



Article

# Analysis of Personality and EEG Features in Emotion Recognition Using Machine Learning Techniques to Classify Arousal and Valence Labels

Laura Alejandra Martínez-Tejada \* , Yasuhisa Maruyama, Natsue Yoshimura and Yasuharu Koike

FIRST Institute of Innovative Research, Tokyo Institute of Technology, Yokohama 226-0026, Japan; maruyama.ya.e@m.titech.ac.jp (Y.M.); yoshimura@pi.titech.ac.jp (N.Y.); koike@pi.titech.ac.jp (Y.K.)

\* Correspondence: martinez.l.aa@m.titech.ac.jp

Received: 6 March 2020; Accepted: 11 April 2020; Published: 13 April 2020



**Abstract:** We analyzed the contribution of electroencephalogram (EEG) data, age, sex, and personality traits to emotion recognition processes—through the classification of arousal, valence, and discrete emotions labels—using feature selection techniques and machine learning classifiers. EEG traits and age, sex, and personality traits were retrieved from a well-known dataset—AMIGOS—and two sets of traits were built to analyze the classification performance. We found that age, sex, and personality traits were not significantly associated with the classification of arousal, valence and discrete emotions using machine learning. The added EEG features increased the classification accuracies (compared with the original report), for arousal and valence labels. Classification of arousal and valence labels achieved higher than chance levels; however, they did not exceed 70% accuracy in the different tested scenarios. For discrete emotions, the mean accuracies and the mean area under the curve scores were higher than chance; however, F1 scores were low, implying that several false positives and false negatives were present. This study highlights the performance of EEG traits, age, sex, and personality traits using emotion classifiers. These findings could help to understand the traits relationship in a technological and data level for personalized human-computer interactions systems.

**Keywords:** EEG; emotion recognition; human computer interactions; feature selection; machine learning; personality

## 1. Introduction

Emotions influence how people process information and make decisions, and they shape their behavior when they interact with their surroundings. When interactions between humans and systems occur, physical, cognitive, and social connections are integrated, including empathetic interactions to enhance users' experience in varied fields [1]. For new human-computer interaction (HCI) paradigms, in which systems are in constant contact with the users, it is important to identify and recognize users' emotional states to improve interactions between digital systems and the users with high recognition accuracy and provide a more personalized experience [2].

From an HCI perspective, it is important to find new ways in which systems can be more personalized to the user and to achieve better cooperation in fields like assistive and companion computing using physiological signals like electroencephalograms (EEG)—a useful tool that describe how cognition and emotional behavior are related at a physiological level [3–5]. Owing to the development of new technology and portable devices to measure EEG, the research is expanding beyond medical applications to areas like e-learning, commerce, entertainment, etc.

Research in emotion recognition using EEG as the main source of information has focused on how to achieve better performance and accuracy in the emotion identification and classification

process, considering different traits that are related on how emotions are managed and regulated. Demographic variables and personality characteristics are useful features to describe the relationship between emotional states and the individual characteristics related to behavior. Personality traits, age and sex are relevant to any computing area involving the understanding and predicting of human behavior [6,7]. From this, it is expected that demographic characteristics and personality traits will foster emotion recognition processes to achieve higher performance.

### 1.1. EEG and Emotion Recognition

Emotion is a psychological state that is accompanied by physiological changes that can lead to the modification of a person's expressions, which are observable and measurable manifestations and can be perceived and evaluated by others as evidence of an emotional state [8]. For the identification of emotional states by HCI systems, varied approaches are grouped under the term *emotion recognition*, which uses affective models (like discrete emotions labels or dimensional emotions scores; i.e., Russell's Affective Model [9]), and measurement methods to identify individuals' behavioral states, which are labeled as emotion or affective states. Emotion recognition can be examined by pattern extraction through machine learning techniques from signals like speech, body movement, and facial expressions, or physiological signals that describe individuals' behavior [10].

One of the physiological signals used for emotion recognition is EEG. EEG signals have gained increasing attention owing to its promise of potential applications in brain-computer interface (BCIs) for assistive technological solutions to overcome physical and speech disabilities. Emotion recognition using EEG signals focuses its development on two main application fields: first, medical applications designed to provide assistance, enhancement, monitoring, assessment, and diagnosis of human psychiatric and neurological diseases; and, second, non-medical applications designed to entertain, educate, and monitor emotional states in a commercial or personal context [11,12].

EEG signals are a powerful method for studying the brain's responses to emotional stimuli because its measurement equipment is noninvasive, fast, and inexpensive. EEG data lacks spatial resolution and requires several electrodes (around 8 to 128 depending on the experiment and the robustness of the equipment used) to be placed on the participant's head; however, it provides great time resolution, allowing researchers to study phase changes in response to emotional stimuli [13]. EEG signals are divided into specific ranges: delta (1–4 Hz), theta (4–7 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (> 30 Hz) bands. To analyze the signal contribution in each band, power spectral densities (PSD) and power spectral asymmetry (PSA) were used.

Regarding emotions recognition using EEG signals, some surveys revealed how EEG signals are related with emotional behavior and used varied methodologies and traits to classify arousal-valence space or discrete emotions [11,13,14]. The main findings revealed that EEG signals and emotions are related to the frequency analysis in different bands. For example, changes in alpha frequency band power and channel asymmetry between the front hemispheres of the brain are traits related to emotional activity: right frontal brain activation is associated with negative emotions, and greater left frontal activation is associated with positive emotions [15–17]. Furthermore, PSD have shown a frontal asymmetrical increase of theta and alpha activity related to the observation of pleasant (unpleasant) stimuli in the left (right) hemisphere [18,19]. Activity in the left frontal hemisphere is related to pleasant stimulus, and right frontal sites show EEG activations for unpleasant stimuli. Moreover, the scalp regions on the left frontal and pre-frontal areas are mostly activated when participants perceive pleasant content, and the right frontal lobe is more activated when people have been exposed to unpleasant content [18]. For discrete emotions, gamma frequency band is related to happiness and sadness [20], as are the alpha waves of the temporal lobe (left for sadness and right for happiness) [21]. However, a limitation with these techniques is that PSD and PSA are related to frontal activation, which can be related to other higher cognitive functions like concentration, planning, judgment, creativity, and inhibition [22].

