L. Santamaria-Granados, M. Munoz-Organero, G. Ramirez-González, E. Abdulhay and N. Arunkumar, "Using Deep Convolutional Neural Network for Emotion Detection on a Physiological Signals Dataset (AMIGOS)," in IEEE Access, vol. 7, pp. 57-67, 2019.

DOI: https://dx.doi.org/10.1109/ACCESS.2018.2883213

Received October 10, 2018, accepted XXXX XX, 2018, date of publication XXXX XX, 2018, date of current version October 10, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Using Deep Convolutional Neural Network for Emotion Detection on a Physiological Signals Dataset (AMIGOS)

LUZ SANTAMARIA-GRANADOS [1], (Fellow, IEEE), MARIO MUNOZ-ORGANERO [2], (MEMBER, IEEE) AND GUSTAVO RAMIREZ-GONZALEZ.[3], (Member, IEEE)

[1]Faculty of Systems Engineering, University of Santo Tomás, Tunja 5878797, Colombia (e-mail: luz.santamaria@ustatunja.edu.co)
[2]Telematics Engineering Department, Universidad Carlos III de Madrid; Av. Universidad, 30, 28911 Leganes, Spain (e-mail: munozm@it.uc3m.es)
[3]Telematics Department, University of Cauca, Popayán, Colombia (e-mail:gramirez@unicauca.edu.co)

Corresponding author: Luz Santamaria-Granados (luz.santamaria@usantoto.edu.co)

ABSTRACT

Recommender systems have been based on context and content, now the technological challenge of making personalized recommendations based on the user emotional state arises, through physiological signals that are obtained from devices or sensors. This study applies the deep learning approach using a Deep Convolutional Neural Network (DCNN), on a dataset of physiological signals (Electrocardiogram -ECG- and Galvanic Skin Response -GSR-), in this case, the AMIGOS dataset. The detection of emotions is done by correlating these physiological signals with the data of arousal and valence of this dataset, to classify the affective state of a person. In addition, an application for emotion recognition based on classic machine learning algorithms is proposed to extract the features of physiological signals, in the domain of time, frequency and non-linear. This application uses a CNN for the automatic features extraction of the physiological signals and through fully connected network layers (FCN), the emotion prediction is made. The experimental results on the dataset AMIGOS, show that the method proposed in this study achieves a better precision of the classification of the emotional states, in comparison with the originally obtained by the authors of this dataset.

INDEX TERMS Emotion recognition, Deep Convolutional Neural Network, physiological signals, machine learning, AMIGOS dataset.

I. INTRODUCTION

During the last two decades, the MIT's affective computing research group has aroused great interest in scientific and academic communities that seek to improve the human emotional experience with technology [1]. Some challenges focus on deepening machine learning and deep algorithms, to ensure that the emotion recognition system has a high precision and robustness in the processing of physiological data [2]. The emotions computational models [3] have been applied to the recognition of affective states through physiological measures, such as Heart Rate Variability (HRV), Blood Volume Pulse (BVP), Skin Temperature (SKT) [4], Electrocardiogram (ECG), and Electrodermal Activity (EDA) [5], that come from the peripheral nervous system and central nervous system.

Affective states are subjective experiences classified in valence and arousal focuses [6]. Similarly, both focuses reflect the degree to which a person incorporates emotions into their conscious affective experience [7]. The stimulus of valence focus is associated with pleasurable or unpleasant aspects, in contrast with arousal focus that induces the activation or deactivation of an emotion. Some databases correlate the affective states with physiological signals [8] [9] [10], which are the result of emotions self-reported by people. The emotional categories are established in a circular structural model that contain basic emotions (for example, excited, happy, pleased, relaxed, peaceful, calm, sleepy, bored, sad, nervous, angry, and annoyed) to define the arousal and valence dimensions [11] [12].

The emergence of sensors and wearable devices as mechanisms for the acquisition of physiological data of people in their daily lives [13] has made possible the research in the recognition of emotional patterns, for the improvement of the user experiences in diverse contexts. Research on the management of the tourism industry highlights the importance of this type of devices for emotional recognition, such as the improvement of the tourism experience through the services personalization [14] [15], where the tourist's expectation is analyzed in three phases (before, during and after the tourist visit) for different dimensions or tourist activities. With regard to the dimension of the tourist attraction, the recommendation systems are an important tool before visiting the tourist destination. In the same way, the World Tourism Organization recognizes that in the market of the increasingly competitive tourist destination, the tourist attractions are more inclined towards the emotional benefits than the physical features and price of the destination [16].

For affective recognition, this paper focuses on exploring models of Deep Convolutional Neural Network (DCNN) [17] in comparison with traditional machine learning algorithms, which can be used as a framework for the emotions detection. The experimental tests for the classification of the emotional dimensions of arousal and valence were made with the dataset AMIGOS [10]. For the purpose of transforming the physiological signals, the QRS detection methods in [18] were applied in the prepocesing stage, which provides the RR intervals of the ECG. Likewise, the temporal series of the Skin Conductance Response (SCR) peaks of the GSR signals [19] were identified. A determining factor in the effectiveness of the emotion prediction is defined in the extraction and correlation of the features of the physiological signals ECG and GSR.

In this study, the authors present an analysis of the statistical techniques used for the manual features extraction of the physiological signals with respect the automatic features extraction that is better correlated with the states emotional. This paper is organized as follows: section 2 presents a review of literature related to the emotions detection from physiological signals. Section 3 describes the AMIGOS dataset used in the process of affective states recognition. Section 4 provides methods based on automatic learning and deep learning algorithms for the emotions classification. Finally, sections 5 and 6 present the results and conclusions of the experiment generated during this research.

II. RELATED WORK

This section presents researches on datasets for the multimodal emotions recognition and the affective states detection through physiological responses.

A. MULTIMODAL DATASET

Emotion is the degree to which a person reacts to changes in the context as a response to the elicitation that manifests itself in their affective states [5]. People use the senses to express the emotion experienced through gestures, speech or physiological responses. The correlation between emotions and physiological data determines the multimodal affect recognition. The contents of images [20], movie clips [21] and music videos [8] have been used to induce emotions that users appraisal with explicit measurements [22], in order to verify the arousal and valence levels. On the other hand, emotions elicited by multimedia content are implicitly recognized by means of physiological and brain signals, enabling the consolidation of a multimodal affective dataset that compares the affective response of people [23].

Precisely, the dataset ASCERTAIN [24] effects the personality and emotion recognition induced by 36 movie clips that have a duration of 58 to 128 seconds, with the registration of physiological signals (ECG and GSR), EEG and activity facial of 58 participants. AMIGOS dataset [10] detects the mood, affect and personality of 40 participants with the registration of their EEG, ECG, and GSR signals, as a result of the stimulus caused during the viewing of short and long videos.

Abadi et al. [25] for the affect detection analyzes the physiological response of the ECG, Electrooculogram (EOG) and trapezius-Electromyogram (EMG), and contrasts the brain signals (EEG and Magnetoencephalogram) of 30 participants who watched 36 movie clips from 80 seconds and 40 segments of one-minute music videos that are part of the DEAP dataset [8]. In the emotional state's recognition of 32 participants, DEAP includes physiological signals (GSR, BVP, SKT, EOG, and EMG) and EEG. Similarly, the multimodal database MAHNOB-HCI [26] contains physiological signals (ECG, GSR, SKT, and Respiration), eye gaze and EEG from 27 participants, who evaluated the emotion through various stimuli (20 emotional videos, 14 short videos, and 28 images).

Both DEAP and MANNOB-HCI demonstrate better EEG effectiveness in predicting arousal and physiological signals obtained a better outcome with valence. AMIGOS has the same behavior with EEG signals, but unlike [8] and [26], it obtained better f1-score outcome with arousal. The physiological features in DECAF had a better arousal recognition in the movie clips and a better valence outcome in the music clips. In ASCERTAIN the multimodal results (ECG and GSR) had a better performance in contrast to the EEG.

