

Article

Affective State Assistant for Helping Users with Cognition Disabilities Using Neural Networks

Luis Muñoz-Saavedra ^{1,*}, Francisco Luna-Perejón ¹, Javier Civit-Masot ¹,
Lourdes Miró-Amarante ^{1,2}, Anton Civit ^{1,2} and Manuel Domínguez-Morales ^{1,2}

¹ Robotics and Technology of Computers Lab, University of Seville, ETSII, 41004 Sevilla, Spain;
fralunper@atc.us.es (F.L.-P.); jcivit@atc.us.es (J.C.-M.); lmiro@us.es (L.M.-A.); civit@us.es (A.C.);
mjdominguez@us.es (M.D.-M.)

² I3US: Research Institute of Computer Engineering, University of Seville, 41004 Sevilla, Spain

* Correspondence: luimunsa@atc.us.es

Received: 28 September 2020; Accepted: 31 October 2020; Published: 3 November 2020

Abstract: Non-verbal communication is essential in the communication process. This means that its lack can cause misinterpretations of the message that the sender tries to transmit to the receiver. With the rise of video calls, it seems that this problem has been partially solved. However, people with cognitive disorders such as those with some kind of Autism Spectrum Disorder (ASD) are unable to interpret non-verbal communication neither live nor by video call. This work analyzes the relationship between some physiological measures (EEG, ECG, and GSR) and the affective state of the user. To do that, some public datasets are evaluated and used for a multiple Deep Learning (DL) system. Each physiological signal is pre-processed using a feature extraction process after a frequency study with the Discrete Wavelet Transform (DWT), and those coefficients are used as inputs for a single DL classifier focused on that signal. These multiple classifiers (one for each signal) are evaluated independently and their outputs are combined in order to optimize the results and obtain additional information about the most reliable signals for classifying the affective states into three levels: low, middle, and high. The full system is carefully detailed and tested, obtaining promising results (more than 95% accuracy) that demonstrate its viability.

Keywords: deep learning; feature extraction; discrete wavelet transform; affective state; cognitive disorders; autism; ASD; EEG; ECG; GSR

1. Introduction

High-context communication relies heavily on sensitivity to non-verbal behaviors and environmental cues to decipher meaning, while low-context exchanges are more verbally explicit, with little reliance on tacit or nuanced [1].

In interpersonal interactions, both emotional and cognitive processes are included. As emotions and related phenomena such as desires, moods, and feelings can be revealed through nonverbal behavior, as well as spoken with words, nonverbal behavior has a significant role in those interactions [2].

Thus, nonverbal behavior includes a variety of communicative behaviors that have no linguistic content. These include (but are not limited to): facial expressiveness, smiles, eye contact, head nodding, hand gestures, postural positions (open or closed body posture and forward to backward body lean); paralinguistic speech characteristics such as speech speed, volume, pitch, pauses, and lack of fluency of speech; and dialogic behaviors such as interruptions. It is widely recognized that nonverbal behavior

conveys affective and emotional information, although it also has other functions (such as regulating the turn of the conversation). As examples, a frown may convey disapproval or a smile may convey approval or agreement. A blank expression can also convey an emotional message to a listener, such as indifference, boredom, or rejection. Nonverbal behaviors often (but not always) accompany words and therefore give words meaning in context (for example, by amplifying or contradicting the verbal message). Thus, the interpretation of a verbal message of agreement (like “That’s okay”) can be interpreted differently depending on whether the statement is accompanied by a frown or a smile or a blank expression [3].

However, certain mental and physical disabilities can provoke that non-verbal communication cannot be captured or understood by the receiver. This is the case of people with Autism Spectrum Disorders (ASD).

Several investigations reaffirm the importance of emotional aspects in social integration and, consequently, in the quality of life in people with ASD [4]. However, these people present a deficit in the process of understanding non-verbal language; and, as many of these expressions are related to the emotional state of the transmitter, consequently, they present a deficit when understanding emotions. Various research works like [5,6] proved these affirmations.

Thus, in order to help in these cognitive deficiencies, our objective is to design an affective states recognition aid system, but, before that, we must delve into how to distinguish these moods.

Human emotions represent physiological phenomena that include feelings, memories, evaluations, unconscious reactions, body gestures, vocalization, postural orientations, etc [7]. Emotional theories have evolved throughout time, distinguishing between those that defend that emotions come from the perception of physiological states (firstly suggested by James and Lange in 1880 [8,9] and known as ‘somatic theories’), and those that consider emotions as cognitive evaluations [10] that are known as ‘cognitive theories’. All these theories relate emotions with particular locations of the brain: cognitive theories relate emotions to the neocortex and somatic theories to the limbic system.

Actually, a large amount of authors consider both theories valid, since the mind is not cognitive or emotional individually, but rather a combination of both. Both emotion and cognition are unconscious processes that are transformed into conscious experiences [11]. Furthermore, emotions are related to unconscious reactions in various parts of the body (not only in brain activity); this implies that, if we were able to detect and discretize these unconscious reactions, we could be able to infer about the emotional state of the person. However, to get to that point, it is necessary to force certain emotions in the user in order to be able to detect the physiological variations.

There are different techniques to induce emotions either through images and sounds [12,13], expressive behaviors [14,15], social interactions [16,17] and music [18] among others [19]. In these studies, when working with multiple users, it is important that the stimuli can be reproduced multiple times without change; this is why the use of expressive behaviors and social interactions are ruled out. The most efficient and widely used ways are image sampling and video playback. In addition, the use of videos implies the inclusion of auditory elements, which enhance the emotional state of the user. Thus, we are going to focus on video playback [20].

Related to emotions classification, two theories are distinguished: the ‘categorical approach’ and the ‘dimensional approach’. The categorical approach is based on the amount and type of emotions such as fear, anger, sadness, and joy. However, the dimensional approach describes emotions as points within a multidimensional space, which can be the activation space and valence [12], and the activation space and approach-distance [21].

The ‘activation’ value represents the intensity of the emotion. The ‘valence’ determines whether the emotion is positive or negative. On the other hand, the approach/away determines whether the emotion makes the person move towards the cause of the emotion or away from said cause. The affective state’s

classification according to those metrics is deeply evaluated by Posner et al. [22] and is summarized in Figure 1. As can be observed, the emotional state depends on the Valence and Arousal values.

Focusing our attention on the dimensional approach, it is very important to remark that, although the emotions' study requires a multidimensional approach [23] such as questionnaires to determine the emotion perceived by the user, the physiological responses and the analysis of the user's behavior (analysis of physiological signals) are the best approximation to an emotional response. This is because the questionnaires have several limitations: the users could not be able to understand or express what they feel [16], or their answer may be influenced by questionnaires from previous experiences [24]. Thus, the analysis of user behavior using questionnaires does not guarantee the same emotional response in different contexts or in different people, so it is difficult to obtain a behavior pattern for different emotions.

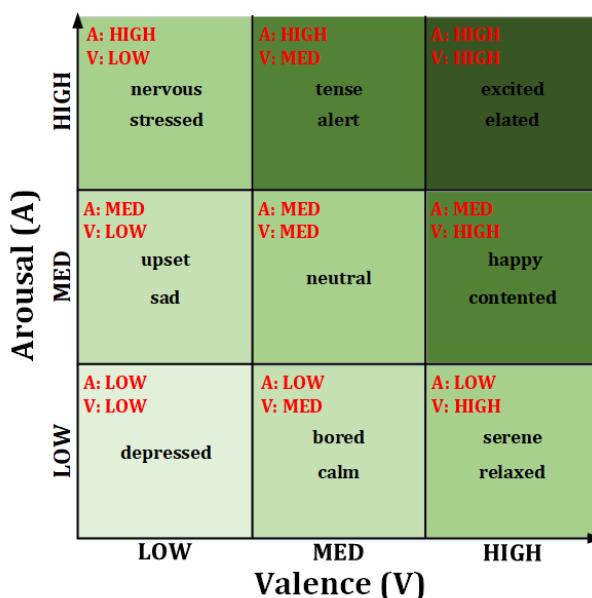


Figure 1. Classification of the different affective states according to Arousal and Valence. Information obtained from Posner et al. [22].

Focusing on using tangible measures that allows us to obtain a more precise measurement with less variability, there are several physiological signals related to the unconscious reactions from the user's body that can be logged and studied: facial's muscles activity, heart activity, skin conductivity, brain activity, among others. These signals will be described deeply in the next section.

Moreover, extracting useful information from physiological signals can be a difficult task as many elements are involved. Some works extract useful features from the data [25], others use thresholds in decision trees [26,27] and others convert the signal's temporal representation to images and apply computer vision techniques [28,29].

On the other hand, the study of medical signals and images has experienced a great progress with the inclusion of Machine Learning (ML) systems capable of automatically identifying and extracting the relevant characteristics to make a correct diagnosis, obtaining better results than those obtained by classical diagnostic systems [30,31].