### 1.2. HCI and Personality Traits

Personality is a behavior pattern that maintains over time and context, differentiating one person from other [23]. To measure and identify an individual's personality, there are a wide variety of psychometric tests that correspond to various psychological theories about behavioral patterns and innate characteristics. One of these theories is Eysenck's personality model, which posits three independent factors to describe personality: psychoticism, extraversion, and neuroticism (PEN) [24]. Another widely used personality model is the five factor model (FFM). Paul Costa and Robert McCrae [25] devised the FFM and the Neuroticism-Extraversion-Openness Five-Factor Inventory questionnaire. They posited that there are five personality traits: extraversion (social vs. reserved), agreeableness (compassionate vs. dispassionate and suspicious), conscientiousness (dutiful vs. easy-going), neuroticism or emotional stability (nervous vs. confident), and openness to experience (curious vs. cautious). These traits describe the frequency or intensity of feelings, thoughts, or behaviors of an individual compared with other people. From this model, individuals' personality is described through these traits in some varying degree. In contrast, Jeffrey Alan Gray built the Bio-psychological Theory of Personality (behavioral inhibition system (BIS)/ behavioral activation system (BAS)) [26]. This is a model of the general biological processes relevant to human psychology, behavior, and personality. This model describes the existence of two brain-based systems for controlling a person's interactions with their environment: the BIS and BAS.

The FFM is considered a standard in science; however, based on statistical analysis of the data, some researchers argue that the FFM should be expanded to include a sixth trait: honesty-humility. The HEXACO model of personality conceptualizes human personality considering this sixth dimension: honesty-humility (H), emotionality (E), extraversion (X), agreeableness (A), conscientiousness (C), and openness to experience (O). Although the HEXACO model has received growing support from scientists, the necessity of this sixth trait is still a matter of debate [27].

Emotion – personality relationships have been studied in psychology for a long time; however, their relationship in HCI is still under discussion. Some studies used varied kinds of behavioral data to identify emotional states, for example, using participants' text and annotation [28,29], identifying personal characteristics in voice and speech [30–32], body language [33,34], and also from digital footprints [35,36]. In the next section, we present some works that aimed to include personality information in different HCI applications.

### 1.3. Related Works

In [30], an approach to detect an user's interaction style in spoken conversation was presented. It combined emotional labeling of conversation-based affective speech corpus of 53 students from both sexes and International Personality Item Pool to measure personality for intelligent speech-based HCIs. Callejas-Cuervo and colleagues [37] proposed a system architecture where videogames can stimulate participants to extract characteristics that can correlate with information from emotion and personality traits, using electrocardiogram (ECG), galvanic skin response (GSR), electromyography (EMG) signals, and the PEN model together with Russell's affective model. Furthermore, in [28], the standard cognitive appraisal model (OCC emotion model) and the FFM personality model were combined in a natural language processing tools to analyze language for affect. In [38], researchers proposed an intensity-based affective model that incorporates the FFM for personality and the OCC model for affect, from predetermined answers related to image and labels; then, they performed personality processing and modeling to predict emotion.

Guo and Ma [35], proposed a modeling personality system from big data coming from different sources including participants' location, heartbeat rate, and browser data to describe person's conditions and activity to accurately identify participants' personality. They proposed a human model of four layers: state, pattern related to daily activity, emotions, and personality. Wei and colleagues [39] focused their attention on the "apparent personality analysis"—using short human-centered video sequences and developing an algorithm for recognizing personality traits from those videos — using

deep modal regression. In [40], emotion recognition and varied characteristics (i.e., personality traits, age, and sex) were used to create a car interphase that takes actions when it identifies an emotional reaction (neutrality, panic/fear, frustration/anger, and boredom/sleepiness) in participants using GSR, temperature, and heart rate. At the time of the publication, one major concern was scenarios in which the emotion occurred, which the system could not identify. They noted that the incorporation of demographic characteristics and personality traits can enhance and increase the accuracy of emotion recognition; however, no data involving these were provided.

Robot interaction with people is another vast field of interest for emotion recognition and personality traits. Anzalone and colleagues [41] examined extraversion in human-humanoid interactions using nonverbal behavior (i.e., upper-body movements and interaction duration). Additionally, Bhin and colleagues discussed building an automated psychophysical personality data acquisition system for human-robot interaction under the premise that, to build more natural interaction between human and robots, systems need the ability to recognize the psychological state (i.e., personality) of users [42]. They proposed a system for personality recognition using nonverbal cues through a commercial webcam to record participants' body movement and facial expression, a microphone to record speech signals, a wristband (Microsoft Band or Empatica E4) to obtain physiological signals such as heart rate and body temperature, and the FFM (BFI-K-44) to measure personality.

The influence of personality traits in works regarding affective computing and HCI has been expanding in recent years, which benefit from examining emotion, mood, and personality by employing physiological data, facial and audio recognition, body movement, etc. Works that use EEG signals as a main source of information have been increasing in recent years. The following section summarizes their findings.

#### 1.4. EEG-Related Works

Cai and colleagues [43] evaluated the behavior and personality of 42 participants using physiological data from wearables devices that measured heart rate, respiration, and EEG while participants watched a 20-minute video or gave an 8-minutes presentation. Their main objective was not to predict personality or emotion, but to correlate these two characteristics and analyze the relationship between personality traits and behavior through the influence of emotional states. They used Pearson's correlation coefficient to determine the relationship between respiration rate and personality traits, and Spearman's rank correlation coefficient to determine the relationship between facial expressions and personality traits under different emotion states. They yielded evidence of said correlations; however, their results were not definitive.

Rukavina and colleagues [2] examined personality, sex, age and gender roles to improve the emotion recognition accuracy. Age and personality dimensions were correlated with all extracted features during each of the five affective states or core affects from the valence, arousal, and dominance space: for each core affect, they presented two blocks of 20 pictures, 10 pictures in a 2-minute time window (20-second fixation cross, 20-seconds picture presentation belonging to the same core affect, and 20-second fixation cross), using a total of 100 pictures. First, they performed a correlation analysis to consider only meaningful variables for the classification analysis. They concluded that sex and age were significantly correlating with affective states; however, they did not find a correlation between personality traits and affective states. One possible explanation was the high significance level of the Bonferroni correction ( $p < 0.007$ ). Although, their experiment was limited because the time and methods of presenting the emotional stimuli material to the participants can affect the outcome.