The works related to the affective recognition establish the experimentation of the users with diverse stimuli and the influence of the emotions in their physiological behaviors, therefore need arises to identify emotional patterns in the physiological features that improve the detection of the states affective. Moreover, section III describes the experiment with short videos of the AMIGOS affective dataset, used for the emotions recognition with the machine learning approaches proposed in this survey.

B. EMOTIONAL STATES DETECTION

The publications related to the affective recognition from physiological data have the purpose of constructing reliable models supported by techniques and machine learning algorithms, to discover patterns of the emotional states that are hidden in the physiological signals. Various methodologies have been explored for the preprocessing of data, the extraction, and selection of physiological features, as stages prior to the classification of emotion.

Some studies for the affect recognition of have implemented supervised classification approaches [27] such as k-Nearest Neighbor (k-NN) [28] [18], and Support Vector Machine (SVM) [9] [29]. The researchers defined keywords to validate the user's emotional responses through the valence and excitation model. The physiological signals are processed by sliding window technique [30] and the process of reducing the dimensionality of the features is based on the Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) techniques [29].

On the other hand, the Deep Learning approach applies non-linear transformations to physiological signals for the detection of features of human emotional behavior. In this context, CNN [31] techniques have been used for the automatic extraction of SCR and BVP features and 70 to 75% accuracy results have been obtained in the prediction of emotion (relaxation, anxiety, excitement, and fun). Other investigations validated the performance of affection models with deep learning using the multimodal DEAP database [32] [33] and adopted a multiple-fusion-layer-based ensemble classifier of stacked autoencoder (MESAE) framework, to extract the physiological features that were merged into an SAE network. The accuracy results in arousal and valence were 0.83 and 0.84 respectively.

Regarding to semi-supervised learning methodologies SAE was integrated with Deep Belief Network (DBN) using a Bayesian inference classification based decision fusion method [34], results of arousal were obtained in 73.1% and valence in 78.8%. In [35] they defined a hybrid model composed of a CNN and a Recurrent Neural Network (RNN). As a requirement for the sequential processing in the CNN, the features were extracted and the prediction was made in the Long Short-Term Memory (LSTM) unit of the RNN. This model obtained an accuracy of 74.1% for arousal and 72.1% for valence. The models based on CCN and DNN [36] showed better results in the affective classification when using the image domain of the EEG signals [37].

The related works deal with the trend of deep learning for the emotions detection related to heart disease, mental disorder, and stress. However, to validate the affective models there is a limitation in the access to small physiological datasets [38] or there is a problem in obtaining correct data [17]. Therefore, it is necessary to publish repositories of physiological datasets, so that researchers can test the classification models that can be used in the personalization of tourist services or any domain.

III. AMIGOS DATASET

The validation of the emotional classifier is done with A dataset for Mood, personality and affect research on Individuals and GrOupS (AMIGOS) [10]. This dataset is the result of two experiments related to the multimodal study of emotional responses. In the first, 40 participants watched 16 short videos (duration < 250 seconds), in the second, 17 people individually and five groups of four participants watched four long videos (duration > 14 minutes). In both experiments, neuro-physiological signals were captured from the subjects during the elicitation of emotion [21].

TABLE 1. Classification of the 16 short videos with the physiological signals instances that were recorded during the presentation of the stimuli of each subject [10].

Video	Instances	Duration	Quadrant	Film	Clips
10	12225	96	LAHV	August Rush	6
13	7229	57	LAHV	Love Actually	4
138	15610	122	LALV	The Thin Red Line	7
18	10575	83	LAHV	House of Flying Daggers	5
19	16106	126	LALV	Exorcist	8
20	8335	65	LALV	My girl	5
23	14265	112	LALV	My Bodyguard	7
30	9717	76	HALV	Silent Hill	5
31	19886	155	HALV	Prestige	9
34	8417	66	HALV	Pink Flamingos	5
36	8698	68	HALV	Black Swan	5
4	11621	91	HAHV	Airplane	6
5	14347	112	HAHV	When Harry Met Sally	7
58	8181	64	LAHV	Mr Beans Holiday	4
80	13047	102	HAHV	Love Actually	6
9	9630	75	HAHV	Hot Shots	5

Electroencephalogram (EEG) signals were recorded using the Emotiv EPOC Neuroheadset containing 14 electrodes for AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4 channels. The physiological signals were recorded with the ECG Shimmer 2R5 platform from three electrodes for the Electrocardiogram (ECG right and ECG left channels) and two electrodes for the Galvanic Response of the Skin (GSR channel). The physiological data were preprocessed with a sampling frequency of 128 Hz.

The affective levels of the participants were reported in a self-assessment (arousal, valence, dominance, liking, familiarity and seven basic emotions) and in an external annotation (arousal and valence). The five dimensions are measured in the scale of 1(low) to 9 (high), the basic emotions (neutral, disgust, happiness, surprise, anger, fear, and sadness) are binary values. Specifically, this study focuses on the experiment with the 16 short videos, due to the fact that a long video is more likely to elicit diverse emotional states according to the scenes presented. That is, the emotion appraisal is determined by the changes that the subject can experience in the context. The experienced emotions can change through a process of regulating emotion, which determines the effects on human behavior [39].

The classification of the 16 short videos by quadrants of valence and arousal (high and low) was performed by [10] according to the elicitation of the emotion, for each participant 94 clips were recorded according to the duration of each video (see table 1). The first 20 seconds of each clip, included five seconds from the beginning of the stimuli, then were generated non overlapping intermediate segments of 20 seconds, excepting for the final clip.

Figure 1 shows the distribution of the valence and arousal mean of the self-assessing participants during the experiment, 40 * 16 = 640 instances are available. However, it is observed that the videos 20 and 23 tend to a neutral value of arousal, that is, the intensity of the emotion is not so

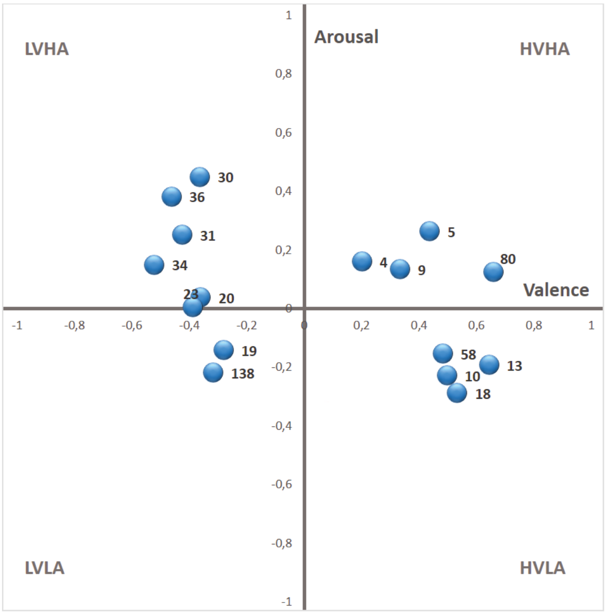

FIGURE 1. Distribution of mean ratings of valence and arousal of self-assessment of the 16 short videos. Scale from -1 to 1.

marked. Therefore, using the k-means classification method, we define the four clusters with the thresholds for the labels of arousal and valence [32]. Figure 2 shows the clusters with a threshold of (5, 5) for the two or four classes of low or high emotion and were obtained with the K-means clustering method [27]. In the current study, the emotional classification is defined as low and high subjective scale for the valence and arousal dimensions.

