Abstract: Estimation of human emotions plays an important role in the development of modern brain-computer interface devices like the Emotiv EPOC+ headset. In this paper, we present an experiment to assess the classification accuracy of the emotional states provided by the headset’s application programming interface (API). In this experiment, several sets of images selected from the International Affective Picture System (IAPS) dataset are shown to sixteen participants wearing the headset. Firstly, the participants’ responses in form of a self-assessment manikin questionnaire to the emotions elicited are compared with the validated IAPS predefined valence, arousal and dominance values. After statistically demonstrating that the responses are highly correlated with the IAPS values, several artificial neural networks (ANNs) based on the multilayer perceptron architecture are tested to calculate the classification accuracy of the Emotiv EPOC+ API emotional outcomes. The best result is obtained for an ANN configuration with three hidden layers, and 30, 8 and 3 neurons for layers 1, 2 and 3, respectively. This configuration offers 85% classification accuracy, which means that the emotional estimation provided by the headset can be used with high confidence in real-time applications that are based on users’ emotional states. Thus the emotional states given by the headset’s API may be used with no further processing of the electroencephalogram signals acquired from the scalp, which would add a level of difficulty.

Keywords: artificial neural network; brain-computer interface; electroencephalography; emotional state; assessment

1. Introduction

Emotions play an important role in human relations and they have gained interest in human-computer interaction as well in recent years. For instance, the possibilities of including users’ emotional states as input to computing systems are being explored to dynamically adapt the interface to their needs at each moment [1–5]. Psychologists distinguish between psychological excitement, expression of behaviour and conscious experience of emotions [6]. Traditional approaches analyse changes in facial expression and/or voice to infer emotional states [7–9].

However, emotions are not always easily exhibited through these cues. This is why specialised devices capturing brain activity are used more and more to detect emotions. A brain-computer interface...
(BCI), which uses electroencephalography (EEG) techniques to record the electrical activity of the brain, is one such device. Among the manifold applications, EEG signals allow one to create emotional models capable of reporting a person’s emotional state when an external stimulus is presented [6].

Traditionally, EEG techniques have been applied to medical applications [10–12]. Nonetheless, they are now being used in different fields such diverse as marketing, video games and e-learning [13,14]. These new fields of application have pushed the evolution of EEG devices, following the new users’ needs in terms of usability, affordability and portability. The Emotiv EPOC+ headset (https://www.emotiv.com/EPOC/) is one of these new devices. It is a low-priced and lightweight portable BCI-EEG device that offers great flexibility compared to traditional devices used in medicine. Even so, the device captures, processes and analyses in real-time a relatively high amount of data, which have enabled its use in new tele-health systems [15].

The purpose of this paper is the assessment of the classification accuracy of the emotional states provided by the application programming interface (API) of the Emotiv EPOC+ headset. For this, an experiment is introduced in which several sets of images, extracted from the International Affective Picture System (IAPS) [16] database, are presented to sixteen participants. Their subjective answers and the values provided by the API are compared with valence, arousal and dominance values of the images as validated and labelled in IAPS. The recordings of these values are analysed through artificial neural networks (ANNs) to validate the emotional model. An ANN is a type of non-linear classifier used in many applications in a wide variety of disciplines [17]. It is not surprising that ANNs have recently gained significant interest in emotion classification through EEG signals in the last few years [18–22].

To the best of our knowledge, there is no previous work assessing the accuracy of the headset’s real-time measurements by comparing them with a validated image database by means of ANNs. The rest of the paper is organised as follows. In Section 2 the materials and methods used in the study are described. Section 3 shows the results obtained. Lastly, Section 4 exposes the more relevant conclusions and discussion derived from the present study.