These techniques have been mainly applied to imaging systems using Convolutional Neural Networks (CNN), which typically use classical Artificial Neural Networks as a final step for providing one classification output for each sample input. However, when working with physiological signals,

temporal dimension is a very relevant source of information and, for this reason, in many cases, it is necessary to analyze a time window to obtain a correct classification.

In order to analyze these time windows, two main approaches are commonly used in Deep Learning (DL): using Recurrent Neural Networks (RNN) with the original data obtained from the physiological sensors, or using feature extraction techniques (before training the neural network) to work only with the most useful information.

Previous works obtained very good results using RNNs with analog and time-dependent sensors (like accelerometers) [32,33], but working with physiological sensors like electrocardiogram (ECG) is a harder task since these signals have a lot of noise (in high and low frequencies), they are not periodic and, in some cases like galvanic skin response (GSR), an incremental peak does not lead to a subsequent decline. This is why it is common to extract information from the frequency components of the physiological signals before using a neural network system.

Therefore, after a detailed introduction, the main objective of this work consists in the design, implementation, and testing of a Deep Learning (DL) system for user's affective state classification, in order to help in the emotional learning process of users with cognitive disabilities. Each physiological signal will be pre-processed by extracting its frequency features, and the obtained coefficients will be used as inputs of a NN system. Each signal will be tested to classify the main affective states (Arousal and Valence) in an independent way. Results will be analyzed, and the signals with better results will be combined in order to obtain a global affective state classification system.

The main novelty of this work is this system's architecture itself, as it allows for evaluating each signal independently, allowing that each network model can be adapted to the particular characteristics of its signal. In this way, two great benefits are obtained: (1) we can observe how each signal affects affective state classification; and, thanks to these partial evaluations, (2) the overall system designed by combining the independent ones obtains better results than others trained by brute force (as they do not take into account the specific characteristics of each type of signal).

The rest of the paper is divided in the following way: first, in the 'Materials and Methods' section, the different physiological signals and datasets analyzed for this work are presented, as well as the system's architecture, including the different implemented subsystems. Next, the results obtained after the training process and the evaluation of each model are detailed and explained in the Results and Discussion section. Finally, conclusions are presented.

2. Materials and Methods

In this section, the classifier architecture used for the task previously explained is described deeply. In order to do that, the information provided as output of the model, necessary for the classification the different affective states, needs to be described and, moreover, the physiological signals that provide this data need to be presented too.

Thus, first, the physiological signals used for obtaining the input data are detailed. Following this, the datasets used in this work are presented and compared, focusing on the signals logged by each one. In addition, finally, the proposed classifier to identify affective states is described.

2.1. Physiological Signals

Speaking about collecting data from physiological signals implies focusing on those medical instruments designed to be able to appreciate those signals. Each instrument can only collect information from a specific type of physiological signal. However, some of these instruments can be placed in different positions on the user's body; thus, the information obtained can be interpreted in different ways (depending on where

the instrument is placed). Next, the four most used physiological signals in emotional studies and their associated instruments are presented.

2.1.1. Brain's Electrical Activity

The use of brain's evoked potentials as a measure of the response to visual stimuli has been used in several works [34,35]. Positive potential modulations are obtained when visualizing both pleasant and unpleasant stimuli. Moreover, a positive slow wave has also been detected that is maintained once the stimulus has ended [35].

The main problem about using these signals is its low amplitude as well as its surrounding noise, which makes it considerably difficult to obtain the emotional response through the detection of evoked potentials patterns. Furthermore, depending on the helmet used to capture these potentials, it can be a very invasive technique for the user. Despite this, the asymmetry of the Electroencephalograph (EEG) allows us to determine the valence against a stimulus or the approach-distance [36].

Brain's electrical activity is measured using an EEG, obtaining a representation of all the electrodes' activity over time (electroencephalography). As detailed before, this physiological signal can help to determine the emotional valence.

2.1.2. Heart's Electrical Activity

A wide variety of studies demonstrate that heart rate variations are related with emotions. Graham [37] found that viewing images for 6 s causes a pattern in heart rate variation. This pattern consisted of three phases: initial deceleration, followed by an acceleration component and finally by a last deceleration component [12].

Some emotions such as fear, anger, or joy have been evaluated using cardiac activity [38]; for example, cardiac frequency accelerates in anger and fear [39].

However, it presents several problems related to the duration and repetition of the stimulus. On the one hand, short stimuli (around half a second) do not modify heart rate [40]. On the other hand, the initial deceleration decreases when the same stimulus is displayed repeatedly during an experiment [41].

Heart's electrical activity is measured using an Electrocardiograph (ECG), obtaining a representation over time (electrocardiography) from which the heart rate can be obtained. As detailed before, this physiological signal has been demonstrated to be related with emotions, and it is important to evaluate time windows in order to detect frequency accelerations and decelerations.

2.1.3. Muscles' Electrical Activity

The muscles most directly related to the user's emotions are those in the face: the mood is expressed through grimaces or facial gestures caused by these muscles. Facial muscles' electrical activity is useful for performing studies of emotions in which emotional arousal is so low that it is insensitive to facial gestures [42]. There are two muscles that can be used for emotional assessment: the major zygomatic, related to smiling, and the superciliary corrugator related to the frown gesture.

Schwartz [43] was the first researcher to link both muscles with emotions, realizing that unpleasant images produced greater activity of the superciliary corrugator, and those pleasant images caused greater activity in the major zygomatic, relating the activity of both muscles with the valence of emotions.

However, the relationship of both muscles to valence differs considerably: some researchers report greater activity in the superciliary corrugator than in the major zygomatic [44]. Furthermore, the relationship with the valence of the electrical signal of these muscles appears to be linear in the case of the superciliary corrugator, but has a "J" shape in the case of the zygomatic [45]. There is less consensus on whether this electrical activity is related to activation level. Some authors suggest that, at least in the case

of the corrugator, this relationship exists, while others suggest the opposite [46]. From the point of view of Cacioppo [47], by recording the facial muscles' electrical activity, both valence and activation can be obtained. Although corrugated and zygomatic electrical activities have been widely used, other muscles such as the orbicularis have also been used to measure valence-positive emotions [48].

Muscles' electrical activity is measured using an Electromyograph (EMG), obtaining a representation over time (electromyography) of the electrical activity of the muscles near the placed electrodes. Thus, in order to detect the activity of the facial muscles, those electrodes need to be placed near the corrugator and the zygomatic muscles. However, as detailed before, there are several discrepancies about the information provided by these muscles.

2.1.4. Dermal Electrical Activity

Changes in skin conductivity are strongly related to variations in the level of activation [49,50]. An increase in the activation level causes an increase in the level of skin conductivity, generating a positive potential that begins about 400 ms after the stimulus [12] and lasts between 400 ms and 700 ms [35,45].

The skin conductivity is also known as 'galvanic skin response' (GSR) or 'electro-dermal activity' (EDA). This GSR signal has two components, a tonic component and a phasic component. On the one hand, the tonic component is a low frequency signal that is associated with the baseline (trend) of the signal and undergoes slight variations over time. On the other hand, the phasic component corresponds to rapid and punctual variations, and is directly associated with the response to a stimulus.

Thus, dermal electrical activity (GSR or EDA) is measured using typical electrodes located in places where the density of the sweat glands is higher (forehead, cheeks, palms, fingers, and the soles of the feet), obtaining a representation over time of the skin's conductivity (varying because of user's sweating). This measure, although simple, has been demonstrated to reflect the affective state of the user in a very reliably way.

2.2. Dataset

In this subsection, several datasets are presented and analyzed in order to find the most adequate set for training and testing our classifier. Several information is obtained from each one: number of people, sensors used (physiological signals recorded), logging frequency and emotional states labeled. As will be observed next, most datasets use similar sensors (EEG, ECG and GSR) and obtain similar emotional states.

For the datasets searching process, we took into account those who met the following criteria: freely accessible, output labeled with emotional states, videos used as stimuli for the test subjects, and information recorded from various sensors. According to these criteria, five datasets were found: AMIGOS, ASCERTAIN [51], DEAP, DREAMER [52], and HR-EEG4EMO.

The main advantage about working with public and labelled datasets in our case is that the video clips that serve as stimuli have been carefully selected by psychologists to provoke a specific affective state. Moreover, in order to check that the labels are correct, at the end of each data collection, a survey is given to the users so that they can detail their affective state at that moment and how they have felt while viewing each video. Thus, we can assume that those labels are correct and work with that.

First, a brief summary about the used datasets is presented; and, after that, a summarizing table with the most relevant information is shown.

2.2.1. ASCERTAIN

This is a multi-modal dataset for measuring the affect, personality and affect recognition using commercial neuro-physiological sensors. This dataset has the information obtained from 58 participant, 38 male and 20 female, with a mean age of 30 years. The data was collected using EEG, ECG, GSR and

Facial Gestures while the participants were watching 36 short musical videos. Data stored were labeled using Arousal, Valence, Dominance, Liking, and Familiarity. The ECG was recorded at 256 Hz, the GSR at 128 Hz and the EEG only at 32 Hz.