IV. PROPOSED METHODS

Affective computing involves the design of machine learning models to discover physiological patterns of affective states from datasets. In this research, we propose the validation of supervised learning algorithms and Deep Learning for the efficient emotion detection. Therefore, in figure 3, it is shown the system with the components to load the dataset in a data frame, as a requirement for the preprocessing of the ECG and GSR signals. Then, the feature extraction stage can be developed explicitly or implicitly. The first uses hand-crafted functions to obtain features in the time or frequency domain, which can be selected with machine learning algorithms. The second, with deep learning, extracts automatic representations of the features. Finally, the models are trained and tested with algorithms from the two approaches.

A. MACHINE LEARNING

1) Data preprocessing

As a previous step to the features extraction of the physiological signals, the detection of peaks of the ECG and GSR signals is performed, because the emotions generate significant changes in these segments. The Heart Rate Variability (HRV) analysis is an affective diagnostic tool to determine the beat to beat interval (RR interval) [18]. The values between a RR interval correspond to the time between two peaks R, which is calculated through a standard wave of the QRS complex. The ECG signal is transformed with the PanTomkins QRS detection algorithm proposed in [40]. The signal is filtered to reduce the noise with cutoff frequencies of 0.5 and 15 Hz and uses an adaptive threshold for the detection of the QRS complex (see figure 4).

TABLE 2. Notation of features extracted from ECG and GSR signals [42], [45]

Signal	Features group	Description of the extracted features
ECG	Time Domain (1 - 13)	meanNN, medianNN, standardDeviationNN, rmSSD, pnn50, pnn20, coeffVariationSD, medianADNN, coeffVariationNN, mCoeffVariationNN, shannonEntropy, HRVtriangular, and numArtifacts.
	Frequency Domain (14 - 24)	peakHF, hfTotalPowerRatio, normalizedHF, peakLF, lfhfRatio, lfTotalPowerRatio, normalizedLF, totalPower, ulfPeak, vhfPeak, and vlfPeak.
	Non Linear (25 - 33)	correlation dimension, entropy (SVD, HF, LF, VLF, and shannon), fractal dimension (higushi and petrosian), and fisher information.
GSR	Mean, standard deviation, max, min, kurtosis, and skew (34 - 87)	EDA at apex, SCR width, amplitude, decay time, half amplitude, half amplitude (index and indexpre), latency, and rise time.

Similarly, the GSR signal is preprocessed using bandpass filters to reduce noise with cutoff frequencies of 0.05 and 19 Hz [41]. Then it is resampled with a digital phase filter of 10 Hz. During SCR peak detection a standard method is used that identifies the max, min and offset indexes of the signal GSR [42]. So, the threshold of the amplitude is determined and the features between SCR peaks are calculated (see figure 5).

2) Extraction and selection of features

The affect detection requires an adequate features extraction of the signals, which correlate with the emotional states recorded by the participants in the self-assessment. That is, the relationship between features and emotions determines the physiological reaction [43] and is taken as input to the predictor. Parametric measurements of the ECG signals in the time domain quantify the variability of interbeat intervals (IBI) measurements successive. The power distribution is determined in the frequency domain and the unpredictability of a series IBI is quantified in the non-linear according to [44] [45].

For the case of GSR signals, are extracted statistics in the time domain related to amplitude, rise time, decay time, latency, mean amplitude indexes and SCR peak indexes. Because each GSR signal produces a set of measurements by the amount of detected SCR peaks, some measures of central tendency, dispersion variation and distribution are applied. In Table 2, the features generated from the peaks of the ECG and GSR signals are described.

After the process of extracting features, machine learning

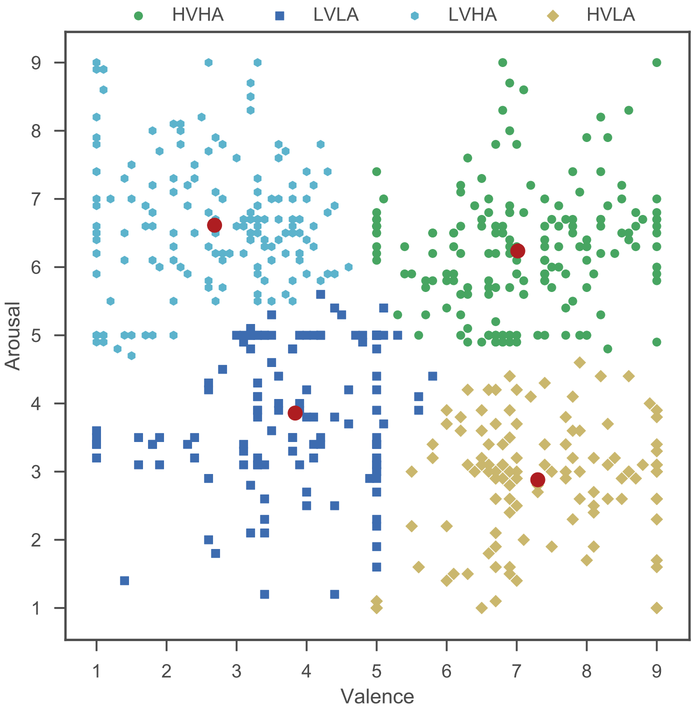

FIGURE 2. Clustering of valence and arousal of self-assessment of the short videos. Scale from 1 to 9.

algorithms are used to filter the redundant features that can cause overfitting in the classification model [46].

B. DEEP CONVOLUTIONAL NEURAL NETWORK

Deep learning is an area of machine learning based on algorithms and techniques for modeling high-level abstractions in datasets [47], such as the patterns recognition in images, text or emotions. The learning levels take as input the results of the previous levels, which are transformed into insights, to train and validate the classification model.

The DCNN architecture proposed for the emotion detection system was adapted from the work of [48], with the Keras framework [49]. The DCNN involves a sequence of CNN layers and pooling layers to automatically extract features from the physiological signals. Fully connected layers are located in front of CNN, operate on all nodes and are used to predict the affective state.

In this study, CNN layers are considered fuzzy filters [50] that reduce noise and discover particular morphological patterns in the R peaks of the ECG signals and the SCR peaks of the GSR signals. Initially, the transformation implemented by the neuronal layers is parameterized by its weight w, since the neurons learn to discover the correct values (convolution kernel), without affecting the behavior of the other layers [51]. That is, in the 1D convolutional layer the features vector of the physiological signals resulting from the transformation of the input data x, is defined in equation 1.

$$x_i^l = f\left(\sum_j w_{ij}^l x_j^{l-1} + b_i^l\right) \tag{1}$$

Where x_j^{l-1} represents the input vector to the convolutional function, w_{ij}^l denotes the kernel weight between the i^{th} and j^{th} neurons of the layers l and $l-1$ respectively. b_i^l is the bias coefficient of the neuron i^{th} in the layer l and x_i^l indicates the output of the convolutional layer.

In the CNN layers and fully connected layers, the activation function of the Rectified Linear Unit (ReLU) is set, which handles a threshold of 0 for the negative values. This $ReLU(x)$ function is calculated as equation 2:

$$f(x) = max(0, x_i) \tag{2}$$

The max-pooling layers are alternated between the CNN layers because of to segment a convolutional region that can increase the robustness of the features and reduce the dimensionality of the physiological signals vector. As a regularization technique to decrease the overfitting in the layers of the neural network, the dropout with a value of 0.5 is added. The output layers of the fully connected network are configured with the softmax classifier, with the purpose that the hidden layers verify the probability of predicting the emotion.