Miranda-Correa and Patras [44] propose a multi-task cascade deep learning approach and performed binary classification for emotional states (arousal and valence) and personal factors (personality, mood, and social interaction) from EEG signals. Forty participants watched short affective videos and thirty-seven participants watched long videos (ranging from 51 seconds to nearly 24 minutes), in individual or group sessions. The researchers analyzed the time and frequency domain features

from EEG data trough segments (20 second time window), to obtain the affective levels (arousal and valence) using convolutional neural networks (CNNs) and recurrent neural networks (RNNs); then, they estimated the Big Five factors and their relationship with mood; i.e., the Positive Affect and Negative Affect Schedule (PANAS) from  $n$  consecutive segments' affective levels using a second deep network with recurrent layer of eight units with "sigmoid" output function. Using the fusion affect sub-network (from CNNs and RNNs), they achieve 0.59 and 0.61 F1-scores ( $p < 0.001$ ) for valence and arousal recognition, respectively.

Mittermeier and colleagues [45], studied if there is an emotion-specific neural correlation between positive and negative auditory emotional stimuli and attention through auditory-evoked potentials (AEPs) and if there is a specific relationship between AEPs evoked by emotional stimuli and the personality dimension extraversion–introversion. Differing from the other studies, this work focusses on the auditory emotional stimuli to analyze the correlation between reaction times to the stimuli, evoke potentials, and personality (extraversion). They found that extraversion correlated with the EPN 170 amplitude in the emotional paradigms. Compared to participants in the introverted subgroup, extroverted persons exhibited significantly higher EPN 170 amplitudes in the P3 channel for emotional paradigms (syllables (Pz channel) and words (P3 channel)) but not in the tones task.

Subramanian and colleagues [46] built a multimodal database from 58 participants for implicit personality and affect recognition using commercial physiological sensors to understand the relationship between emotional attributes and personality traits and characterize both by physiological responses. The paper described the influence of personality differences on users' affective behavior using the ASCERTAIN database to understand the relationship between emotional attributes from an arousal/valence model and the Big Five personality model by measuring users' physiological responses. Their main goal was to assess personality traits via affective physiological responses instead of questionnaires. They compiled valence and arousal ratings reflecting user's affective impression: a seven-point scale was used with a -3 (*very negative*) to 3 (*very positive*) scale for valence, and a 0 (*very boring*) to 6 (*very exciting*) scale for arousal. Ratings concerning engagement (*did not pay attention to totally attentive*), liking (*I hated it to I loved it*), and familiarity (*never seen it before to remember it very well*) were also acquired with the five traits from the FFM. They found that arousal was moderately correlated with extraversion, while valence correlated strongly with linking ( $0.68 p < 0.05$ ). GSR features obtained higher recognition performance for both arousal and valence (0.68 with Naïve Bayer's (NB) classifier), while ECG features obtained the worst recognition performance (0.56 for valence and 0.57 for arousal using Support Vector Machine (SVM)). EEG features had better performance recognizing arousal (0.61) as compared to valence. GSR, ECG and EMG achieved better recognition of valence. Peripheral (ECG+GSR) features performed better than unimodal features for arousal recognition, while the best multimodal F1-score (0.71) was obtained for valence. Finally, comparing the two employed classifiers, NB achieved better recognition performance than linear SVM for arousal (0.69 using peripheral signals) and valence (0.68 for GSR signal).

Mueller and Kuchinke [47] examined the individual differences in implicit processing of emotional words (happy, neutral, and fear-related) on a lexical decision task; i.e., deciding whether a letter string is a correct word or a non-word via a button. They argued that several participant-specific variables (personality traits and neurological foundation) are known to modulate processing of emotional information. The main task comprised 35 trials in pseudo-randomized order displaying faces of five individuals in each of different emotional expressions. A correlation analysis was performed between happy and neutral, happy and fear-related, and fear-related and neutral. Difference scores were calculated for response times (RTs), error rates (ERs), and drift rates (DRs), which correlated with all nine variables of emotion processing (RTs, ERs, and DRs for Happy–Fear, Happy–Neutral, and Fear–Neutral). Additionally, they performed three multiple linear regression analyses for RTs, ERs, and DRs as dependent variables to predict individual emotional effects. Results revealed that BAS-Drive was the variable that explained most of the variance regarding Happy to Fear RT (H-F<sub>RT</sub>) differences. RTs for happy words were generally shorter than RTs to fear-related words resulting

in negative difference scores on average. The negative relationship between H-F<sub>RT</sub> differences and BAS-Drive scores revealed that participants with larger BAS-Drive scores showed greater H-F<sub>RT</sub> differences. Instead, BAS-Drive scores were positively correlated with Fear to Neutral RTs.

Although the literature shows a relationship between emotion, mood, affective states, and personality [48–50], how to effectively use demographic characteristics and personality traits to ascertain emotion recognition remains unclear. There is no standard for choosing emotion or personality models in recognition techniques, and variables and classification approaches differ between studies, thus yielding inconsistent results. This is understandable given the newness of this field, and each outcome offers novel insight into new approaches that can be developed. The presented papers do not provide conclusive information about a strong correlation between emotion classification and demographic characteristics and personality traits. It is still not clear how singular personality traits can be measured from psychological signal and emotion, even though literature says they correlate at biological level [24].

For this study, we aim to test the hypothesis that age, sex, and personality traits, can improve the classification accuracies for arousal and valence levels, when they are used alongside EEG data for emotion recognition processes by machine learning algorithms. Using the information from the AMIGOS dataset [51], we analyzed (1) the contribution of the different EEG traits, demographic characteristics, and personality traits in the classification process of arousal, valence, and discrete emotions labels using varied machine learning techniques, (2) the contribution of the demographic characteristics and personality traits in emotion classification, as relevant information related to behavior and individual characteristics, and (3) the performance of simple classification models with new EEG traits that were not considered in the AMIGOS study.