2.2.2. DREAMER

This is a multi-modal database consisting of EEG and ECG signals recorded during affect elicitation by means of audio-visual stimuli. Signals from 23 participants watching 18 video-clips were recorded along with the participants' self-assessment of their affective state after each stimuli, in terms of valence, arousal, and dominance. All the signals were captured using portable, wearable, wireless, low-cost and off-the-shelf equipment at 256 Hz for the ECG and 128 Hz for EEG.

2.2.3. Summary

The most relevant information about the analyzed datasets is summarized in Table 1.

As can be seen, all datasets have several similarities. On the one hand, they used a good amount of participants and several sensors (two of them are common to all: EEG and ECG, although GSR is present in almost all of them); so the sensors we will use are EEG, ECG, and GSR. On the other hand, regarding the emotional states labeled and the studies done in previous works and detailed in the introduction, the main affective states are "Arousal" and "Valence".

Table 1. Datasets summary (superscripts indicate logging frequency for the used datasets).

Dataset	#Participants	Physiological Signals	Logging Frequency (Hz)	Tagged Labels
AMIGOS	40	EEG, ECG, GSR	128 256	Valence, Arousal, Dominance, Liking and Familiarity
ASCERTAIN [51]	58	EEG ¹ , ECG ² , GSR ³ and Facial EMG	¹ 32 ³ 128 ² 256	Valence, Arousal, Engagement, Liking and Familiarity
DEAP	32	EEG, ECG, GSR EOG and Facial EMG	128	Valence, Arousal, Dominance, Liking and Familiarity
DREAMER [52]	23	EEG ⁴ and ECG ⁵	⁴ 128 ⁵ 256	Valence, Arousal and Dominance
HR-EEG4EMO	40	EEG, ECG, GSR, SPO ₂ , breath and heart rate	100 1000	Valence

Thus, according to this, HR-EEG4EMO cannot be used because it only labeled "Valence". Another problem is that DREAMER does not have information about GSR, but its information about EEG and ECG can be combined with the others. Thus, according to these filters, our database should be composed by the datasets AMIGOS, ASCERTAIN, DEAP, and DREAMER. However, after trying several times to contact the creators of the datasets AMIGOS and DEAP, and after waiting several months for an answer, we had to discard these datasets due to not having access to their data. Thus, our database is composed by the datasets ASCERTAIN [51] and DREAMER [52]. However, four adaptations are needed so that the information given by both datasets are completely the same:

- Sensors: only EEG, ECG, and GSR data are used. Facial EMG from ASCERTAIN is discarded, and GSR data are not provided by DREAMER (using only the GSR information from ASCERTAIN).
- Data frequency: based on the different frequencies used, it will be 32 Hz for EEG, 128 Hz for GSR, and 256 Hz for ECG. These are the lowest frequencies from each sensor of both datasets. The others will be down-sampled in order to balance them.
- Data amount: In order to perform a balanced training, which gives equal weight to the information from all the datasets, we must use a similar amount of information from each of the datasets. In this

way, there is no type of bias when training. Thus, the number of samples used from each of the datasets will be restricted by the dataset that has the least number of samples. In the case of GSR, the full amount of data provided by ASCERTAIN will be used.

- Labels: both datasets share the main two labels (“Valence” and “Arousal”). Due to the studies done before, with the information of both affective states is enough to extract information of the emotional state of the user.
- EEG: eight channels are used in order to design our own device in a near future with OpenBCI platform. These channels are (according to the 10–20 international placement system) Fp1, Fp2, C3, C4, T5, T6, O1, and O2. The information from other channels is discarded.

According to the number of classes used for training and testing each affective state, some previous considerations are needed to build the dataset used in our classifier:

- Most previous works divide each affective state in two classes (this fact will be observed in the final comparison made in the “Results and Discussion” section). We think that this is not enough, as we need more classes in order to discretize correctly the different states.
- Among the works that use three classes, the central zone (‘neutral’) is given greater weight than the extremes (‘low’ and ‘high’); i.e., in one of the works a division of 25% (low), 50% (medium), and 25% (high) is used. We thought that this division was not realistic and, although these values depend on the subject, we made an equitable division between those classes: we use the lower 30% values for ‘low’, 40% for ‘medium’, and the upper 30% values for ‘high’.

2.3. Affective State Classifier

The proposed system is based on several neural networks (one for each sensor) trained for each of the affective states (Arousal and Valence), obtaining six neural networks that, after evaluating, are combined in order to improve the classification results. A complete vision of the proposed system is shown in Figure 2.

For all the neural networks, we use a 500-epoch training, a learning rate of 0.001, an Adam optimizer and the sigmoid function as activation function for all layers except the output (where a softmax is used). The loss function used is the Mean Square Error.

In order to provide veracity to the results obtained, and due to their variability depending on the distribution between the training and testing subsets, each test has been performed 10 times. The information presented in the “Results and Discussion” section correspond to the mean of those ten repetitions.

In the following sections, this process will be specified step by step, detailing from the extraction of useful information from the sensors to the design of the experiments carried out.

2.3.1. Pre-Processing

First, the information from each dataset is filtered. Creators of the ASCERTAIN dataset specify in one external file the confidence of the information obtained from each video-clip of each user, providing information about the noise level and the problems detected. In order to avoid classifier malfunctioning, all the data streams with high noise levels or labeled with low confidence by the creators have been discarded.

After that, the information from both datasets is combined in an equitable proportion for each sensor, obtaining three separate datasets: one for EEG (with eight channels), one for ECG (with two channels), and one for GSR (one channel). For ECG and GSR, the amount of data are down-sampled to 64 Hz, using a mean window filter of size 4 and 2, respectively. The goal for this filter is to obtain a smoothed signal, reducing the noise of both data streams. However, it cannot be applied to EEG signal since it is sampled

at 32 Hz in an ASCERTAIN dataset, and reducing the frequency of this signal can cause irreversible loss of information.

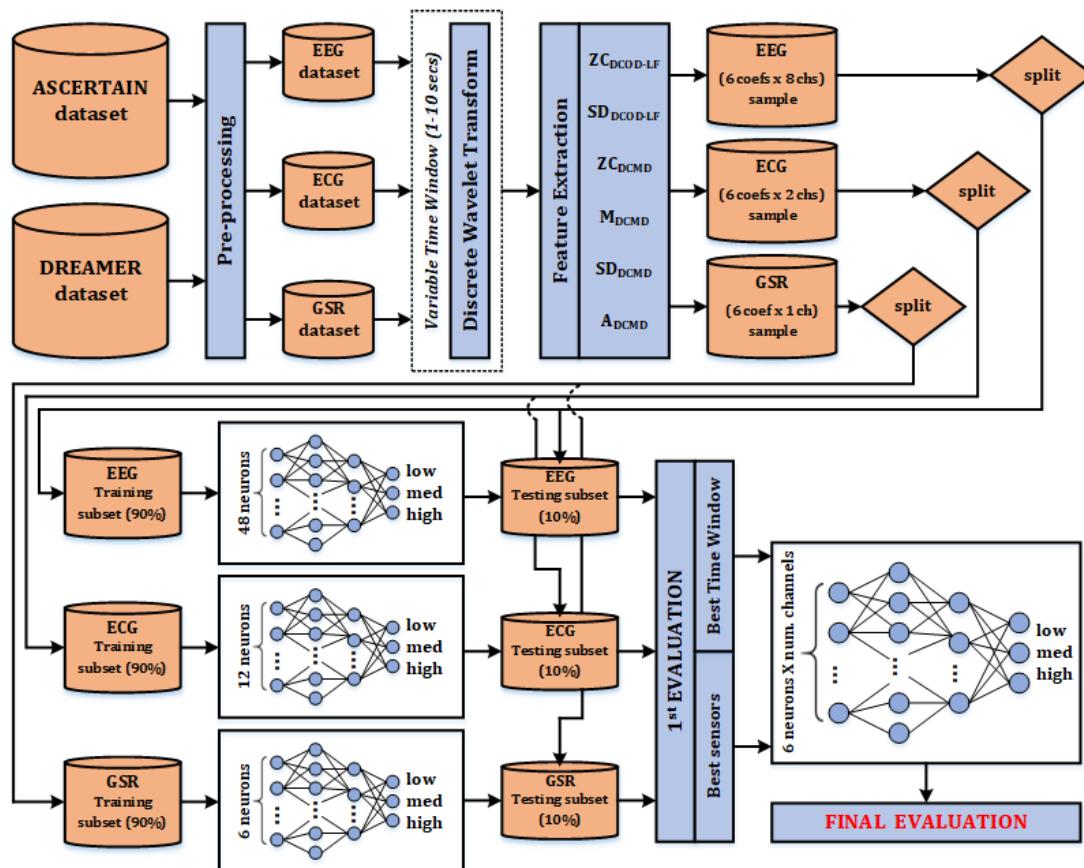


Figure 2. Full system implemented in this work.