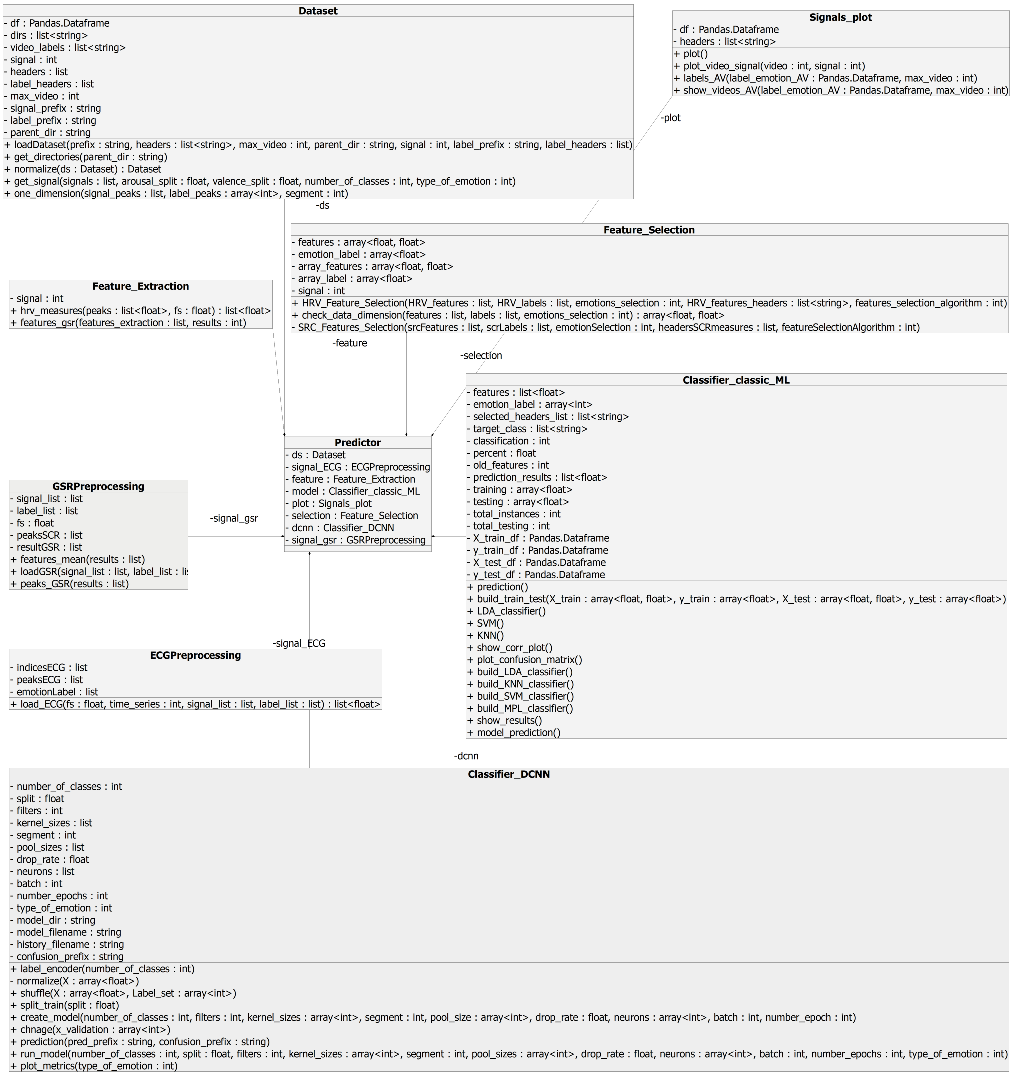
Dataset
- df : Pandas.Dataframe
- dirs : list<string>
- video_labels : list<string>
- signal : int
- headers : list
- label_headers : list
- max_video : int
- signal_prefix : string
- label_prefix : string
- parent_dir : string
+ loadDataset(prefix : string, headers : list<string>, max_video : int, parent_dir : string, signal : int, label_prefix : string, label_headers : list)
+ get_directories(parent_dir : string)
+ normalize(ds : Dataset) : Dataset
+ get_signal(signals : list, arousal_split : float, valence_split : float, number_of_classes : int, type_of_emotion : int)
+ one_dimension(signal_peaks : list, label_peaks : array<int>, segment : int)
Signals_plot
- df : Pandas.Dataframe
- headers : list<string>
+ plot()
+ plot_video_signal(video : int, signal : int)
+ labels_AV(label_emotion_AV : Pandas.Dataframe, max_video : int)
+ show_videos_AV(label_emotion_AV : Pandas.Dataframe, max_video : int)
-plot
-ds
Feature_Selection
- features : array<float, float>
- emotion_label : array<float>
- array_features : array<float, float>
- array_label : array<float>
- signal : int
+ HRV_Feature_Selection(HRV_features : list, HRV_labels : list, emotions_selection : int, HRV_features_headers : list<string>, features_selection_algorithm : int)
+ check_data_dimension(features : list, labels : list, emotions_selection : int) : array<float, float>
- SRC_Features_Selection(srcFeatures : list, scrLabels : list, emotionSelection : int, headersSCRmeasures : list, featureSelectionAlgorithm : int)
Feature_Extraction
- signal : int
+ hrv_measures(peaks : list<float>, fs : float) : list<float>
+ features_gsr(features_extraction : list, results : int)
-feature
-selection
Classifier_classic_ML
- features : list<float>
- emotion_label : array<int>
- selected_headers_list : list<string>
- target_class : list<string>
- classification : int
- percent : float
- old_features : int
- prediction_results : list<float>
- training : array<float>
- testing : array<float>
- total_instances : int
- total_testing : int
- X_train_df : Pandas.Dataframe
- y_train_df : Pandas.Dataframe
- X_test_df : Pandas.Dataframe
- y_test_df : Pandas.Dataframe
+ prediction()
+ build_train_test(X_train : array<float, float>, y_train : array<float>, X_test : array<float, float>, y_test : array<float>)
+ LDA_classifier()
+ SVM()
+ KNN()
+ show_corr_plot()
+ plot_confusion_matrix()
+ build_LDA_classifier()
+ build_KNN_classifier()
+ build_SVM_classifier()
+ build_MPL_classifier()
+ show_results()
+ model_prediction()
Predictor
- ds : Dataset
- signal_ECG : ECGPreprocessing
- feature : Feature_Extraction
- model : Classifier_classic_ML
- plot : Signals_plot
- selection : Feature_Selection
- dcnn : Classifier_DCNN
- signal_gsr : GSRPreprocessing
GSRPreprocessing
- signal_list : list
- label_list : list
- fs : float
- peaksSCR : list
- resultGSR : list
+ features_mean(results : list)
+ loadGSR(signal_list : list, label_list : li
+ peaks_GSR(results : list)
-signal_gsr
-signal_ECG
ECGPreprocessing
- indicesECG : list
- peaksECG : list
- emotionLabel : list
+ load_ECG(fs : float, time_series : int, signal_list : list, label_list : list) : list<float>
-dcnn
Classifier_DCNN
- number_of_classes : int
- split : float
- filters : int
- kernel_sizes : list
- segment : int
- pool_sizes : list
- drop_rate : float
- neurons : list
- batch : int
- number_epochs : int
- type_of_emotion : int
- model_dir : string
- model_filename : string
- history_filename : string
- confusion_prefix : string
+ label_encoder(number_of_classes : int)
- normalize(X : array<float>)
+ shuffle(X : array<float>, Label_set : array<int>)
+ split_train(split : float)
+ create_model(number_of_classes : int, filters : int, kernel_sizes : array<int>, segment : int, pool_size : array<int>, drop_rate : float, neurons : array<int>, batch : int, number_epoch : int)
+ chnage(x_validation : array<int>)
+ prediction(pred_prefix : string, confusion_prefix : string)
+ run_model(number_of_classes : int, split : float, filters : int, kernel_sizes : array<int>, segment : int, pool_sizes : array<int>, drop_rate : float, neurons : array<int>, batch : int, number_epochs : int, type_of_emotion : int)
+ plot_metrics(type_of_emotion : int)

FIGURE 3. Software components for the emotion recognition system, with a deep learning approach and classic machine learning algorithms.