2. Materials and Methods

2.1. Materials

2.1.1. Emotiv EPOC+ Headset

The Emotiv EPOC+ BCI-EEG headset (see Figure 1a) was originally marketed for video games where players control various aspects within the game environment. Over the years, its use has increased within new research areas due to its low price and portability. Emotiv EPOC+ headset has 14 sensors and 2 references (see Figure 1b). The sampling frequency for each of the channels is 128 Hz and the battery life is approximately 12 hours with regular use. The captured signals are transferred to a computer through a 2.4 GHz wireless connection [23]. The device has copper electrodes to facilitate the detection of the signal. They must be slightly moistened with a saline solution to allow the sensors to make good contact with the user’s scalp, as well as to act as a disinfectant. The device implements a control system to monitor the quality of the contacts with the skin. This monitoring makes it possible to detect if any electrode has dried during use [6].

The device’s API monitors six different cognitive states in real-time: excitement (arousal), interest (valence), stress (frustration), engagement/boredom, attention (focus) and meditation (relaxation). The API uses two libraries (EmoEngine and EmoKey) that estimate the emotional state of the user. EmoEngine is a library that stores signals coming from the EEG channels, while EmoKey processes these signals and provides numerical values for each channel. All values are normalised between 0 (lower intensity) and 1 (higher intensity). The person in charge of interpreting the data does not require previous training to correctly read emotions [23,24].

Emotiv EPOC+ headset captures brainwaves but cannot acquire thoughts, feelings or intentions. To date, Emotiv has not made public how the classification algorithm works, so the emotional model
implemented in the API remains unknown. The only information is that hundreds of volunteers were
engaged by the company to capture data while watching movies and playing video games. After this,
the volunteers filled out a questionnaire with their psychological experiences [24].

![Image of Emotiv EPOC+ headset and spatial mapping of electrodes on the scalp.]

Figure 1. (a) Emotiv EPOC+ headset; (b) Spatial mapping of the electrodes on the scalp.

2.1.2. EmoSys Software Suite

EmoSys is a software suite developed by some of the authors of this paper. EmoSys enables the
integration of several devices that collect (neuro)physiological signals. EmoSys processes EEG data
obtained from EmoEngine and EmoKey, synchronising and storing them in a .csv (comma-separated
values) file, among other possibilities.

Figure 2 shows a basic scheme of EmoSys’ operation. The BCI-EEG device captures EEG signals
of the person under study. The raw signals are stored using EmoEngine, and EmoKey processes them
and obtains their metrics. Finally, the raw EEG signals and emotional metrics are combined in EmoSys,
obtaining two .csv files that are used later on for further analysis.

![Diagram of EmoSys system with EEG input, EmoEngine, EmoKey, Emotion Metrics, and EmoSys outputs.

Figure 2. Acquisition system using the headset, its application programming interface (API)
and EmoSys. Electroencephalography (EEG), brain-computer interface (BCI), comma-separated
values (CSV).

The EmoSys user interface is depicted in Figure 3. Even though it allows the integration of several
devices, this figure only shows data coming from the EPOC+ device. The left side of the picture
offers the electrodes’ current connection status using a simple color code: green for a good connection
quality and black for electrodes not providing information. The right side plots data coming from each
channel. There is also a set of buttons at the top-left of the window to start/stop data recording.
Figure 3. EmoSys user interface showing data received from Emotiv EPOC+ device.
2.2. Methods

2.2.1. Participants

Sixteen people took part in this study, concretely 9 men (56.25%) and 7 women (43.75%). All of them were in good physical and mental health. They were all volunteers and did not receive any financial compensation for their participation. The participants signed a participation agreement informing them about the type of images they would be shown and the possibility of stopping the experiment at any time. This study was conducted in accordance with the Declaration of Helsinki, and, as real patients were not involved, the approval of an Ethics Committee in Clinical Research was not required according to Spanish and European legislation.

The experiment was conducted in a controlled environment. Each participant was seated in a comfortable environment without any external stimuli that might condition him/her. The sensors were moistened with a saline solution to improve their contact with the scalp and the headset was fitted to each participant straightaway. The EmoSys application was launched to store the data captured by the headset. Once the experiment started, the participant was left alone in order not to be conditioned in any way.

2.2.2. E-Prime and IAPS

E-Prime was used to control the conditions of the experiment. E-Prime is a tool widely used in psychological experimentation. It covers from the design phase of an experiment to the collection and exportation of the data for a later analysis. The design is done through slides where text, images, video clips and even personalised questionnaires are included [25]).