Thus, after the pre-processing step, three datasets are obtained: EEG dataset (sampled at 32 Hz), ECG dataset, and GSR dataset (both sampled at 64 Hz).

For the labeled information about Arousal and Valence, each dataset discretizes its values into a rank between 0 and N-1 (using a N-level resolution that vary according to the dataset). In almost all the works about emotions classification using Arousal and Valence, only two levels are taken into account for each affective state (low/deactivation and high/activation) but, in this case, we discretize both affective states in three levels: LOW, MEDIUM and HIGH. In the “Results and Discussion” section, the results obtained in this work will be compared with other works, and those differences will be detailed.

2.3.2. Feature Extraction

In order to train the system, and according to the explanation done in the Introduction section, a feature extraction process will be applied to each dataset independently.

This process will consist of representing the information of each signal in time windows (whose width will be studied for the implementation of the neural networks), in order to extract the frequency characteristics of these windows. With the information represented in the frequency space, the coefficients obtained can be selected to obtain the main features of each time window. These features will be used as inputs of the neural networks.

The classical way of extracting the frequency components of a digitized periodic analog signal is their decomposition using the Discrete Fourier Transform (DFT). However, with physiological signals, this type of transformation is not efficient, since the components of these signals vary their periodicity continuously. For these cases, the Discrete Wavelet Transform (DWT) is used.

After applying a third-order Daubechies wavelet (db3) to the time window, two coefficient sets are obtained: the approximation coefficients (high frequencies) and the detailed coefficients (low frequencies). As the only high frequencies in a physiological signal come from the device noise or the environment alternating current, these are removed for the final coefficients (so the approximation coefficients are not taken into account). The obtained set of coefficients (detailed coefficients of the first-level decomposition) will be denoted as “Detailed Coefficients of the Original Decomposition” or DCOD. Moreover, to reduce spurious noise from the signal, the lowest frequencies can be removed: to do this, the maximum useful level of decomposition for the given input data length is calculated in order to extract the detailed components of this level (which represents the lowest frequencies of the signal) and those coefficients (denoted as “Detailed Coefficients of the Maximum Decomposition” or DCMD) are removed from the DCOD. Thus, after this processing, the coefficients obtained are separated and filtered, erasing the highest and lowest frequencies from them; and the resulting set of coefficients will be denoted as “Detailed Coefficients of the Original Decomposition without Lowest Frequencies” or DCOD-LF. However, DCMD are stored too as some basic information (like signal offset) is given by them.

With both processed sets of coefficients (DCOD-LF and DCMD), the features used as input of the neural network can be extracted. We have followed the studies done by JeeEun Lee and Sun K. Yoo [53,54], where different features are analyzed to extract information from physiological signals that have been previously processed by DWT. Following those studies, in our work, we have used these features: zero-crossing of DCOD-LF coefficients ($ZC_{DCOD-LF}$), standard deviation of DCOD-LF coefficients ($SD_{DCOD-LF}$), zero-crossing of DCMD coefficients (ZC_{DCMD}), mean of DCMD coefficients (M_{DCMD}), standard deviation of DCMD coefficients (SD_{DCMD}) and amplitude of DCMD coefficients (A_{DCMD}).

2.3.3. Single Neural Networks

The neural network architectures used in this work are based on a classical Multilayer Perceptron (MLP) network, using an input layer (whose width depends on the number of features extracted from the signal), two hidden layers, and an output layer (with three neurons according to the different levels used for the affective state classification: low, medium and high). Softmax activation function is applied to the output layers so it converts the resulting vector to a vector of categorical probabilities. We choose the category with more probability as the output to obtain a unique classification.

Thus, according to the physiological signal, the single MLP neural network used in the first step is:

- EEG: input layer (6 features \times 8 channels = 48 nodes), hidden layer 1 (96 nodes), hidden layer 2 (24 nodes), output layer (three nodes).
- ECG: input layer (6 features \times 2 channels = 12 nodes), hidden layer 1 (24 nodes), hidden layer 2 (six nodes), output layer (three nodes).
- GSR: input layer (6 features \times 1 channel = 6 nodes), hidden layer 1 (12 nodes), hidden layer 2 (six nodes), output layer (three nodes).

Each network is trained for Valence and Arousal independently, using 90% data of each dataset for training and 10% data for testing purposes. Moreover, all these neural networks are trained and tested using different sizes of the data time window varying from 1 second to 10 s. The results are analyzed and the best time window is selected.

2.3.4. Full Classification System

The work done in this investigation is not limited to evaluate the classification confidence for each affective state using each physiological sensor independently. After the single evaluation detailed before, the two physiological signals with the best results are selected and their coefficients are combined in order to obtain a full classifier that mix the information from both of them.

Thus, a final MLP neural network (that represents the final classification system) is trained (90% of the data) and tested (10% of the data) using the coefficients from both physiological signals, and, finally, the results for Arousal and Valence are detailed deeply. The architecture of this neural network is similar to the ones detailed before, using one input layer, two hidden layers, and one output layer. The width of the input and the hidden layers depends on the number of coefficients (that depends on the signals selected)—so, in order not to anticipate the results of the first part of this work, they will be indicated in the “Results and Discussion” section.

3. Results and Discussion

First, the results obtained after training each single neural network are shown. As detailed before, the time window evaluated for each physiological signal and affective state vary between 1 s and 10 s in order to determine the best time window for the full classification system.

After presenting the single results, the two best single systems will be combined and tested, obtaining the final results of this work. In this case, the size of the time window will not be evaluated, since it will be chosen from the single neural networks evaluation (4 s).

Finally, the results obtained in the full classification system will be compared with the results obtained from similar works in the last few years.

3.1. Single Neural Networks Results

As detailed before, each neural network will be tested using time windows from 1 to 10 s width.

The results of training GSR data with Arousal and Valence are detailed in Tables 2 and 3, respectively.

According to GSR neural network, a $\approx 83\%$ accuracy is obtained for Arousal with a loss less than 0.10, and a $\approx 76\%$ accuracy is obtained for Valence with a loss less than 0.13. Thus, GSR obtains better results when classifying Arousal. Moreover, the best time windows for both classifications are 4 s and 5 s width (remarked in red). It is important to remark that accuracy and error difference between training and testing is almost nil, so the network would work right with new data.

The same process is done for ECG dataset, and the results after training that network are detailed in Tables 4 and 5.

According to ECG classification results, a $\approx 81\%$ accuracy is obtained for Arousal with a loss around 0.10 and a $\approx 80\%$ accuracy is obtained for Valence with a loss around 0.10 too. Thus, ECG obtains better results when classifying Arousal, but those results are a little bit worse than the ones obtained with GSR. However, the results obtained for Valence are significantly better than the ones obtained by GSR. In this case, the best time windows for both classifications are 3 s and 4 s width.

In addition, finally, the same process is repeated for EEG neural network and the results obtained are detailed in Tables 6 and 7.

Table 2. Results obtained with GSR network classifying Arousal (best time windows are marked).

Time Window (s)	Train			Test		
	#Samples	Accuracy (%)	Loss	#Samples	Accuracy (%)	Loss
1	54,816	79.95	0.1080	5486	78.80	0.1083
2	27,272	80.28	0.1056	2730	79.65	0.1071
3	18,040	81.34	0.1014	1805	81.20	0.1033
<u>4</u>	<u>13,488</u>	<u>83.48</u>	<u>0.0964</u>	<u>1349</u>	<u>82.59</u>	<u>0.0987</u>
<u>5</u>	<u>10,704</u>	<u>83.06</u>	<u>0.0965</u>	<u>1073</u>	<u>82.93</u>	<u>0.0983</u>
6	8872	82.12	0.0988	889	81.14	0.1001
7	7576	81.09	0.1035	759	80.36	0.1078
8	6608	80.27	0.1083	661	79.69	0.1107
9	5844	79.84	0.1091	584	79.33	0.1134
10	5212	79.10	0.1110	521	79.02	0.1131

Table 3. Results obtained with GSR network classifying Valence (best time windows are marked).

Time Window (s)	Train			Test		
	#Samples	Accuracy (%)	Loss	#Samples	Accuracy (%)	Loss
1	54,816	74.55	0.1262	5486	74.50	0.1282
2	27,272	74.83	0.1257	2730	74.68	0.1263
3	18,040	75.05	0.1249	1805	74.91	0.1255
<u>4</u>	<u>13,488</u>	<u>75.88</u>	<u>0.1245</u>	<u>1349</u>	<u>75.60</u>	<u>0.1233</u>
<u>5</u>	<u>10,704</u>	<u>76.22</u>	<u>0.1241</u>	<u>1073</u>	<u>75.83</u>	<u>0.1225</u>
6	8872	75.53	0.1240	889	74.94	0.1241
7	7576	75.08	0.1252	759	74.93	0.1256
8	6608	74.80	0.1259	661	74.29	0.1267
9	5844	74.51	0.1261	584	73.06	0.1297
10	5212	74.30	0.1273	521	73.02	0.1346

Table 4. Results obtained with ECG network classifying Arousal (best time windows are marked).