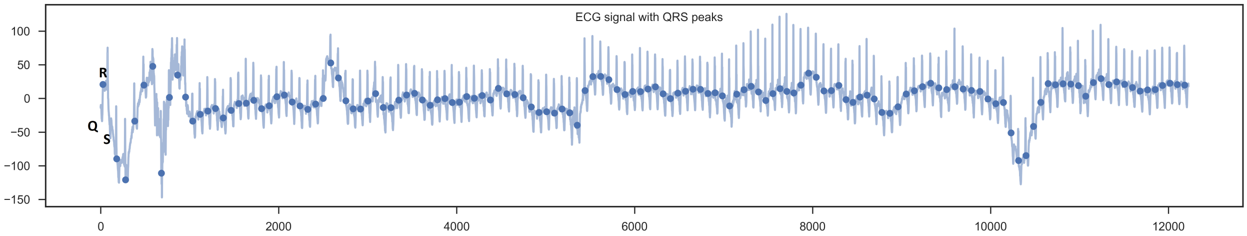

FIGURE 4. Detection of RR interval in the ECG signal. Dataset AMIGOS [10], participant 1, video 10.

During the supervised training, the loss is minimized with the Root Mean Square Propogation (RMSProp) [52] optimizer, since it adjusts the learning rate adaptively. Initially, the learning rate is set to 0.001. Once the model is executed, the knowledge base is consolidated between the vector of physiological features and the class vector. Then, to evaluate the emotion recognition, in the fully connected layer the cross-entropy loss function is set, which determines the degree of correspondence of the target output vector y_i, with the predicted output vector c_j, as follows in equation 3:

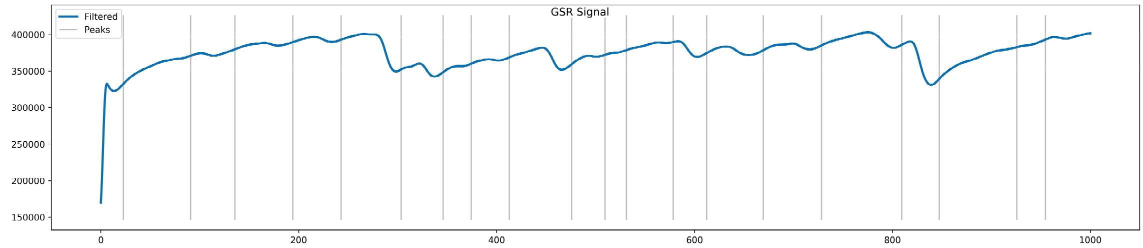

FIGURE 5. Detection of peaks in the GSR signal. Dataset AMIGOS [10], participant 1, video 10.

$$E = \frac{1}{2}\sum_{j=1}^{N}(y_i - c_j)^2 \tag{3}$$

The emotion recognition model based on deep learning algorithms is shown in figure 6. The structure is defined by an input layer that connects the vector of physiological features with the neurons of the first convolutional layer. Which, in turn, connects with three consecutive convolutional layers, to extract the features of the ECG and GSR signals. further, it is appreciated the transformation process of the input vectors in local patches inside a convolution window [51]. Each 1D CNN contains a sequence of temporal data for the recognition of local patterns, which can be learned from the physiological signals morphology. The functionality of the CNN layers is given by the convolution kernel that obtains the local patches and the Max pooling extracts the windows from the feature vectors to generate the downsampling output vector.

The vector resulting from physiological features extraction state is sent to the input neurons of the three FCN, to perform the training and testing process of the model. The last FCN layer is used to predict the affective state.

V. EXPERIMENTAL RESULTS

Emotion recognition models are tested through the AMIGOS dataset. In the first validation with the deep learning algorithms, the automatic extraction of the features is performed from the R peaks and SCR peaks. In contrast, with the instances of physiological signals that are loaded directly from the data frame to the convolutional layers. The second experiment is based on some classic machine learning algorithms to extract, select and detect emotions. Each physiological signal is made up of 640 instances (40 participants * 16 videos), but at the time of consolidating the data frame, null values were found, therefore, it was reduced to 603 instances.

A. EMOTION DETECTION WITH DCNN

During the experiment, the configuration parameters previously explained were defined for the training and testing of the deep learning model. Once the physiological signals have been preprocessed, it is defined a segment of length of 200 R peaks for the input vector of the ECG signal. For the GSR signal, it is specified an input segment of 20 SCR peaks. The values of each vector were normalized with the calculation of the mean and the standard deviation of all the points of the signal segment. The sizes of the kernel and the filter of the CNN layers affect the features detection that is represented in a convolution vector.

TABLE 3. The classification accuracy for the CNN model

Physiological data		Arousal		Valence	
Signal	**Input lenght**	**Train Acc.**	**Test Acc.**	**Train Acc.**	**Test Acc.**
ECGL	200	0.83	0.82	0.75	0.71
ECGL-ECGR	200	0.83	0.76	0.79	0.75
ECGL	15000	0.82	0.82	0.66	0.72
GSR	15000	0.66	0.69	0.66	0.67
GSR	20	0.71	0.71	0.73	0.75

For the ECG vector, the kernel size for the four convolutional layers is defined at 15, 10, 5 and 1. In GSR vector, it was configured at 10, 3, 1 and 1. The max-poling sizes were defined in 5, 2, 2 and 2,1,1 respectively for the ECG and GSR signals. Kernel filter sizes were set to 256. The epochs number used to train the model was 200. Table 3 shows the accuracy results that were obtained for the best model during training and testing.

In the experimentation process, two types of input data segments were configured for each ECG signal. The first was transformed to 200 R peaks, the second was normalized and segmented to 15.000 points. In the case of the ECGL signal, the results obtained for the arousal dimension were similar, although different processing techniques were applied, mainly in terms of dimensionality reduction.

Since the length of the segment of the ECGL signal is not significant, it is evident that the convolutional layers extract emotionally discriminatory features for the detection of arousal levels (low and high). With the valence dimension, better results were obtained when the ECG signals were integrated (ECGL and ECGR), than when using only the ECGL signal. In a similar way for the GSR signal, regarding the type of segments that were used during the experiment, it can be seen that with the length of 20 SCR peaks, the valence levels (low and high) have a better performance.

B. DCNN VS SHALLOW MACHINE LEARNING ALGORITHMS

In this section, we compare the performance results in the prediction of the affective states originally obtained by the authors of the dataset AMIGOS [10], with the algorithms