## 2. Materials and Methods

According to the brain-computer interface cycle, where it is the common approach to perform emotion recognition using EEG signals, we followed the basic phases proposed in [52] and used across different works [53–55] to analyze the performance of the different classifiers implemented. The first phases related to experiment implementation where participants are exposed to the emotional stimuli, recording of the EEG signals and preprocessing of the raw data, which are being retrieved from the dataset AMIGOS. The phases related to feature extraction and classifier implementation are being performed by the authors of this work.

### 2.1. AMIGOS Dataset Experiment

AMIGOS is a dataset to study the relationship between affect, personality, and mood [51]. The dataset consists of multimodal recordings of participants and their responses to fragments of emotional videos. Participants took part in two experimental setups while watching long and short videos: first, in an individual scenario, and, second, in a group scenario with other participants. While watching the videos, EEG, ECG, GSR, frontal high-definition video, and both RGB and depth full body videos were recorded. Personality (Big-Five), mood (PANAS), internal annotation (participants' self-assessment affective levels), and external annotation (off-line annotations by three annotators; valence and arousal scales) scores were obtained. The participants read and signed a consent form to take part in the study.

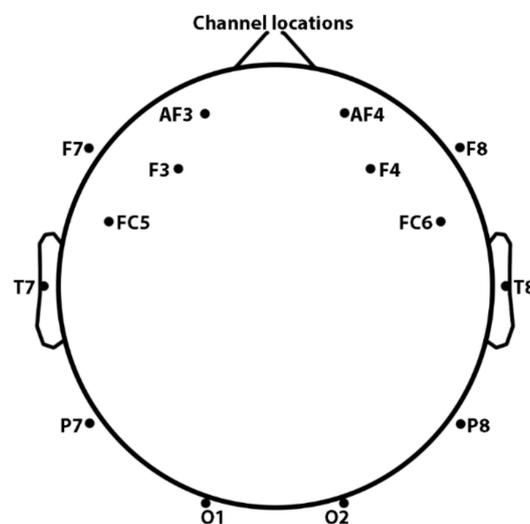
From the AMIGOS dataset, we used information from the individual–short videos scenario, in which 40 participants (male = 27, female = 13, aged 21–40 years, mean age = 28.3 years) watched 16 videos (duration < 250 s)—four from each high and low arousal–valence emotional levels combination: high arousal and high valence (HAHV), high arousal and low valence (HALV), low arousal and high valence (LAHV), and low arousal and low valence (LALV). The experiment consisted of an initial self-assessment session for arousal, valence, and dominance scores, as well as a selection of basic emotions (neutral, happiness, sadness, surprise, fear, anger, and disgust) that participants felt before any stimuli were shown. Next, 16 videos were presented in a random order in 16 trials, each consisting

of (1) a five-second baseline recording showing a fixation cross; (2) The display of one video; and (3) self-assessment of arousal, valence, dominance, mood, liking, and familiarity, as well as the selection of basic emotions. After the 16 trials, the recording session ended.

## 2.2. AMIGOS Features

For the input features, we used the 14 EEG signals from Emotiv EPOC Neuroheadset as information source, which were recorded at 128-Hz sample rate and 14-bit resolution (electrode distribution is shown in Figure 1). We used the demographic characteristics (age and sex) and personality traits which were acquired before the experiment using an online form.

- For EEG features, we used the preprocessed signals from the AMIGOS dataset. The signals were averaged to the common reference, filtered with a band-pass frequency filter from 4.0 Hz to 45 Hz, EOG removal was applied and then segmentation was performed. We calculate the 105 EEG features reported in the AMIGOS experiment, which correspond to PSD and PSA between pairs of electrodes. PSD corresponds to the five bands correlated with emotion response: theta (3–7 Hz), slow alpha (8–10 Hz), alpha (8–13 Hz), beta (14–29 Hz), and gamma (30–47 Hz) for each electrode (70 features). PSD was obtained by Welch’s method (time window = 128 samples corresponding to 1 second) between 3 and 47 Hz and averaged over the frequency bands. PSA was calculated between each of the seven pairs of electrodes in the five frequency bands correlated with emotion response (35 features). These pair of electrodes comprised two electrodes located in the same scalp region, but on the opposite side of the head: AF3/AF4, F3/F4, F7/F8, FC5/FC6, T7/T8, P7/P8, and O1/O2.
- We also utilized age, sex, and the Big Five personality traits [56] (i.e., 7 features).



**Figure 1.** Electroencephalogram (EEG) channel location according to the International 10–20 System.

A total of 112 features from the AMIGOS dataset were used in this study.

## 2.3. Added EEG Features

For this study, we calculate the fractal dimension (FD) and the differential entropy (DE) for each one of the electrodes in the five frequency bands mentioned before. Moreover, the rational asymmetry (RASM) and differential asymmetry (DASM) for each of the seven pairs of electrodes in the five bands were calculated (70 features). Because in previous literature [57–59], these EEG traits are related to participants’ emotional responses, and reports about EEG emotion recognition used the same kind of features to obtain classification above the chance level [14]. We wanted to include these EEG traits

to analyze if they can improve the classification performance in contrast with the ones used in the AMIGOS work.

- FD is a measure of signal complexity. Because EEG signals are nonlinear and chaotic, a FD model can be applied in EEG data analysis [60]. We compute FD using the Higuchi algorithm for each of the 14 EEG signals (14 features).
- DE can be defined as the entropy of continuous random variables and is used to measure its complexity [61]. DE is equivalent to the logarithm of the energy spectrum (ES) in a certain frequency band for a fixed length EEG sequence [62]. We calculated ES as the average energy of EEG signals in the different five frequency bands for each electrode and applied the logarithm to obtain DE (70 features). DASM and RASM were calculated as the differences and ratios between the DE of the seven pairs of asymmetry electrodes (35 features for each trait).

In total, we added 154 features from EEG signals to complement the ones already obtained in the AMIGOS base experiment, thus, 266 features were used in the emotion classification models.

We applied feature selection methods [58] to analyze how the different features are related with the classification labels and to obtain a reduced set of features (from the total 266 features), to analyze the improvement in the classification performance. We applied feature importance to analyze in how much percentage the features contribute to predict the different label scenarios. Also, we implemented univariate selection and a recursive feature elimination with cross validation to select the features that improve the classification rates and built a second set of features.