As images are the input to evoke emotions in participants, the well-known and validated International Affective Picture System (IAPS) database was chosen. In fact, the IAPS database is one of the most commonly used image databases to perform emotional experiments. It consists of a set of colour images depicting different objects or situations that are grouped into categories associated with a specific emotional state [26]. The exposure to these images seeks to incite the same emotion in the subject as the one suggested in the photograph. The database was originally validated using a graphic scale self-assessment manikin (SAM) questionnaire [27] (see Figure 4) by asking participants to rate how pleasant/unpleasant, calm/excited and controlled they felt when looking at each of them. Therefore, the mean value and standard deviation for valence, arousal and dominance are known for each picture. This opens the door to comparing these values with those offered by the API of the Emotiv EPOC+.

![Figure 4. Self-assessment manikin. A set of images is used to value valence (top row), arousal (middle row) and dominance (bottom row).](image-url)
In our specific experiment, several IAPS images with different levels of valence, arousal and dominance are selected. Four types of experimental conditions are selected: HH condition, characterised by high valence and high arousal, HL by high valence and low arousal, LH by low valence and high arousal, and LL by low valence and low arousal. Dominance in all conditions is usually medium. Table 1 shows the average values for each group consisting of 25 images.

Table 1. Mean value and standard deviation for valence, arousal and dominance of each group of International Affective Picture System (IAPS) images.

<table>
<thead>
<tr>
<th>Experimental Condition</th>
<th>Valence</th>
<th>Arousal</th>
<th>Dominance</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH (high valence, high arousal)</td>
<td>7.09 (1.61)</td>
<td>6.43 (2.03)</td>
<td>5.34 (2.19)</td>
</tr>
<tr>
<td>HL (high valence, low arousal)</td>
<td>7.23 (1.54)</td>
<td>3.26 (2.22)</td>
<td>6.44 (2.10)</td>
</tr>
<tr>
<td>LH (low valence, high arousal)</td>
<td>1.67 (1.21)</td>
<td>6.93 (2.22)</td>
<td>2.79 (2.11)</td>
</tr>
<tr>
<td>LL (low valence, low arousal)</td>
<td>3.48 (1.51)</td>
<td>3.65 (1.99)</td>
<td>4.80 (2.08)</td>
</tr>
</tbody>
</table>

2.2.3. Experiment Design

As already stated, the aim of this experiment was to compare the values of the emotional states obtained from Emotiv EPOC+ API with the previously known and validated values evoked by images belonging to the IAPS benchmark. The study uses all the tools described before.

As also pointed out, E-Prime was used to control the execution of the experiment. The experiment started by informing the participant about the procedure of the experiment and how to respond to the SAM scale. After that, a set of images with neutral values of valence and arousal was presented, followed by a distracting task aimed at eliminating the image-based emotion induced. Next, the system entered into a loop that was repeated four times. First the participant was shown a block of images related to one of the groups (HH, HL, LH or LL; see Table 1). In this case, ten images of each group were randomly chosen for display. After that, a distracting task was proposed to the participant, and he/she completed the SAM questionnaire before starting a new loop. Once all image blocks had been displayed, along with their corresponding distracting task, the test was finished, and the participant was thanked for his/her time spent. The steps are depicted in the flowchart illustrated in Figure 5.

![Figure 5. Flowchart of the experimental design used in E-Prime. Self-assessment manikin (SAM).](image-url)

Simultaneously, the EmoSys Software Suite captures and stores both EEG signals (raw signals) and emotional states obtained from the Emotiv EPOC+ API. In this study, only the emotional states obtained from Emotiv EPOC+ API are utilised. Once emotional states have been stored, segmented and
synchronised with E-Prime data, the relationship between each emotional state and its corresponding output is calculated by using artificial neural networks (see Figure 6).