Time Window (s)	Train			Test		
	#Samples	Accuracy (%)	Loss	#Samples	Accuracy (%)	Loss
1	54,820	80.20	0.1035	5487	79.80	0.1055
2	27,264	80.33	0.1033	2730	80.11	0.1041
<u>3</u>	<u>17,892</u>	<u>80.86</u>	<u>0.1031</u>	<u>1889</u>	<u>80.31</u>	<u>0.1048</u>
<u>4</u>	<u>13,368</u>	<u>81.72</u>	<u>0.1024</u>	<u>1411</u>	<u>80.69</u>	<u>0.1038</u>
5	10712	80.45	0.1026	1071	80.18	0.1057
6	8876	80.40	0.1040	890	80.04	0.1061
7	7588	80.19	0.1057	759	78.53	0.1132
8	6612	80.12	0.1047	660	79.59	0.1121
9	5844	79.85	0.1068	584	78.92	0.1153
10	5208	80.05	0.1063	519	78.24	0.1151

Table 5. Results obtained with ECG network classifying Valence (best time windows are marked).

Time Window (s)	Train			Test		
	#Samples	Accuracy (%)	Loss	#Samples	Accuracy (%)	Loss
1	54,820	76.92	0.1151	5487	75.65	0.1167
2	27,264	76.97	0.1136	2730	76.09	0.1166
<u>3</u>	<u>17,892</u>	<u>78.19</u>	<u>0.1051</u>	<u>1889</u>	<u>77.47</u>	<u>0.1112</u>
<u>4</u>	<u>13,368</u>	<u>80.89</u>	<u>0.0998</u>	<u>1411</u>	<u>79.29</u>	<u>0.1028</u>
5	10712	76.23	0.1166	1071	76.29	0.1189
6	8876	75.41	0.1184	890	75.64	0.1240
7	7588	76.17	0.1158	759	76.16	0.1206
8	6612	76.16	0.1179	660	75.19	0.1227
9	5844	76.93	0.1122	584	76.30	0.1174
10	5208	75.42	0.1203	519	75.69	0.1206

Table 6. Results obtained with EEG network classifying Arousal (best time windows are marked).

Time Window (s)	Train			Test		
	#Samples	Accuracy (%)	Loss	#Samples	Accuracy (%)	Loss
1	54,637	69.37	0.1416	5755	67.32	0.1582
2	27,195	69.65	0.1411	2865	67.35	0.1541
<u>3</u>	<u>17,927</u>	<u>70.53</u>	<u>0.1325</u>	<u>1888</u>	<u>65.33</u>	<u>0.1657</u>
<u>4</u>	<u>13,374</u>	<u>70.97</u>	<u>0.1366</u>	<u>1409</u>	<u>63.72</u>	<u>0.1672</u>
5	10,662	71.54	0.1358	1123	62.98	0.1683
6	8841	72.26	0.1300	931	65.21	0.1744
7	7559	72.27	0.1302	796	63.00	0.1799
<u>8</u>	<u>6581</u>	<u>74.06</u>	<u>0.1210</u>	<u>693</u>	<u>63.46</u>	<u>0.1749</u>
<u>9</u>	<u>5813</u>	<u>75.63</u>	<u>0.1161</u>	<u>612</u>	<u>62.32</u>	<u>0.1913</u>
10	5195	65.34	0.1641	547	65.29	0.1652

Table 7. Results obtained with EEG network classifying Valence (best time windows are marked).

Time Window (s)	Train			Test		
	#Samples	Accuracy (%)	Loss	#Samples	Accuracy (%)	Loss
1	54,637	74.81	0.1216	5755	75.22	0.1215
2	27,195	75.05	0.1178	2865	73.99	0.1270
<u>3</u>	<u>17,927</u>	<u>76.36</u>	<u>0.1099</u>	<u>1888</u>	<u>72.09</u>	<u>0.1340</u>
<u>4</u>	<u>13,374</u>	<u>76.21</u>	<u>0.1096</u>	<u>1409</u>	<u>72.38</u>	<u>0.1416</u>
5	10,662	76.64	0.1114	1123	73.35	0.1307
6	8841	77.15	0.1043	931	73.28	0.1432
7	7559	77.04	0.1063	796	69.82	0.1423
<u>8</u>	<u>6581</u>	<u>77.27</u>	<u>0.1046</u>	<u>693</u>	<u>69.43</u>	<u>0.1505</u>
<u>9</u>	<u>5813</u>	<u>79.87</u>	<u>0.0956</u>	<u>612</u>	<u>68.05</u>	<u>0.1559</u>
<u>10</u>	<u>5195</u>	<u>80.34</u>	<u>0.0915</u>	<u>547</u>	<u>67.67</u>	<u>0.1684</u>

In the case of an EEG network, a $\approx 75\%$ accuracy is obtained for Arousal with a loss around 0.12 and a $\approx 80\%$ accuracy is obtained for Valence with a loss less than 0.10. The results obtained for Arousal are the worst of all physiological signals, so EEG is discarded for that. However, in the case of Valence classification, training results are similar that the ones obtained by ECG. However, there are two main problems about using EEG signal for Valence classification:

- The time window needed to obtain such results is very high compared with the one used for GSR and ECG. Thus, if the full system uses EEG, the data rate of the system will be reduced significantly.
- The difference between the accuracy of the training set and the accuracy of the testing set is too high (and the error too). This fact may mean the system is overtrained and it is not ready for new data, so testing the system with other users can cause bad results.

Thus, according to the results obtained for each neural network independently, for the case of classifying Arousal and Valence, GSR and ECG obtains better results than EEG. In fact, GSR obtains the best results for Arousal and ECG obtain the best results for Valence. About the optimum time window, if we overlapped the results from Tables 2–5, the best results are distributed between the 4 s and 5 s windows for GSR and between 3 s to 4 s windows for ECG; thus, the common time window that will be selected is 4 s.

Next, the full classification system is detailed and evaluated.

3.2. Full Classification System

According to the results obtained previously, features from GSR and ECG will be combined for the full classification system, obtaining a MLP neural network with these characteristics: input layer (6 features × 3 channels = 18 nodes), hidden layer 1 (36 nodes), hidden layer 2 (12 nodes), and output layer (three nodes).

Using a 4-s time window, the results obtained are detailed in Table 8. It is important to remark that, in order to obtain reliable results, samples from both sensors are not combined randomly: we use the time stamps of each video-clip that are stored with the sensors data; thus, in this way, we make sure that each GSR sample is combined with the ECG sample collected in parallel.

Table 8. Full classification system results: GSR and ECG combined.

Affective State	Train		Test	
	Accuracy (%)	Loss	Accuracy (%)	Loss
Arousal	95.87	0.0279	91.71	0.0643
Valence	94.64	0.0249	90.36	0.0589

Moreover, in order to verify that EEG is not suitable to this task, we have created a system composed by the features from the three sensors (EEG, ECG, and GSR). Results obtained after training and testing this system using a 4-s time window are detailed in Table 9.

Table 9. Classification system results with GSR, ECG, and EEG combined demonstrates that the inclusion of EEG reduces the accuracy.

Affective State	Train		Test	
	Accuracy (%)	Loss	Accuracy (%)	Loss
Arousal	81.95	0.0591	75.07	0.1351
Valence	83.38	0.0576	77.88	0.1349

As can be observed in Table 9, and comparing the results with Table 8, it is demonstrated that, for the goal of this work, the combination ECG + GSR gives more information about the affective state of the user than the combination of the three sensors (EEG + ECG + GSR).

After the final training with ECG and GSR, results showed that the classification system obtained by the combination of GSR and ECG coefficients improves the results obtained individually by each sensor significantly. The trend of the training process is growing for Arousal and Valence, as can be observed in Figure 3. Moreover, the error obtained for both affective states is much lower than the ones obtained individually. It is important to remark that there is a small difference between train and test subsets ($\approx 3\text{--}4\%$), but there is not a significant difference as the results obtained for the test subset are above 90–91%.

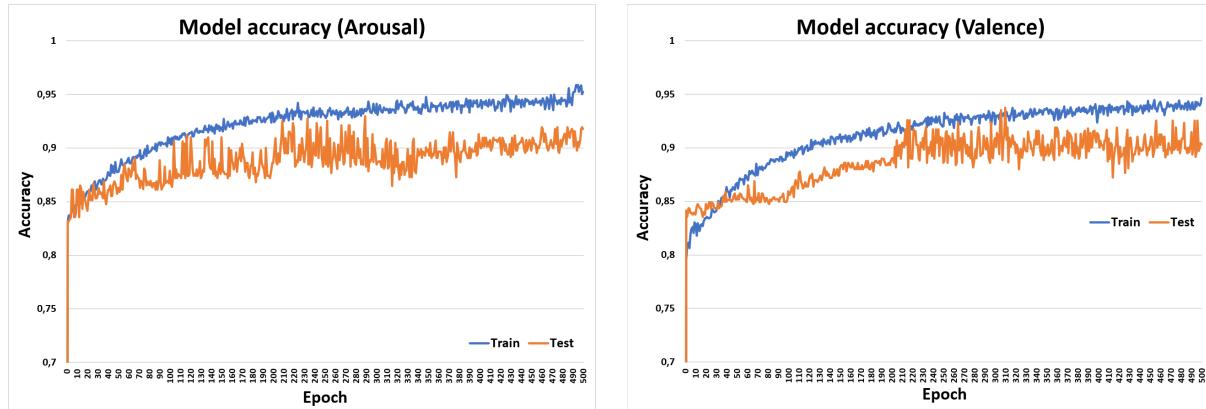


Figure 3. Full system accuracy for Arousal (left) and Valence (right).