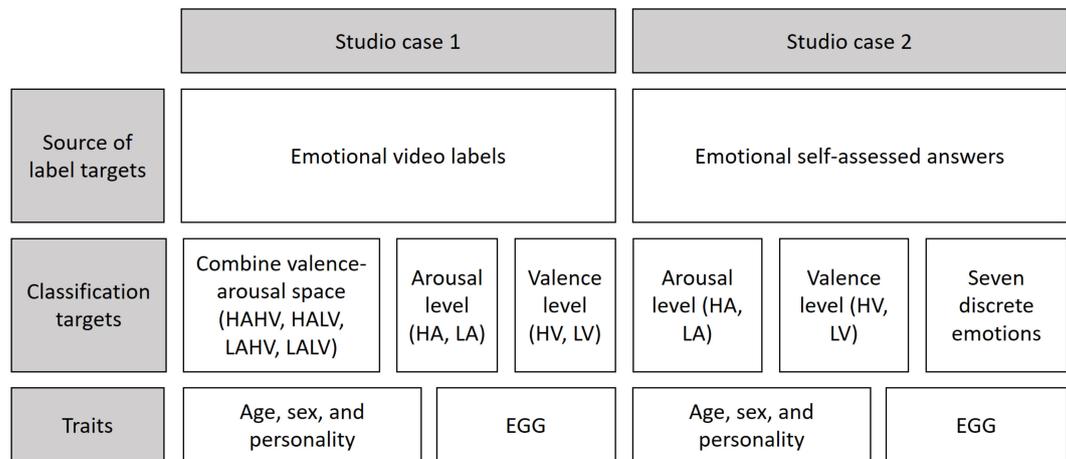
#### 2.4. Classifiers

We wanted to focus our analysis in two different studio cases: first, analyzing the classification performance of different machine learning algorithms using all 266 features to classify the emotional stimuli video labels (arousal and valence levels). In this case, our motivation was to analyze what features can predict the videos emotional labels based on participants' personal data. This would help us identify to what degree it is possible to classify emotions when the self-assessment arousal and valence scores from the participants are not available. Second, We analyzed the classification performance using only EEG data and participants' sex, age, and personality traits to classify self-assessment emotional answers using self-assessment manikins [63] and the seven basic categorical emotions, which were reported by the participants at the end of each video. Our motivation in this second case was to analyze the performance of different classifiers when using only information related to EEG signals and characteristics like age, sex, and personality.

For the first studio case (Figure 2a), we tested three different classification scenarios in which we select two sets of input features. For the classification scenarios, we considered the labels from the videos used as emotional stimuli: first, we combined valence-arousal space labels (HAHV, HALV, LAHV, and LALV); second, we considered arousal labels (HA and LA); and third, we considered valence space labels (HV and LV). To transform the arousal and valence responses into classification labels, we use a threshold of 5.0 to convert the response values into binary labels to obtain categorical data. For the input features, we considered a first set of features with only EEG data, demographic characteristics and personality traits (266 features) and a second set of features with EEG data, demographic characteristics and personality traits reduced using feature selection. From the 640 AMIGOS short-videos observations (16 videos × 40 participants), we exclude the observations that had missing personality and EEG data.

For the second studio case (Figure 2b), we tested 9 different classification scenarios corresponding to the different self-assessment traits related directly to emotions (arousal and valence labels and the seven emotions). To transform the arousal and valence responses into classification labels, we use a threshold of 5.0 to convert the response values into binary labels (HA and LA; HV and LV) to obtain categorical data. For the input features, we considerate: a first set of features with EEG data and demographic characteristics and personality traits (266 features), and a second set of features with EEG and demographic characteristics and personality traits reduced using feature selection. From the

640 AMIGOS short videos observations (16 videos  $\times$  40 participants), we exclude the observations that had missing personality and EEG data.



(a) Studio case 1 related to the emotional video labels as classification target.

(b) Studio case 2 related to the emotional self-assessment response labels as classification target.

**Figure 2.** Studio cases for emotional classification. Classification targets: high arousal (HA) and low arousal (LA), high valence (HV), and low valence (LV).

The classifiers were chosen to test and compare the emotion recognition accuracy using simple machine learning models.

- SVM is a linear model that use a decision boundary as a linear function to separate two classes with a line, a plane, or a hyperplane, fitting two parameters: regularization or margin maximization (C), and kernel. C determines the strength of the regularization. Higher values of C correspond to less regularization, trying to fit the training set as best as possible to each individual data point. With lower values of C, the algorithms will try to adjust to the majority of the data points. Kernels are mathematical functions that take data as input and transform it into the required form (i.e., linear radial basis function).
- Naïve Bayes is faster than linear models by looking at each feature individually, collecting simple per-class statistics from each feature.
- Random Forest, is a collection of decision trees, where each tree is slightly different from the others. With many trees (estimators) it is possible to reduce the overfitting by averaging the results of each tree. And with the tree deepness it is possible to splits the tree capturing more information about the data.
- Artificial neural network is a multi-layer fully-connected neural nets that consist of an input layer, multiple hidden layers with units, and an output layer. Each layer has an activation function to discriminate the data (i.e., relu, sigmoid).

Our goal was to identify if the accuracy improved in any of the scenarios using different feature sets compared with the accuracy reported in AMIGOS work using PANDAS framework under python language. For the combine valence–arousal space label scenario, we applied SVM with linear (C = 100) and radial basis function (RBF) kernel (C = 100, gamma = 0.1). For the other scenarios we applied SVM with linear (C = 100) and RBF kernel (C = 100, gamma = 0.1), Naïve Bayes, Random Forest (estimators = 2000, max\_depth = 300) and an artificial neural network (ANN) with 134 hidden units, one “relu” activation function hidden layer, and, for the output layer we used a “sigmoid” activation function (optimizer = “rmsprop”, batch size = 32, epochs = 100). Parameters were tuned using grid search with

cross-validation. To evaluate the classifier accuracy, we obtained the mean accuracy, mean F1, and mean area under the curve (AUC) scores using a 10-fold cross-validation approach over the training set of features (75% of all the dataset).

### 3. Results

Feature construction and feature selection are key steps in the data analysis process—in most cases, conditioning the success of any machine learning endeavor [64]. Previous works have shown how applying feature selection process in emotion recognition tasks using EEG traits [57,58], increases the performance of the classifiers while the computational power is reduced. For the purpose of this work, we wanted to perform feature selection process to reduce the number of features, preventing overfitting and improving the classification process.