**Figure 6.** Flowchart of the experimental design. Artificial neural network (ANN).

### 2.2.4. Multilayer Perceptron Architecture

Given the nature of the problem at hand, which is not very complex, we decided to use the multilayer perceptron (MLP) architecture. An MLP consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilises a supervised learning technique called backpropagation for training. Moreover, the connections of an MLP are always directed forward, hence they are also called feed-forward networks [28,29]. The connections define a relationship between the input and output variables of the network. Each neuron in the network processes the information received by its inputs and produces a response or activation that propagates to the neurons of the next layer. In our particular case, the values of the emotional states provided by the Emotiv EPOC+ API define the input layer. The values of the emotional states belonging and validated for each of the photographs in the IAPS library are used in the output layer.

The automatic determination of the number of neurons and the number of hidden layers that optimise the resolution of the problem at hand has been debated [30]. Indeed, it is not possible to demonstrate that using architectures in which connections from one layer are removed or added to layers that are not immediately subsequent will produce better results [29]. Therefore, there is no method or rule that determines the optimal number of hidden layers and/or neurons to solve a given problem. Generally, the number of hidden layers and neurons in each layer are determined by trial and error. For the number of neurons at a given layer, we have used a method called inverted geometric pyramid [31] (see Figure 7).

**Figure 7.** Flowchart of a multilayer perceptron (MLP) artificial neural network.
The learning rule is the mechanism by which the neural network learns how to classify the data. Each neuron adapts and modifies its network parameters so that the current output is as close as possible to the expected output. Thus, network learning is formulated as a problem of error minimisation. In our case, error minimisation is performed by using Levenberg-Marquard (L-M) and Bayesian Regularisation (BR) methods [28]. These two methods were selected because they have several advantages over others. The first advantage is that they are fast and present in most neural network libraries and toolboxes. Second, they are also easily configurable and their adjustment parameters are simple and easily understandable. As is usual in ANN learning, data are randomly separated into three sets: training, validation and test. We use 70% for training, 15% for validation and 15% for testing. Due to the non-deterministic nature of ANNs, each test is repeated 30 times. We randomly distribute data in each test with a random separation algorithm [17,29].

3. Results

This section compares the results obtained with different ANN configurations. The target outputs of the tested networks are the values of valence, arousal and dominance, which are known in advance for each IAPS image selected for experimentation. The configurations use different number of layers, neurons per layer and learning methods. Moreover, the inverted pyramid method is used to determine the number of neurons in the first layer. As a first approximation, the product of the number of input and output variables is used and then increased to find the maximum performance of the network. The process stops when the performance starts decreasing. Table 2 shows ANN configurations that offer the best classification accuracies in terms of the purposes established in this study. These are single-layer (L = 1) with N = 15 or N = 30 neurons, and multiple-layer with L = 3 layers and N = 15-8-3 or N = 30-8-3 neurons. Both L-M and BR learning methods have been used for the configurations. In all cases, the activation function used is the sigmoid.

<table>
<thead>
<tr>
<th>Hidden Layer Configuration</th>
<th>Neurons per Hidden Layer</th>
<th>Learning Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-Layer (L = 1)</td>
<td>N = 15</td>
<td>L-M</td>
</tr>
<tr>
<td></td>
<td>N = 15</td>
<td>BR</td>
</tr>
<tr>
<td></td>
<td>N = 30</td>
<td>L-M</td>
</tr>
<tr>
<td></td>
<td>N = 30</td>
<td>BR</td>
</tr>
<tr>
<td>Multiple-Layer (L = 3)</td>
<td>N = 15-8-3</td>
<td>L-M</td>
</tr>
<tr>
<td></td>
<td>N = 15-8-3</td>
<td>BR</td>
</tr>
<tr>
<td></td>
<td>N = 30-8-3</td>
<td>L-M</td>
</tr>
<tr>
<td></td>
<td>N = 30-8-3</td>
<td>BR</td>
</tr>
</tbody>
</table>

3.1. Assessment of SAM Responses vs. IAPS Values

Firstly, the answers to the SAM scale obtained from the participants of the present study are compared to the values provided by the original IAPS database validation as an initial step towards determining the classification accuracy of the emotional model implemented by the API of the Emotiv EPOC+. Should the correlation between the two sets of results be high enough, this would mean that the subset of selected IAPS images is a good representation of the emotions that were intended to be evoked to the participants. This is an excellent starting point for checking the effectiveness of the model developed for the study.