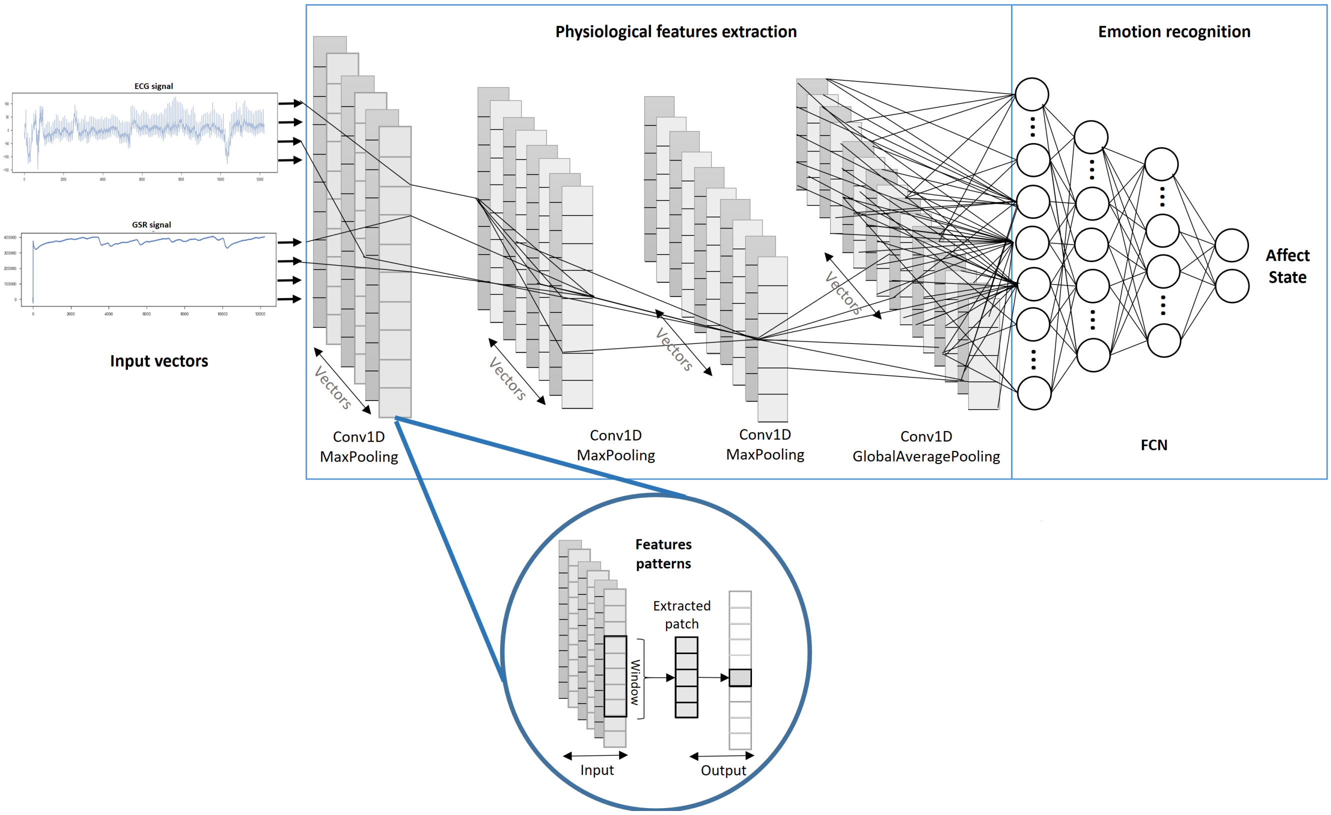

FIGURE 6. An schema of emotion recognition process based in deep learning.

TABLE 4. Performance comparison of DCNN with classical ML algorithms for emotion recognition based on ECG signals

ECGL Classifier	Arousal		Valence	
	Accuracy	F1-Score	Accuracy	F1-Score
Naive Bayes [10]		0.59		0.57
Nearest Neighbors	0.69	0.66	0.58	0.57
Linear Discriminant Analysis	0.72	0.63	0.67	0.65
Linear Support Vector	0.68	0.60	0.61	0.55
Multi-Layer Perceptron	0.68	0.59	0.61	0.51
AdaBoost	0.70	0.66	0.61	0.58
Random Forest	0.68	0.67	0.59	0.59
DCNN	0.81	0.76	0.71	0.68

TABLE 5. Performance comparison of DCNN with classical ML algorithms for emotion recognition based on GSR signals

GSR Classifier	Arousal		Valence	
	Accuracy	F1-Score	Accuracy	F1-Score
Naive Bayes [10]		0.54		0.53
Nearest Neighbors	0.68	0.64	0.69	0.68
Linear Discriminant Analysis	0.67	0.61	0.64	0.55
Linear Support Vector	0.69	0.56	0.68	0.55
Multi-Layer Perceptron	0.68	0.60	0.64	0.55
AdaBoost	0.64	0.59	0.66	0.65
Random Forest	0.58	0.58	0.64	0.64
DCNN	0.71	0.67	0.75	0.71

proposed in this study. Unlike CNN, the features of the physiological signals of the ECG and GSR were extracted manually, as explained in the section on extraction and selection of features. In most cases with machine algorithms, similar prediction results were obtained or a little higher than the previous study of [10].

Therefore, with DCNN a better performance in arousal recognition is achieved through the ECGL signals (see table 4), in contrast to the GSR signal that shows better results in valence prediction (see table 5).

Taking into account the physiological data limitation of the AMIGOS dataset, it was proposed to validate the deep learning model with the data of the EEG and ECG signals. Each signal was segmented and normalized by 10,000 points. For the training, 90% of the data was used and the rest for the testing, that is, 965 instances were assigned to validate the model. Due to the size of the dataset, the processing of each epoch lasted 550 seconds, that is, in comparison with the other tests, the computational effort was increased to generate a more robust model. The categorical recognition of emotions was evaluated from 4 classes (HALV, HAHV, LALV and LAHV).

The figure 7 shows the exponential behavior of the accuracy during the training and testing for the 500 epochs. Similarly, in the figure 8 the values of loss are displayed during the learning that is decreasing for each epoch. The confusion matrix is showing the results of prediction for the four classes of arousal and valence respectively (see figure 9).

The unification of the EEG and ECG signals ratifies the trend of the prediction results for arousal compared to valence because better results were obtained when the values of the labels are high. Possibly, by the subjective evaluation of the participants in the self-assessment of the emotion elicitation during the experiments of the short videos.

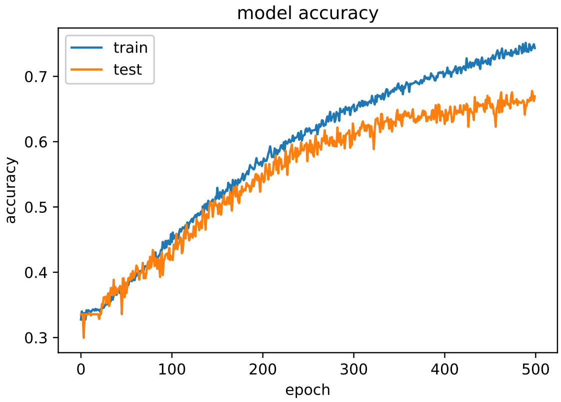

FIGURE 7. Accuracy result for DCNN model, using the EEG and ECG signals for the emotion recognition. [10], participant 1, video 10.

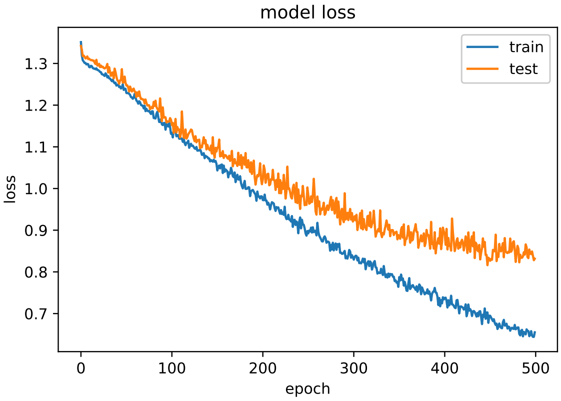

FIGURE 8. Loss result for DCNN model, using the EEG and ECG signals for the emotion recognition.

Table 6 shows the comparative results of studies similar to this research. In [38] describes a recognition method of arousal from ECG signals of various datasets represented in a common spectrum-temporal space to train a deep neural network. Derived from the results of the affect prediction with DECAF, it is can conclude that the stimulus is an indispensable factor to induce emotion. Also, other studies [34] [32] [36] for the arousal and valence detection used diverse EOG, EMG and, EEG signals from DEAP dataset, and they have obtained the same or better results than the reported in this work (see table 3). Unlike AMIGOS, DEAP dataset is one of the most explored datasets for emotional recognition, since different machine learning models have been developed for the automatic extraction of physiological features, features fusion and classification of the affective state. Hence, the performance outcome of emotional recognition models are subject to the number of physiological signals, the stimuli selection to elicit emotion, the reliability of the emotional assessment labels (self-evaluation) and the participants' number in the experiment.