Feature selection methods can generally be divided into filter and wrapper methods. While wrapper methods select features based on interaction with a classifier, filter methods are model-independent [58]. Filter techniques assess the relevance of features by looking only at the intrinsic properties of the data. Advantages of filter techniques are that they easily scale to very high-dimensional datasets, they are computationally simple and fast, and they are independent of the classification algorithm. In this case, feature selection needs to be performed only once, and then different classifiers can be evaluated [65].

For feature selection, we used different feature selection filter approaches to understand how they affect the overall emotion classification process. We analyzed how the features contribute in percentage to predict the different label scenarios using feature importance selection. We also performed univariate selection and a recursive feature elimination (RFE) with cross-validation to select the features to build the second set of features [66].

#### 3.1. Feature Selection and Analysis for EEG Data, Demographic Characteristics, and Personality Traits to Predict Video Emotional Labels

##### 3.1.1. Feature Importance

Feature importance [66] provides a percentage score for each feature of the dataset, the higher the score, the more important or relevant is the feature towards the output variables—using forests of trees to evaluate the importance of features on a classification task and identify the features more related to each of the labels. The EEG traits contribute around 0.5%–0.3% percent to each of the different scenarios. In contrast, the importance percentage of the personal trait labels have the lowest scores. Table 1 shows the scores of demographic characteristics, and personality traits for the different scenarios, which do not exceed 0.32%, implying that they are not relevant to the classification process.

**Table 1.** Feature importance score (%) from sex, age, and personality traits for video emotional labels classification.

Arousal–Valence		Arousal		Valence	
Feature	Score (%)	Feature	Score (%)	Feature	Score (%)
Agreeableness	0.2936	Agreeableness	0.2936	Emotional stability	0.3120
Extroversion	0.2752	Extroversion	0.2752	Agreeableness	0.2907
Emotional stability	0.2636	Emotional stability	0.2636	Conscientiousness	0.2896
Age	0.2557	Age	0.2557	Extroversion	0.2790
Creativity/openness	0.2550	Creativity/openness	0.2550	Age	0.2779
Conscientiousness	0.2304	Conscientiousness	0.2304	Creativity/openness	0.2613
Sex	0.2054	Sex	0.2054	Sex	0.2316

##### 3.1.2. Univariate Selection

When a finite training sample is provided, the statistic of the relevance is assessed by performing a statistical test with null hypothesis, “H0: the feature is individually irrelevant”; that is, X and

Y are statistically independent. Feature selection based on individual feature relevance is called univariate [67]. In univariate selection each feature is considered separately, intended to select single variables that are associated in most degree with the target variable according to a statistical test. The advantage of this technique is that it is fast and scalable; however, it ignores feature dependencies. Higher scores and *p*-values indicate that the variable is associated and consequently it is useful to the target [68].

Using the univariate feature selection algorithm propose by [66], we obtained the best features based on an analysis of variance, *F*-test, and *p*-value of the features related to the three arousal and valence labels scenarios, selecting 10% of significant features [68]. Inspecting the features, we found that, for valence–arousal scenario, only one EEG trait was selected for the algorithm—PSD from EEG channel AF4 in the theta band. For the other scenarios, no features were selected.

### 3.1.3. RFE with Cross-Validation

RFE with cross-validation is a RFE with automatic tuning of the number of features selected, it returns the most suitable features based on SVM classifier with linear kernel. For RFE, the SVM will be retrained several times with a decreasing number of features [64,68]. The features selected differed from the ones identified by the feature importance and the univariate selection algorithms because, in the RFE, an external estimator assigns weights to features. This estimator is trained on the initial set of features and the importance of each feature is obtained by a coefficient attribute; then, the least important features are discarded from the current set of features. That procedure is recursively repeated on the discarded set until the desired number of features is eventually reached [66].

Performing RFE with personal and EEG traits, we obtained 3 features for valence–arousal label, 15 features for arousal label, and 1 feature for valence. In this case, no demographic characteristics nor personality traits were selected by the algorithm. For valence–arousal label, AF3/AF4 PSA index in the beta band; DE in the theta band channel AF4; and FC5/FC6 RASM in the slow alpha band were selected by the RFE. For arousal label, PSD from slow alpha (AF3), alpha (AF3), and gamma (FC5) bands; F3/F4 and F7/F8 PSA index in the theta band; FD of channel P8; DE in the theta (AF3, O1) and gamma (T8) bands; RASM in theta (T7/T8), slow alpha (F3/F4, O1/O2), beta (FC5/FC6) and gamma (F7/F8) bands were selected by the RFE. For valence label only, DE from F4 channel in the beta band was selected.

Finally, for the second set of features we built one dataset combining the results from the univariate selection and the RFE feature selection process to determine how the performance of the classifiers behave in contrast to the original sets of traits.

For valence-arousal label, information of the frontal (AF3, AF4, FC5, FC6) region of the scalp were selected in theta, slow alpha and beta bands. For arousal label, information of the frontal (AF3, F3, F4, FC5, FC6, F7, F8), temporal (T7, T8), and occipital (O1, O2) regions of the scalp were selected. In general, PSD, PSA DE and RASM features were commonly selected for valence-arousal and arousal labels.

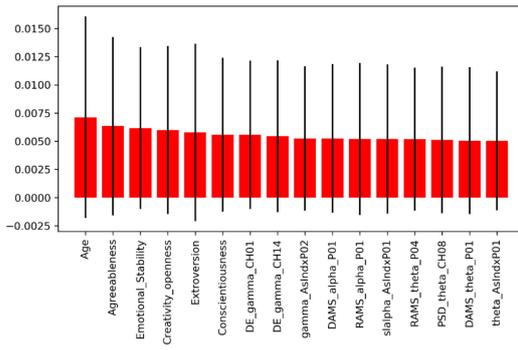
## 3.2. Feature Selection and Analysis for EEG, Demographic Characteristics, and Personality Traits to Predict Self-Assessed Traits Labels

### 3.2.1. Feature Importance

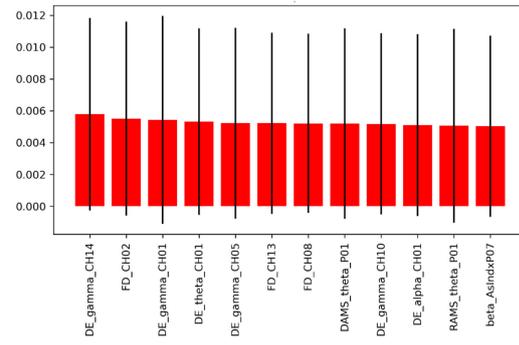
In Figure 3, we show the features that exceeded 0.5% of importance for each of the classification labels. The red bars are the feature importance of the forest, along with their inter-trees variability. In Table 2, we show the notation for the EEG channels and pair of electrodes used in Figure 3.