In order to evaluate the errors obtained by the classifier with both affective states, confusion matrices for the train and test subsets are shown in Figure 4.

	LOW	MID	HIGH		LOW	MID	HIGH
LOW	3957	237	0	LOW	379	23	7
MID	55	5091	49	MID	19	592	12
HIGH	2	209	3768	HIGH	5	51	323
	LOW	MID	HIGH		LOW	MID	HIGH
LOW	3773	298	7	LOW	331	62	3
MID	75	5521	96	MID	25	561	11
HIGH	1	239	3358	HIGH	8	27	383

Figure 4. Confusion matrices: (top-left) train set with Arousal; (top-right) test set with Arousal; (bottom-left) train set with Valence; and (bottom-right) test set with Valence.

As shown in Figure 4, most of the errors occur between the extreme values (low and high) with the medium value. There are only a few cases where a high value is classified as low or vice versa.

In order to evaluate our system deeply, other metrics are commonly taken into account in neural network systems: sensitivity, specificity, precision, and F1-score. Those metrics have been calculated with the final system composed by ECG+GSR and the results are shown in Tables 10–13.

As can be observed in previous tables, results obtained for the training set are, indeed, better than the ones obtained for the testing set. Even so, F1-score results obtain $\approx 95\%$ for Valence and Arousal for the training set, and $\approx 90\%$ for the testing set. Some important aspects that can be observed with those final metrics are detailed next:

- F1-score: For the training sets, the extreme classes (LOW and HIGH) obtain better results than the MIDDLE one; but, for the testing sets, this is not true.
- Sensitivity: the class MIDDLE obtains better results than the others in all the cases.
- Specificity: the class MIDDLE obtains worse results than the others in all the cases.
- Precision: the class MIDDLE obtains worse results than the others in all the cases.

Table 10. Additional metrics obtained for Arousal (Train) classification with GSR + ECG system.

Class	True Positives	False Positives	True Negatives	False Negatives	Sensitivity	Specificity	Precision	F1-Score
LOW	3957	57	9117	237	0.943	0.994	0.986	0.964
MED	5091	446	7727	104	0.979	0.945	0.919	0.948
HIGH	3768	49	9340	211	0.947	0.995	0.987	0.966

Table 11. Additional metrics obtained for Arousal (Test) classification with the GSR + ECG system.

Class	True Positives	False Positives	True Negatives	False Negatives	Sensitivity	Specificity	Precision	F1-Score
LOW	379	24	978	30	0.926	0.976	0.941	0.933
MED	592	74	714	31	0.950	0.906	0.889	0.918
HIGH	323	19	1013	56	0.853	0.981	0.944	0.896

Table 12. Additional metrics obtained for Valence (Train) classification with the GSR + ECG system.

Class	True Positives	False Positives	True Negatives	False Negatives	Sensitivity	Specificity	Precision	F1-Score
LOW	3773	76	9214	305	0.925	0.992	0.980	0.952
MED	5521	537	7139	171	0.970	0.930	0.911	0.939
HIGH	3358	103	9667	240	0.933	0.989	0.970	0.951

Table 13. Additional metrics obtained for Valence (Test) classification with the GSR + ECG system.

Class	True Positives	False Positives	True Negatives	False Negatives	Sensitivity	Specificity	Precision	F1-Score
LOW	331	33	982	65	0.836	0.967	0.909	0.871
MED	561	89	725	36	0.939	0.890	0.863	0.899
HIGH	383	14	979	35	0.916	0.986	0.965	0.940

The analysis carried out previously is very interesting, since it describes that, in most of the metrics, the MIDDLE class obtains worse results than the others. This circumstance can be due to the fact that this class is not located in any extreme of the scale, so there are some occurrences that are close to the LOW and HIGH classes, and those ones can be confused with them. However, in the case of classes located at the extremes of the scale (LOW and HIGH), they only have a border where its occurrences can be confused. Moreover, usually, the occurrences with high or low values of Valence and/or Arousal tend to be extreme cases with extreme values (practically at the limits of their scale); so this is why they are commonly classified better than the occurrences of the MIDDLE class.

In order to achieve our final goal, we need the best results we can obtain, and it is true that a 100% accuracy is not obtained, but it is very difficult to obtain a perfect result when working with analog signals (which have a lot of noise) from multiple users and with multiple samples taken from each of them. Moreover, working with affective states, the variability between users is huge. For example, for the same video clip, two users can transmit completely opposite affective states (since it is something completely

subjective), but, in addition, the same affective state provided by the same user can fluctuate in intensity. This is why the results obtained can be considered enough for our final purpose.

3.3. Comparison

Finally, after presenting the results obtained with the full classification system, they will be compared with results obtained by similar works in the last years. This comparison is synthesized in Table 14. Results obtained for our system are detailed for training and testing subsets independently, although the other works do not distinguish between them.

One important improvement in the present work is the data resolution: since most works only use a two-level classification system for Arousal and Valence, we increase to three levels of classification: LOW, MEDIUM, and HIGH. The works that use a three-level resolution obtain an accuracy lower than our system (maximum values $\approx 88\%$ accuracy for Valence and $\approx 91\%$ accuracy for Arousal). On the other hand, the work with the best results is the one presented by JeeEun Lee and Sun K. Yoo in 2020 [54] with a $\approx 98\%$ accuracy, but it presents some deficiencies compared with our system, and it is important to comment on them:

- Data resolution: it uses only two levels for Arousal and Valence (negative and neutral). We use three (low, medium, and high).
- Time window: it uses a 30-s time window, obtaining a maximum processing rate of 0.033 samples per second. We use 4 s for the time window, so our system has around a 7.5x data rate (0.25 samples per second).
- Training epochs: the results are provided after a 5000-epoch training. We use only 500 epochs.
- Architecture complexity: it uses four hidden layers and a Recurrent Neural Network (RNN). We use a classical MLP neural network with only four hidden layers, so it is easier to be implemented in embedded systems.

Because of all these differences, we can affirm that our system is an improvement over other works with similar purposes. It is important to comment that several works have been discarded for this comparison because of the following reasons: they do not use physiological signals, they do not classify Arousal and Valence, or they do not provide accuracy results.

Table 14. Results comparison with previous works.

Work	Published	Output Resolution	Sensors	Technology	Accuracy
García, H. et al. [55]	2016	3 levels (Low, Medium, High)	EEG, EMG and EOG ¹	SVM ⁴	Valence: 88.3% Arousal: 90.6%
Liu, J. et al. [56]	2016	2 levels (Low, High)	EEG	KNN ⁵ and RF ⁶	Valence: 69.6% Arousal: 71.2%
Li, X. et al. [57]	2016	2 levels (Low, High)	EEG	C-RNN ⁷	Valence: 72.1% Arousal: 74.1%
Zhang, J. et al. [58]	2016	2 levels (Low, High)	EEG	PNN ⁸	Valence: 81.2% Arousal: 81.2%
Mirmohamadsadeghi, L. et al. [59]	2016	2 levels (Low, High)	ECG and Respiration	SVM ⁴	Valence: 74.0% Arousal: 74.0%
Zheng, W. et al. [60]	2017	3 levels (Negative, Neutral, Positive)	EEG	KNN ⁵ , LR ⁹ and SVM ⁴	Mean: 79.3%
Girardi, D. et al. [61]	2017	2 levels (Low, High)	EEG, GSR, and EMG	SVM ⁴	Valence: 63.9% Arousal: 58.6%
Lee, J. et al. [53]	2018	2 levels (Neutral, Negative)	ECG, GSR and SKT ²	NN ¹⁰	Mean: 92.5%
Lee, M. et al. [62]	2019	2 levels (Low, High)	PPG ³	CNN ¹¹	Valence: 75.3% Arousal: 76.2%
Sonkusare, S. et al. [63]	2019	2 levels (Low, High)	ECG, GSR and SKT ²	CNN ¹¹	Mean: 92%
Lee, J. et al. [54]	2020	2 levels (Neutral, Negative)	ECG, GSR and SKT ²	B-RNN ¹²	Mean: 98.4%
This work	2020	3 levels (Low, Medium, High)	ECG and GSR	NN ¹⁰	Valence (train): 94.6% Arousal (train): 95.9% Valence (test): 90.4% Arousal (test): 91.7%

¹ EOG: Electrooculography; ² SKT: Skin Temperature; ³ PPG: Photoplethysmography (Blood Pressure); ⁴ SVM: Supported Vector Machine; ⁵ KNN: k-nearest neighbours; ⁶ RF: Random Forest; ⁷ C-RNN: Convolutional Recurrent Neural Network; ⁸ PNN: Probabilistic Neural Network; ⁹ LR: Logic Regression; ¹⁰ NN: Classical MLP Neural Network; ¹¹ CNN: Convolutional Neural Network; ¹² B-RNN: Bidirectional Recurrent Neural Network.

