetics*. 2021; 51(9): 4386-99. doi: [10.1109/TCYB.2020.2987575](https://doi.org/10.1109/TCYB.2020.2987575).

Further reading

39. Huang Y, Yang J, Liu S, Pan J. Combining facial expressions and electroencephalography to enhance emotion recognition. *Future Internet*. 2019; 11: 1-17. doi: [10.3390/fi11050105](https://doi.org/10.3390/fi11050105).
40. Kwon YH, Shin SB, Kim SD. Electroencephalography based fusion two-dimensional (2d)-convolution neural networks (CNN) model for emotion recognition system. *Sensors (Switzerland)*. 2018; 18: 1383. doi: [10.3390/s18051383](https://doi.org/10.3390/s18051383).
41. Karanchery S, Palaniswamy S. Emotion recognition using one-shot learning for human-computer interactions. In: 2021 International Conference on Communication, Control and Information Sciences (ICCISc); 2021. p. 1-8. doi: [10.1109/ICCISc52257.2021.9485024](https://doi.org/10.1109/ICCISc52257.2021.9485024).
42. Kuruvayil S, Palaniswamy S. Emotion recognition from facial images with simultaneous occlusion, pose and illumination variations using meta-learning. *J King Saud Univ - Computer Inf Sci*. 2021. ISSN 1319-1578. doi: [10.1016/j.jksuci.2021.06.012](https://doi.org/10.1016/j.jksuci.2021.06.012) (In press).
43. Sasidharakurup H, Nutakki C, Rajendran A, Venugopal P, Sumon M, Navaneethkumar L, Madhu H, Bipin GN, Shyam D. Spectral correlations in speaker-listener behavior during a focused duo conversation using EEG. In: *Proceedings of the Seventh International Conference on Advances in Computing, Communications and Informatics (ICACCI-2018)*, Bangalore, Karnataka, India, Sept 19-22, 2018.
44. Bodda S, Maya S, Potti M, Naryanan E, Sohan U, Bhuvaneshwari Y, Mathiyoth R, Diwakar S. Computational analysis of EEG activity during stance and swing gait phases. In: *Proceedings of the Third International Conference on Computing and Network Communications (CoCoNet'19)* (accepted), Trivandrum, Kerala, India, 2019.
45. Keshari T, Palaniswamy S. Emotion recognition using feature-level fusion of facial expressions and body gestures. In: 2019 4th International Conference on Communication and Electronics Systems (ICCES 2019), Coimbatore, TamilNadu, India; 2019. p. 1184-9.

-
46. Lawrance D, Palaniswamy S. Emotion recognition from facial expressions for 3D videos using siamese network. In: 2021 International Conference on Communication, Control and Information Sciences (ICCISc); 2021. p. 1-6. doi: [10.1109/ICCISc52257.2021.9484949](https://doi.org/10.1109/ICCISc52257.2021.9484949).

Emotion
recognition –
EEG and
physiological

Corresponding author

Suja Palaniswamy can be contacted at: p_suja@blr.amrita.edu

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com