**Table 2.** (a) Notation for the EEG electrodes. (b) Notation for the EEG pair of electrodes.

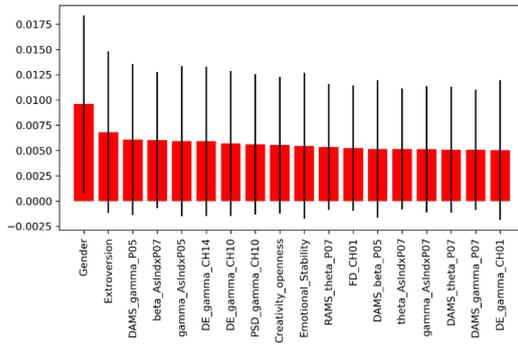
(a)														
n	01	02	03	04	05	06	07	08	09	10	11	12	13	14
Channel	AF3	F7	F3	FC5	T7	P7	O1	O2	P8	T8	FC6	F4	F8	AF4
(b)														
n	01		02		03		04		05		06		07	
Pair	AF3/AF4		F3/F4		F7/F8		FC5/FC6		T7/T8		P7/P8		O1/O2	



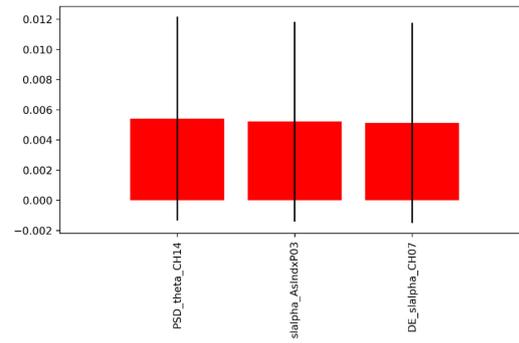
(a) Arousal



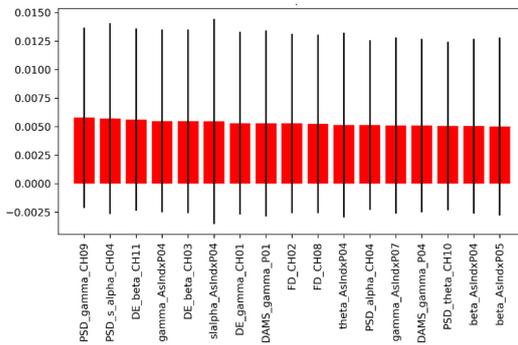
(b) Valence



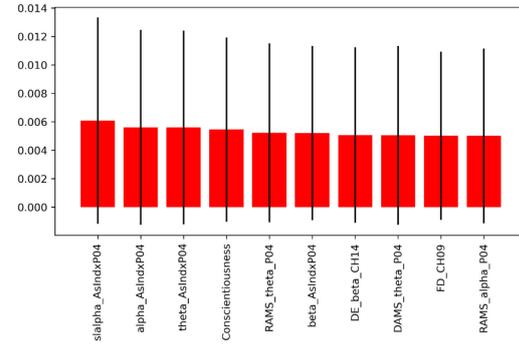
(c) Sadness



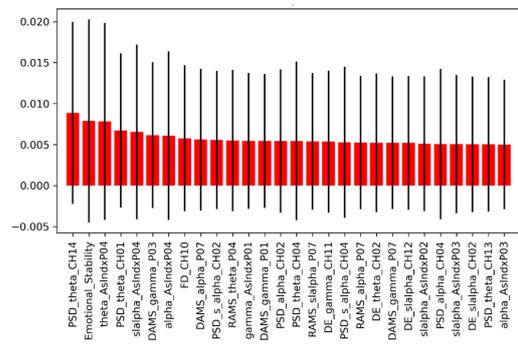
(d) Fear



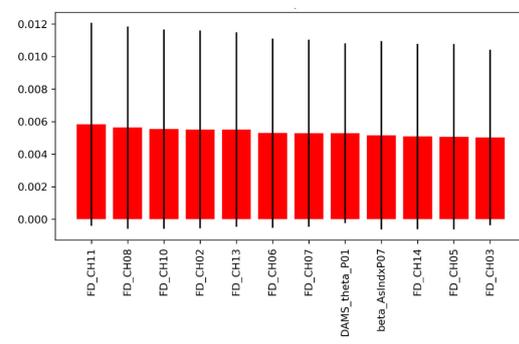
(e) Happiness



(f) Neutral

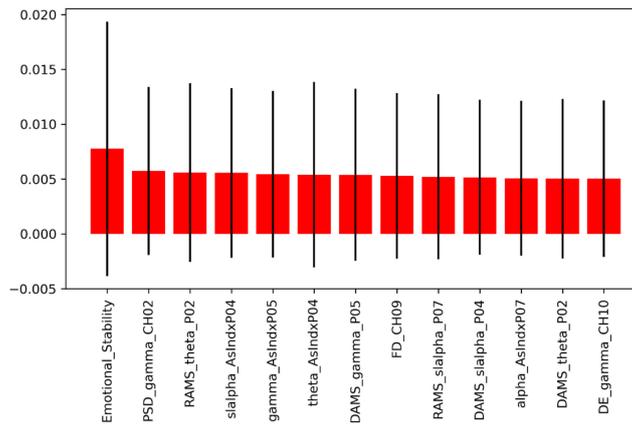


(g) Disgust



(h) Anger

Figure 3. Cont.