4. Conclusions

After analyzing the information stored from the various physiological sensors that appear in the databases and after implementing and testing an independent classification system for each one, it is concluded that the GSR and ECG sensors provide more information about the two main affective states: Arousal and Valence. Therefore, in future works, the use of EEG can be discarded, which is an advantage because this sensor is the most complex to use, the one that requires the most preparation, the most expensive, the one with the greatest number of channels provided (and therefore the most computationally expensive), and the one with the greatest variability between users.

The accuracy results obtained for Arousal and Valence from the system composed of GSR and ECG show that the pre-processing step carried out has been correct for these types of problems, and that the study of the time window for data processing has been decisive to obtain positive results.

Comparing this work with other similar works over the last years, we can find notable improvements in all areas. The architecture used in our system has a low complexity compared to other architectures used in previous works, allowing our system to have a data-rate higher than all of them and to be able to be integrated with relative ease in an embedded system. The accuracy obtained by our system clearly surpasses most of the works developed previously except in the case of the work presented by Lee and Sun K. Yoo in 2020; however, this last work has a much more complex architecture and the results are discretized only in two levels, while our work performs a three-level classification: low, medium, and high.

These improvements not only allow for integrating our classifier into an embedded system, as mentioned above, but also for extending the classification of the affective state to high-level emotions like sadness, happiness, nervousness, etc. (according to the classification indicated in the Introduction section and shown in Figure 1) thanks to the improvement in the resolution of the classification.

As indicated at the beginning, our long-term objective is to develop a portable or wearable system for the detection of emotions that can be used as a non-verbal communication learning aid for users with cognitive disabilities. To achieve this goal, our next step is to reduce the device to a more manageable size to make it more comfortable when using it. Furthermore, as can be seen from the results, the EEG sensor is discarded because it obtains the worst classification results. Thus, when designing the device, it is not necessary to mount the EEG helmet and, therefore, the other sensors can be located in a more comfortable area, such as the arm or hand.

Thanks to the advances presented in this work, it is demonstrated that this objective is feasible.

Author Contributions: Conceptualization: L.M.-S., M.D.-M., A.C.; methodology: L.M.-S., L.M.-A., M.D.-M.; software: L.M.-S., F.L.-P., J.C.-M.; formal analysis: L.M.-S., L.M.-A., A.C.; investigation: L.M.-S., F.L.-P., J.C.-M.; resources: L.M.-A., M.D.-M., A.C.; data curation: L.M.-S., F.L.-P., JCM; writing—original draft preparation: L.M.-S.; writing—review and editing: L.M.-S., F.L.-P., J.C.-M., L.M.-A., M.D.-M., A.C.; visualization: L.M.-A., M.D.-M., A.C.; supervision: L.M.-A., M.D.-M.; project administration: A.C.; funding acquisition: A.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially funded by the Telefonica Chair “Intelligence in Networks” of the Universidad de Sevilla, Spain.

Acknowledgments: The authors would like to thank the support of the Telefonica Chair “Intelligence in Networks” of the Universidad de Sevilla, Spain.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Argyle, M. Non-verbal communication in human social interaction. In *Non-Verbal Communication*; Cambridge Universit Press: Cambridge, UK, 1972.
2. Knapp, M.L.; Hall, J.A.; Horgan, T.G. *Nonverbal Communication in Human Interaction*; Cengage Learning: Boston, MA, USA, 2013.
3. Isbister, K.; Nass, C. Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics. *Int. J. Hum.-Comput. Stud.* **2000**, *53*, 251–267. [[CrossRef](#)]
4. Tirapu-Ustároz, J.; Pérez-Sayes, G.; Erekatxo-Bilbao, M.; Pelegrín-Valero, C. ¿Qué es la teoría de la mente. *Revista de Neurología* **2007**, *44*, 479–489. [[CrossRef](#)] [[PubMed](#)]
5. Volkmar, F.R.; Sparrow, S.S.; Rende, R.D.; Cohen, D.J. Facial perception in autism. *J. Child Psychol. Psychiatry* **1989**, *30*, 591–598. [[CrossRef](#)] [[PubMed](#)]
6. Celani, G.; Battacchi, M.W.; Arcidiacono, L. The understanding of the emotional meaning of facial expressions in people with autism. *J. Autism Dev. Disord.* **1999**, *29*, 57–66. [[CrossRef](#)] [[PubMed](#)]
7. Hatfield, E.; Cacioppo, J.T.; Rapson, R.L. Emotional contagion. *Curr. Dir. Psychol. Sci.* **1993**, *2*, 96–100. [[CrossRef](#)]
8. James, W. William James writings 1878–1899, chapter on emotion. *Libr. Am.* **1992**, 350–365.
9. Lange, C. Über Gemüthsbewegungen. Leipzig, Thomas. In *The Emotions: A Psychophysiological Study*; Hafner Publishing: New York, NY, USA, 1885; pp. 33–90.
10. Cannon, W.B. The James-Lange theory of emotions: A critical examination and an alternative theory. *Am. J. Psychol.* **1927**, *39*, 106–124. [[CrossRef](#)]
11. LeDoux, J.E. Emotion circuits in the brain. *Ann. Rev. Neurosci.* **2000**, *23*, 155–184. [[CrossRef](#)]
12. Lang, P.; Bradley, M.M. The International Affective Picture System (IAPS) in the study of emotion and attention. *Handb. Emot. Elicitation Assess.* **2007**, *29*, 70–73.
13. Wiens, S.; Öhman, A. Probing Unconscious Emotional Processes on Becoming A Successful Masketeer. In *Handbook of Emotion Elicitation and Assessment*; Oxford University Press: Oxford, UK, 2007.
14. Ekman, P. The directed facial action task. *Handb. Emot. Elicitation Assess.* **2007**, *47*, 53.
15. Laird, J.D.; Strout, S. Emotional behaviors as emotional stimuli. In *Handbook of Emotion Elicitation and Assessment*; Oxford University Press: Oxford, UK, 2007; pp. 54–64.
16. Amodio, D.M.; Zinner, L.R.; Harmon-Jones, E. Social psychological methods of emotion elicitation. *Handb. Emot. Elicitation Assess.* **2007**, *91*, 91–105.
17. Roberts, N.A.; Tsai, J.L.; Coan, J.A. Emotion elicitation using dyadic interaction tasks. *Handb. Emot. Elicitation Assess.* **2007**, *106*–123.
18. Eich, E.; Ng, J.T.; Macaulay, D.; Percy, A.D.; Grebneva, I. Combining music with thought to change mood. In *Handbook of Emotion Elicitation and Assessment*; Oxford University Press: Oxford, UK, 2007; pp. 124–136.
19. Rottenberg, J.; Ray, R.; Gross, J. Emotion elicitation using films. In *Handbook of Emotion Elicitation and Assessment*; Oxford University Press: Oxford, UK, 2007.
20. Tooby, J.; Cosmides, L. The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethol. Sociobiol.* **1990**, *11*, 375–424. [[CrossRef](#)]
21. Coan, J.A.; Allen, J.J. Frontal EEG asymmetry as a moderator and mediator of emotion. *Biol. Psychol.* **2004**, *67*, 7–50. [[CrossRef](#)] [[PubMed](#)]
22. Posner, J.; Russell, J.A.; Peterson, B.S. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev. Psychopathol.* **2005**, *17*, 715. [[CrossRef](#)]
23. Berridge, K.C. Pleasures of the brain. *Brain Cogn.* **2003**, *52*, 106–128. [[CrossRef](#)]
24. Berkowitz, L.; Jaffee, S.; Jo, E.; Troccoli, B.T. On the correction of feeling-induced judgmental biases. In *Feeling and Thinking: The Role of Affect in Social Cognition*; Cambridge University Press: Cambridge, UK, 2000; pp. 131–152.
25. Al-Qazzaz, N.K.; Hamid Bin Mohd Ali, S.; Ahmad, S.A.; Islam, M.S.; Escudero, J. Selection of mother wavelet functions for multi-channel EEG signal analysis during a working memory task. *Sensors* **2015**, *15*, 29015–29035. [[CrossRef](#)]