This is the reason why the percentage of hits have been calculated for all the answers given by the sixteen participants. The hits are obtained assuming a normal distribution of the values provided in IAPS. We consider that a response in SAM questionnaire is a hit with regards to the IAPS values when it lies within one standard deviation of the mean. Table 3 shows the percentage of hits for a 68.27% account.
The results of mean hits are not as good as expected (see Table 3). Therefore, in order to try to better adjust the correlation between SAM responses and IAPS values to the specific sixteen participants of this study, artificial neural networks are tested. The use of this type of approach may be oversized. Nevertheless, with this special purpose in mind, ANNs are designed by using the SAM mean values for valence, arousal and dominance declared by the participants as input parameters, and the mean values provided by IAPS pictures as output parameters. As shown in Table 4 the classification accuracies for each of the implemented configurations is above 90% for all performed analyses. According to these results, L-M offers the highest result (0.96% for L = 3/N = 15-8-3, shown in bold). Hence, a higher correlation is reached through training ANNs.

### Table 3. Percentage of hits per condition.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Valence</th>
<th>Arousal</th>
<th>Dominance</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH</td>
<td>93.75</td>
<td>81.25</td>
<td>87.50</td>
</tr>
<tr>
<td>HL</td>
<td>93.75</td>
<td>81.25</td>
<td>93.75</td>
</tr>
<tr>
<td>LH</td>
<td>75.00</td>
<td>87.50</td>
<td>56.25</td>
</tr>
<tr>
<td>LL</td>
<td>81.25</td>
<td>68.75</td>
<td>75.00</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td><strong>85.94</strong></td>
<td><strong>79.69</strong></td>
<td><strong>78.13</strong></td>
</tr>
</tbody>
</table>

3.2. Different ANN Configurations to Compare Emotiv EPOC+ API Outcomes with IAPS Values

In second place, many ANN configurations are tested by varying each parameter present at every layer with the aim of comparing Emotiv EPOC+ API outcomes with IAPS values. Although we are limited by hardware in testing larger ANNs, a very large configuration is not necessary to obtain good results.

In our case, 76% classification accuracy is obtained with an L = 1/N = 15 configuration by using the L-M adjustment method for all emotional states present during the experiment (see Table 5). Conversely, if the number of hidden layers is increased to L = 3 (using the same parameters), the performance of the network decreases 3% for the L-M method. On the other hand, the network performance increases up to 85% (shown in bold) when using the BR method. This increment is due to the fact that this type of approach is usually better for small datasets and a large number of layers.

### Table 4. Classification accuracy of different ANN configurations comparing SAM responses with IAPS values. Levenberg-Marquard (L-M) and Bayesian Regularisation (BR).

<table>
<thead>
<tr>
<th>L = 1/N = 15</th>
<th>L = 1/N = 30</th>
<th>L = 3/N = 15-8-3</th>
<th>L = 3/N = 30-8-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-M</td>
<td>BR</td>
<td>L-M</td>
<td>BR</td>
</tr>
<tr>
<td>Training</td>
<td>0.93</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>Validation</td>
<td>0.92</td>
<td>–</td>
<td>0.91</td>
</tr>
<tr>
<td>Test</td>
<td>0.94</td>
<td>0.91</td>
<td>0.90</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>L = 1/N = 15</th>
<th>L = 1/N = 30</th>
<th>L = 3/N = 15-8-3</th>
<th>L = 3/N = 30-8-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-M</td>
<td>BR</td>
<td>L-M</td>
<td>BR</td>
</tr>
<tr>
<td>Training</td>
<td>0.70</td>
<td>0.44</td>
<td>0.71</td>
</tr>
<tr>
<td>Validation</td>
<td>0.60</td>
<td>0.43</td>
<td>0.87</td>
</tr>
<tr>
<td>Test</td>
<td>0.76</td>
<td>0.41</td>
<td>0.74</td>
</tr>
</tbody>
</table>
4. Conclusions and Discussion

This paper has investigated the field of emotion elicitation, and more concretely the use of artificial neural networks to classify emotions. In this case, we have focused on the outcomes of the API of the Emotiv EPOC+ headset after processing electroencephalogram signals. The emotional states calculated by the API have been compared with validated valence, arousal and dominance values from IAPS database.