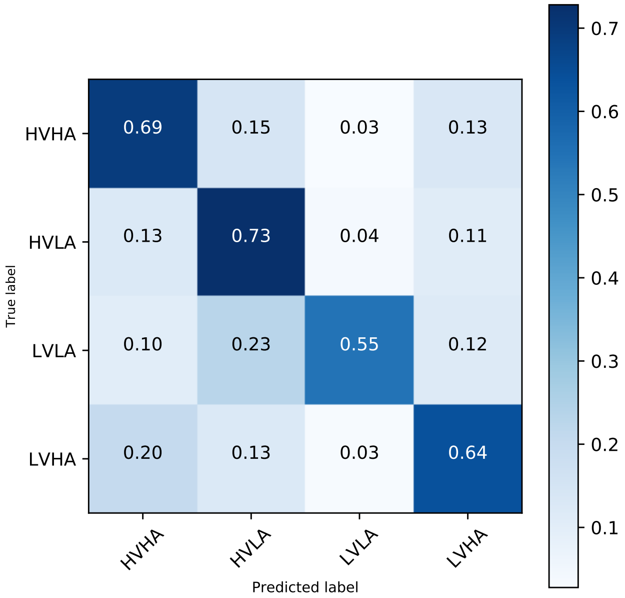

FIGURE 9. Normalized confusion matrix for the prediction of four classes.

TABLE 6. Accuracy comparison with other datasets

Research	Dataset	Arousal	Valence
DNN [38]	DEAP MAHNOB ASCERTAIN DECAF movie DECAF music	0.64 0.66 0.7 0.65 0.79	
SAE and DBN [34]	DEAP	0.73	0.78
MESAE [32]		0.84	0.83
CNN [36]		0.73	0.81
Our work (DCNN)	Amigos	0.76	0.75

VI. CONCLUSION

The convolutional networks in comparison with the classic algorithms of machine learning demonstrated a better performance in the emotion detection in physiological signals, in spite of being conceived for the objects recognition in images. The preprocessing of the peaks of the ECG and GSR signals as an entry vector to the CNN, made possible the identification of morphological features suitable for the affective state prediction. The experimental results validated the proposed methods and improved the performance in the emotion classification for the Dataset AMIGOS.

Physiological datasets with a large number of instances are optimal for the proposed experiments since these directly influence the emotion prediction, a the greater the number of instances, the more effective the model. Consequently, several annotations of arousal and valence must be recorded, since, when subjecting a participant to the stimulus of a short video, it can manifest different levels of emotion during of experiment.

The future work of this research consists in to apply these computational models to data acquired with wearable devices, for the recognition of emotion from physiological signals.

ACKNOWLEDGMENT

The authors wish to thank the team of researchers of the Telematic Applications and Services Group (GAST) of the Universidad Carlos III de Madrid, the people of the AMIGOS dataset project, and the anonymous reviewers by your valuable comments. This research project is financed by the Government of Colombia, Colciencias and the Governorate of Boyacá.

REFERENCES

[1] R. W. Picard, Affective Computing. MIT Press, 1997.

[2] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," IEEE Transactions on Affective Computing, vol. 1, no. 1, pp. 18–37, Jan 2010.

[3] Z. Kowalczuk and M. Czubenko, "Computational approaches to modeling artificial emotion – an overview of the proposed solutions," Frontiers in Robotics and AI, vol. 3, p. 21, 2016. [Online]. Available: http://journal.frontiersin.org/article/10.3389/frobt.2016.00021

[4] E.-H. Jang, B.-J. Park, M.-S. Park, S.-H. Kim, and J.-H. Sohn, "Analysis of physiological signals for recognition of boredom, pain, and surprise emotions," Journal of Physiological Anthropology, vol. 34, no. 1, 2015.

[5] A. Greco, G. Valenza, and E. P. Scilingo, Advances in Electrodermal Activity Processing with Applications for Mental Health: From Heuristic Methods to Convex Optimization, 1st ed. Springer Publishing Company, Incorporated, 2016.

[6] J. A. Rusell, "Chapter 4 - measures of emotion," in The Measurement of Emotions, R. Plutchik and H. Kellerman, Eds. Academic Press, 1989, pp. 83 – 111. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780125587044500104

[7] L. F. Barrett, "Discrete emotions or dimensions? the role of valence focus and arousal focus," Cognition and Emotion, vol. 12, no. 4, pp. 579–599, 1998.

[8] S. Koelstra, C. Muhl, M. Soleymani, J. S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis ;using physiological signals," IEEE Transactions on Affective Computing, vol. 3, no. 1, pp. 18–31, Jan 2012.

[9] M. Wiem and Z. Lachiri, "Emotion classification in arousal valence model using mahnob-hci database," International Journal of Advanced Computer Science and Applications, 2017.

[10] J. A. Miranda, M. Abadi, N. Sebe, and I. Patras, "Amigos: A dataset for affect, personality and mood research on individuals and groups (pdf)," Queen Mary University of London and Università Degli Studi of Trento, Tech. Rep., 2017. [Online]. Available: http://www.eecs.qmul.ac.uk/mmv/datasets/amigos/index.html

[11] R. Thayer, The biopsychology of mood and arousal. Oxford University Press, 1989.

[12] M. Yik, J. A. Russell, and J. H. Steiger, "A 12-point circumplex structure of core affect. emotion," Emotion, vol. 11, no. 4, pp. 705–731, 2011.

[13] M. Makikawa, N. Shiozawa, and S. Okada, "Chapter 7.1 - fundamentals of wearable sensors for the monitoring of physical and physiological changes in daily life," in Wearable Sensors. Oxford: Academic Press, 2014, pp. 517 – 541. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780124186620000076

[14] D. Buhalis and A. Amaranggana, "Smart tourism destinations enhancing tourism experience through personalisation of services," Information and Communication Technologies in Tourism 2015, 2015.

[15] J. Kim and D. R. Fesenmaier, "Measuring human senses and the touristic experience: Methods and applications," Analytics in Smart Tourism Design, pp. 47–63, 2017.

[16] W. T. Organization, A Practical Guide to Tourism Destination Management, R. Carter, Ed. UNWTO publications, 2007.

[17] A. Gulli and S. Pal, Deep Learning with Keras, V. Phadkay, Ed. Packt Publishing, 2017.

[18] L. Zhao, L. Yang, H. Shi, Y. Xia, F. Li, and C. Liu, "Evaluation of consistency of hrv indices change among different emotions," in 2017 Chinese Automation Congress (CAC), Oct 2017, pp. 4783–4786.

[19] D. Bach, G. Flandin, K. Friston, and R. Dolan, "Time-series analysis for rapid event-related skin conductance responses," Journal of Neuroscience Methods, 2009.

[20] M. Bradley and P. Lang, Handbook of Emotion Elicitation and Assessment. Oxford University Press, 2007, ch. The International Affective Picture System (IAPS) in the study of emotion and attention, pp. 29–46.

[21] J. J. Gross and R. W. Levenson, "Emotion elicitation using films," Cognition and Emotion, vol. 9, no. 1, pp. 87–108, 1995.

[22] M. K. Uhrig, N. Trautmann, U. Baumgartner, R. Treede, F. Henrich, W. Hiller, and S. Marschall, "Emotion elicitation: A comparison of pictures and films. frontiers in psychology," Frontiers in Psychology, vol. 7, no. 180, pp. 1664–1078, 2016.

[23] S. Katsigiannis and N. Ramzan, "Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices," IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 1, pp. 98–107, Jan 2018.