26. Mjahad, A.; Rosado-Muñoz, A.; Guerrero-Martínez, J.F.; Bataller-Mompeán, M.; Francés-Villora, J.V.; Dutta, M.K. Detection of ventricular fibrillation using the image from time-frequency representation and combined classifiers without feature extraction. *Appl. Sci.* **2018**, *8*, 2057. [[CrossRef](#)]
27. Ji, N.; Ma, L.; Dong, H.; Zhang, X. EEG Signals Feature Extraction Based on DWT and EMD Combined with Approximate Entropy. *Br. Sci.* **2019**, *9*, 201. [[CrossRef](#)]
28. Ji, Y.; Zhang, S.; Xiao, W. Electrocardiogram classification based on faster regions with convolutional neural network. *Sensors* **2019**, *19*, 2558. [[CrossRef](#)]
29. Oh, S.L.; Vicnesh, J.; Ciaccio, E.J.; Yuvaraj, R.; Acharya, U.R. Deep convolutional neural network model for automated diagnosis of schizophrenia using EEG signals. *Appl. Sci.* **2019**, *9*, 2870. [[CrossRef](#)]
30. Civit-Masot, J.; Domínguez-Morales, M.J.; Vicente-Díaz, S.; Civit, A. Dual Machine-Learning System to Aid Glaucoma Diagnosis Using Disc and Cup Feature Extraction. *IEEE Access* **2020**, *8*, 127519–127529. [[CrossRef](#)]
31. Civit-Masot, J.; Luna-Perejón, F.; Domínguez Morales, M.; Civit, A. Deep Learning System for COVID-19 Diagnosis Aid Using X-ray Pulmonary Images. *Appl. Sci.* **2020**, *10*, 4640. [[CrossRef](#)]
32. Gao, C.; Neil, D.; Ceolini, E.; Liu, S.C.; Delbruck, T. DeltaRNN: A power-efficient recurrent neural network accelerator. In Proceedings of the 2018 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays, Monterey, CA, USA, 25–27 February 2018; pp. 21–30.
33. Luna-Perejón, F.; Domínguez-Morales, M.J.; Civit-Balcells, A. Wearable fall detector using recurrent neural networks. *Sensors* **2019**, *19*, 4885. [[CrossRef](#)]
34. Crites, S.L., Jr.; Cacioppo, J.T. Electrocortical differentiation of evaluative and nonevaluative categorizations. *Psychol. Sci.* **1996**, *7*, 318–321. [[CrossRef](#)]
35. Cuthbert, B.N.; Schupp, H.T.; Bradley, M.M.; Birbaumer, N.; Lang, P.J. Brain potentials in affective picture processing: Covariation with autonomic arousal and affective report. *Biol. Psychol.* **2000**, *52*, 95–111. [[CrossRef](#)]
36. Cacioppo, J.T.; Berntson, G.G.; Larsen, J.T.; Poehlmann, K.M.; Ito, T.A. The psychophysiology of emotion. *Handb. Emotions* **2000**, *2*, 173–191.
37. Graham, F.K.; Clifton, R.K. Heart-rate change as a component of the orienting response. *Psychol. Bull.* **1966**, *65*, 305. [[CrossRef](#)] [[PubMed](#)]
38. Prkachin, K.M.; Williams-Avery, R.M.; Zwaal, C.; Mills, D.E. Cardiovascular changes during induced emotion: An application of Lang's theory of emotional imagery. *J. Psychosom. Res.* **1999**, *47*, 255–267. [[CrossRef](#)]
39. Cacioppo, J.T.; Berntson, G.G.; Klein, D.J.; Poehlmann, K.M. Psychophysiology of emotion across the life span. *Ann. Rev. Gerontol. Geriatr.* **1997**, *17*, 27–74.
40. Codispoti, M.; Bradley, M.M.; Lang, P.J. Affective reactions to briefly presented pictures. *Psychophysiology* **2001**, *38*, 474–478. [[CrossRef](#)] [[PubMed](#)]
41. Bradley, M.M.; Lang, P.J.; Cuthbert, B.N. Emotion, novelty, and the startle reflex: Habituation in humans. *Behav. Neurosci.* **1993**, *107*, 970. [[CrossRef](#)]
42. Cacioppo, J.T.; Tassinary, L.G.; Fridlund, A.J. *The Skeletomotor System*; Cambridge University Press: Cambridge, UK, 1990.
43. Schwartz, G.E.; Fair, P.L.; Salt, P.; Mandel, M.R.; Klerman, G.L. Facial muscle patterning to affective imagery in depressed and nondepressed subjects. *Science* **1976**, *192*, 489–491. [[CrossRef](#)]
44. Lang, P.J.; Greenwald, M.K.; Bradley, M.M.; Hamm, A.O. Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology* **1993**, *30*, 261–273. [[CrossRef](#)]
45. Greenwald, M.K.; Cook, E.W.; Lang, P.J. Affective judgment and psychophysiological response: Dimensional covariation in the evaluation of pictorial stimuli. *J. Psychophysiol.* **1989**, *3*, 51–64.
46. Witvliet, C.V.; Vrana, S.R. Psychophysiological responses as indices of affective dimensions. *Psychophysiology* **1995**, *32*, 436–443. [[CrossRef](#)]
47. Cacioppo, J.T.; Petty, R.E.; Losch, M.E.; Kim, H.S. Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *J. Personal. Soc. Psychol.* **1986**, *50*, 260. [[CrossRef](#)]
48. Ekman, P. Facial expression and emotion. *Am. Psychol.* **1993**, *48*, 384. [[CrossRef](#)]
49. Lang, P.J. The emotion probe: Studies of motivation and attention. *Am. Psychol.* **1995**, *50*, 372. [[CrossRef](#)]

50. Bradley, M.M.; Codispoti, M.; Cuthbert, B.N.; Lang, P.J. Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion* **2001**, *1*, 276. [[CrossRef](#)]
51. Subramanian, R.; Wache, J.; Abadi, M.K.; Vieriu, R.L.; Winkler, S.; Sebe, N. ASCERTAIN: Emotion and personality recognition using commercial sensors. *IEEE Tran. Affect. Comput.* **2016**, *9*, 147–160. [[CrossRef](#)]
52. Katsigiannis, S.; Ramzan, N. DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices. *IEEE J. Biomed. Health Inform.* **2017**, *22*, 98–107. [[CrossRef](#)] [[PubMed](#)]
53. Lee, J.; Yoo, S.K. Design of user-customized negative emotion classifier based on feature selection using physiological signal sensors. *Sensors* **2018**, *18*, 4253. [[CrossRef](#)]
54. Lee, J.; Yoo, S.K. Recognition of Negative Emotion Using Long Short-Term Memory with Bio-Signal Feature Compression. *Sensors* **2020**, *20*, 573. [[CrossRef](#)]
55. García, H.F.; Álvarez, M.A.; Orozco, Á.A. Gaussian process dynamical models for multimodal affect recognition. In Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, 16–20 August 2016; pp. 850–853.
56. Liu, J.; Meng, H.; Nandi, A.; Li, M. Emotion detection from EEG recordings. In Proceedings of the 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Changsha, China, 13–15 August 2016; pp. 1722–1727.
57. Li, X.; Song, D.; Zhang, P.; Yu, G.; Hou, Y.; Hu, B. Emotion recognition from multi-channel EEG data through convolutional recurrent neural network. In Proceedings of the 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Shenzhen, China, 15–18 December 2016; pp. 352–359.
58. Zhang, J.; Chen, M.; Hu, S.; Cao, Y.; Kozma, R. PNN for EEG-based Emotion Recognition. In Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary, 9–12 October 2016.
59. Mirmohamadsadeghi, L.; Yazdani, A.; Vesin, J.M. Using cardio-respiratory signals to recognize emotions elicited by watching music video clips. In Proceedings of the 2016 IEEE 18th International Workshop on Multimedia Signal Processing (MMSP), Montreal, QC, Canada, 21–23 September 2016; pp. 1–5.
60. Zheng, W.L.; Zhu, J.Y.; Lu, B.L. Identifying stable patterns over time for emotion recognition from EEG. *IEEE Trans. Affect. Comput.* **2017**, *10*, 417–429. [[CrossRef](#)]
61. Girardi, D.; Lanobile, F.; Novielli, N. Emotion detection using noninvasive low cost sensors. In Proceedings of the 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), San Antonio, TX, USA, 23–26 October 2017; pp. 125–130.
62. Lee, M.S.; Lee, Y.K.; Pae, D.S.; Lim, M.T.; Kim, D.W.; Kang, T.K. Fast Emotion Recognition Based on Single Pulse PPG Signal with Convolutional Neural Network. *Appl. Sci.* **2019**, *9*, 3355. [[CrossRef](#)]
63. Sonkusare, S.; Ahmedt-Aristizabal, D.; Aburn, M.J.; Nguyen, V.T.; Pang, T.; Frydman, S.; Denman, S.; Fookes, C.; Breakspear, M.; Guo, C.C. Detecting changes in facial temperature induced by a sudden auditory stimulus based on deep learning-assisted face tracking. *Sci. Rep.* **2019**, *9*, 4729. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).