The first step, prior to examining the ANN classification accuracy of the API, was to validate the fit of the responses given by sixteen participants after viewing IAPS images through a SAM questionnaire. The percentage of hits was demonstrated to be not as good as expected, that is 85.94, 79.69 and 78.13% for valence, arousal and dominance, respectively. This may be partially due to the selection of the images related to both low valence conditions, where the level of valence is likely too low.

In general, the selection of images is challenging in this type of experiments, as emotional elicitation depends on personal stereotypes [32]. In this case, the images selected could affect the results achieved in our experiment. Nonetheless, after using an ANN-based approach, up to 96% classification accuracy has been reached for the specific sixteen participants that have taken part in this experiment. Thus, although the images chosen could have some effect in our experiment, ANNs have mitigated the effect.

For the second and most important step, several multilayer perceptron ANN configurations were analysed to evaluate the emotional outcomes of the API of the Emotiv EPOC+ headset. This study demonstrated that multilayer perceptron is sufficient to validate the Emotiv EPOC+ API outcomes. It is not necessary to test with other more complex solutions (convolutional neural networks, deep learning, and so on) for facing the problem at hand. Hence, the main conclusion is that the emotional model implemented in the API offers 85% classification accuracy respect to the validated IAPS values. This result is in line with other research papers in the field of emotion recognition [32,33].

The proposed solution presents a series of advantages and disadvantages. An essential advantage is that multilayer perceptrons with backpropagation provide a simple solution to validate the emotional model implemented for the headset. As shown in this paper, the results obtained are excellent considering that the headset was designed for gaming and has a low price (compared to other devices). Another advantage is that the emotional states given by the headset’s API can be used with no further processing of the electroencephalogram signals acquired from the scalp.

On the other hand, as a clear disadvantage of using the API of the headset is that this software is not transparent. Therefore, it is not possible to further enhance the algorithms related to EEG signal processing. Another limitation of the study performed is the relatively low number of participants. A larger number of people involved in experimentation would guarantee statistically normalised data, closer to the IAPS database volunteers. Nonetheless, the already good results obtained with sixteen participants should be highlighted in this case.

Author Contributions: R.S.-R., A.S.G., L.F.-A., M.T.L. and A.F.-C. conceived and designed the experiments; R.S.-R., A.S.G., M.A.V.-Q. performed the experiments; R.S.-R., A.F.-C. and P.G. analyzed the data; R.S.-R., A.S.G., A.F.-C. and P.G. wrote the paper.

Funding: This work was partially supported by Spanish Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación (AEI)/European Regional Development Fund (FEDER, UE) under DPI2016-80894-R, TIN2016-79100-R and TIN2015-72931-EXP grants, by Castilla-La Mancha Regional Government/FEDER, UE under SAPL/17/180501/000192 grant, and by Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM) of the Instituto de Salud Carlos III. Roberto Sánchez-Reolid holds BES-2017-081958 scholarship from Spanish Ministerio de Educación y Formación Profesional.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.
Abbreviations
The following abbreviations are used in this manuscript:

API Application programming interface
ANN Artificial Neural Network
BCI Brain-Computer Interface
EEG Electroencephalography
EPOC Emotiv EPOC
IAPS International Affective Picture System
SAM Self-assessment manikin
MLP Multilayer perceptron
L-M Levenberg-Marquard
BR Bayesian Regularisation

References
13. Van Erp, J.; Lotte, F.; Tangermann, M. Brain-computer interfaces: Beyond medical applications. Computer 2012, 45, 26–34. [CrossRef]
© 2018 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/).