[24] R. Subramanian, J. Wache, M. K. Abadi, R. L. Vieriu, S. Winkler, and N. Sebe, "Ascertain: Emotion and personality recognition using commercial sensors," IEEE Transactions on Affective Computing, vol. 9, no. 2, pp. 147–160, April 2018.

[25] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, and N. Sebe, "Decaf: Meg-based multimodal database for decoding affective physiological responses," IEEE Transactions on Affective Computing, vol. 6, no. 3, pp. 209–222, July 2015.

[26] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," IEEE Transactions on Affective Computing, vol. 3, no. 1, pp. 42–55, Jan 2012.

[27] S. Raschka, Python Machine Learning. Packt Publishing Ltd., 2016.

[28] H. Ferdinando, T. Seppänen, and E. Alasaarela, "Comparing features from ecg pattern and hrv analysis for emotion recognition system," in 2016 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Oct 2016, pp. 1–6.

[29] M. Matsubara, O. Augereau, C. L. Sanches, and K. Kise, "Emotional arousal estimation while reading comics based on physiological signal analysis," in Proceedings of the 1st International Workshop on coMics ANalysis, Processing and Understanding, ser. MANPU '16. New York, NY, USA: ACM, 2016, pp. 7:1–7:4.

[30] K. Shirahama and M. Grzegorzek, Emotion Recognition Based on Physiological Sensor Data Using Codebook Approach. Cham: Springer International Publishing, 2016, pp. 27–39.

[31] H. P. Martinez, Y. Bengio, and G. N. Yannakakis, "Learning deep physiological models of affect," IEEE Computational Intelligence Magazine, vol. 8, no. 2, pp. 20–33, May 2013. [Online]. Available: https://ieeexplore.ieee.org/document/6496209/

[32] Z. Yin, M. Zhao, Y. Wang, J. Yang, and J. Zhang, "Recognition of emotions using multimodal physiological signals and an ensemble deep learning model," Computer Methods and Programs in Biomedicine, vol. 140, pp. 93 – 110, 2017.

[33] W. Liu, W.-L. Zheng, and B.-L. Lu, "Emotion recognition using multimodal deep learning," in Neural Information Processing, A. Hirose, S. Ozawa, K. Doya, K. Ikeda, M. Lee, and D. Liu, Eds. Cham: Springer International Publishing, 2016, pp. 521–529.

[34] P. Kawde and G. K. Verma, "Multimodal affect recognition in v-a-d space using deep learning," in 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon), Aug 2017, pp. 890–895.

[35] X. Li, D. Song, P. Zhang, G. Yu, Y. Hou, and B. Hu, "Emotion recognition from multi-channel eeg data through convolutional recurrent neural network," in 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Dec 2016, pp. 352–359.

[36] S. Tripathi, S. Acharya, R. Sharma, S. Mittal, and S. Bhattacharya, "Using deep and convolutional neural networks for accurate emotion classification on deap dataset," in Deployed Application Case Studies, 2017.

[37] L. Wenqian, C. Li, and S. Sun, "Deep convolutional neural network for emotion recognition using eeg and peripheral physiological signal," in ICIG 2017: Image and Graphics, 12 2017, pp. 385–394.

[38] M. Gjoreski, H. Gjoreski, M. Luštrek, and M. Gams, "Deep affect recognition from r-r intervals," in Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, ser. UbiComp '17. New York, NY, USA: ACM, 2017, pp. 754–762.

[39] A. Ortony, G. Clore, and A. Collins, The Cognitive Structure of Emotion. Cambridge University Press, 1988.

[40] M. Sznajder and M. Łukowska, "Python online and offline ecg qrs detector based on the pan-tomkins algorithm," Zenodo, Tech. Rep., 2018.

[41] M. S. Perez-Rosero, B. Rezaei, M. Akcakaya, and S. Ostadabbas, "Decoding emotional experiences through physiological signal processing," in 2017 IEEE International conference on acoustics, speech and signal processing (ICASSP). IEEE, 2017, pp. 881–885.

[42] G. Gabrieli, A. Azhari, and G. Esposito, "Pysiology: a python package for physiological feature extraction, special issue of smart innovation, systems and technologies," Python software fundation, Tech. Rep., 2018. [Online]. Available: https://pypi.org/project/pysiology/0.0.9.2/

[43] C. Godin, F. Prost-Boucle, A. Campagne, S. Charbonnier, S. Bonnet, and A. Vidal, "Selection of the most relevant physiological features for classifying emotion," in Proceedings of the 2Nd International Conference on Physiological Computing Systems, ser. PhyCS 2015. Portugal: SCITEPRESS - Science and Technology Publications, Lda, 2015, pp. 17–25. [Online]. Available: https://doi.org/10.5220/0005238600170025

[44] F. Shaffer and J. Ginsberg, "An overview of heart rate variability metrics and norms," Frontiers in Public Health, 2017.

[45] D. Makowski, "Neurokit," Dominique Makowski, Tech. Rep., 2017. [Online]. Available: https://neurokit.readthedocs.io/en/latest/

[46] S. Taylor, N. Jaques, W. Chen, S. Fedor, A. Sano, and R. Picard, "Automatic identification of artifacts in electrodermal activity data," in 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Aug 2015, pp. 1934–1937.

[47] G. Zaccone, R. Karim, and A. Menshawy, Deep Learning with TensorFlow. Explore neural networks with Python., S. Editing, Ed. Packt Publishing, 2017.

[48] B. Pyakillya, N. Kazachenko, and N. Mikhailovsky, "Deep learning for ecg classification," Journal of Physics Conference Series, pp. 1–5, 2017.

[49] Keras, "The python deep learning library," Keras, https://keras.io/, Tech. Rep., 2018.

[50] D. Li, J. Zhang, Q. Zhang, and X. Wei, "Classification of ecg signals based on 1d convolution neural network," in 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom), Oct 2017, pp. 1–6.

[51] F. Chollet, Deep Learning with Python. Manning publications Co., 2018.

[52] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.

LUZ SANTAMARIA-GRANADOS is currently pursuing the Ph.D. degree in telematics engineering with the University of Cauca, Popayán, Colombia. She is also a Professor with the Faculty of Systems Engineering, University of Santo Tomás, Colombia, and also a magister in communication and information sciences. Her research areas are recommender systems and wearable devices.

MARIO MUNOZ-ORGANERO received the M.Sc. degree in telecommunications engineering from the Polytechnic University of Catalonia, Barcelona, Spain, in 1996, and the Ph.D. degree in telecommunications engineering from the Carlos III University of Madrid, Madrid, Spain, in 2004. He is currently a Professor of Telematics Engineering at the Carlos III University of Madrid. He has authored or co-author more than 160 research publications (42 indexed in JCR journals) and participated in more that 20 European-funded and Spanish national-funded research projects. He has also more than four years of experience working for the telecommunications industry in companies such as Telefonica R&D and Lucent Technologies, both in Madrid, Spain. During a sabbatical leave for the academic course 2015-2016 he was a visiting professor at the Universities of Sheffield and Nottingham Trent in the UK. His main research interest is on applied machine learning for wearable sensor data.

GUSTAVO RAMIREZ-GONZALEZ received the B.S. degree in electronic and telecommunications engineering from the University of Cauca, Colombia, in 2001, the M.S. degree in telematics engineering from the University of Cauca, and the Ph.D. degree in telematics engineering from the Universidad Carlos III de Madrid, Spain, in 2010. He is currently a Professor and a Researcher at the Department of Telematics, University of Cauca. He has participated in national and international projects in Colombia and Spain. His research interests include image processing, secure communication, machine learning, and IoT. He has published several research papers in reputed journals and served as a guest editor for several special issues at many journals.