(i) Surprise

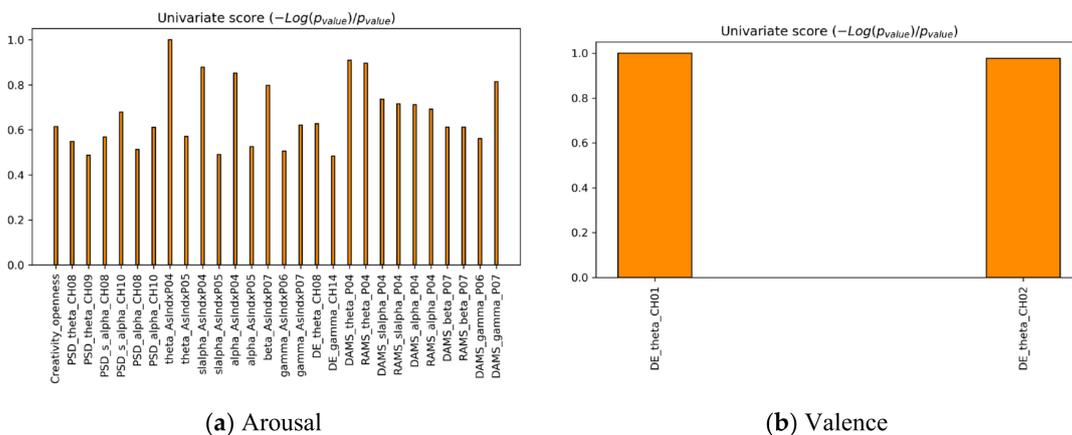
**Figure 3.** Feature importance with first set of features for self-assessed emotional labels classification. On the x-axis, the list of features is represented; on the y-axis, the importance of each feature in percentage is shown.

For arousal label, age, agreeableness, emotional stability, openness, extraversion, and conscientiousness were selected as important features. For the sadness label, sex, extraversion, openness, and emotional stability were selected as important features. For the neutral label, conscientiousness was selected as the important feature. For disgust and surprise labels, emotional stability was selected as an important feature. However, the contribution is still under 0.5%, which is too low compared with the other traits.

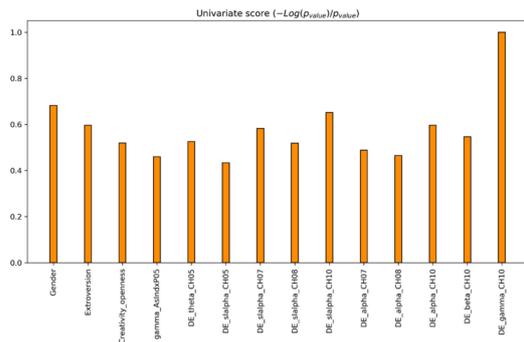
### 3.2.2. Univariate Selection

In Figure 4, we show the ratio for the most significant features. In Table 2, we show the notation for the EEG channels and pair of electrodes used in the Figure 4.

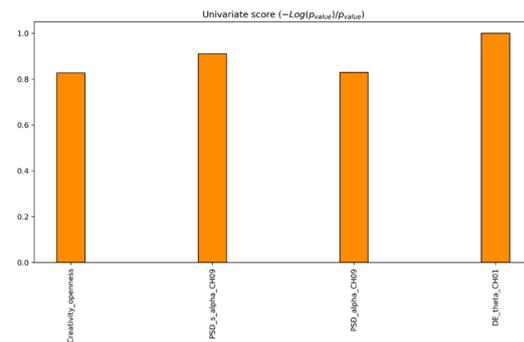
Figure 4 shows that the following demographic characteristics and personality traits were selected: arousal (openness), sadness (sex, extraversion, and openness), fear (openness), surprise (extroversion and emotional stability), disgust (agreeableness and emotional stability), and neutral (conscientiousness and emotional stability).



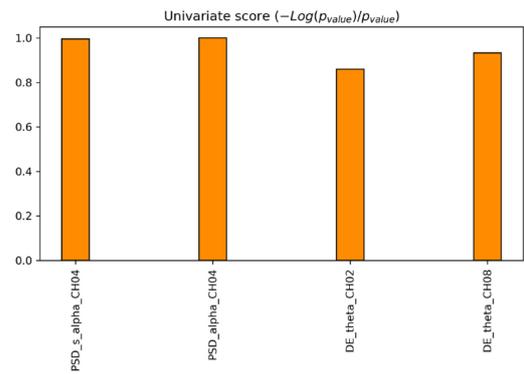
**Figure 4.** Cont.



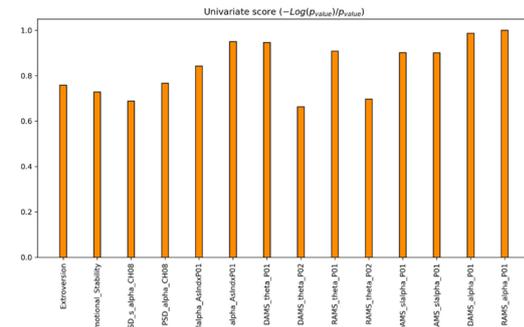
(c) Sadness



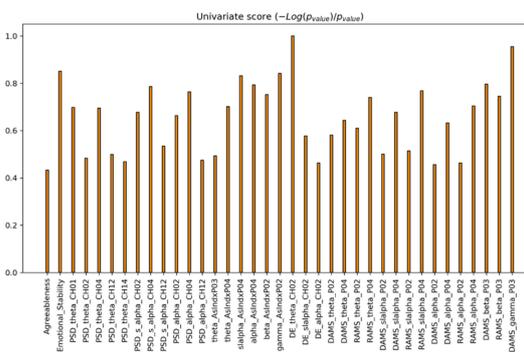
(d) Fear score values  $p < 0.1$  (none  $p < 0.05$ )



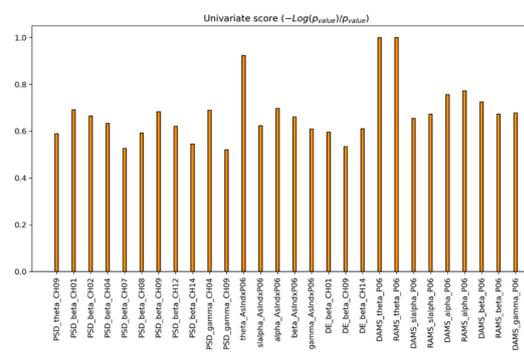
(e) Happiness



(f) Surprise

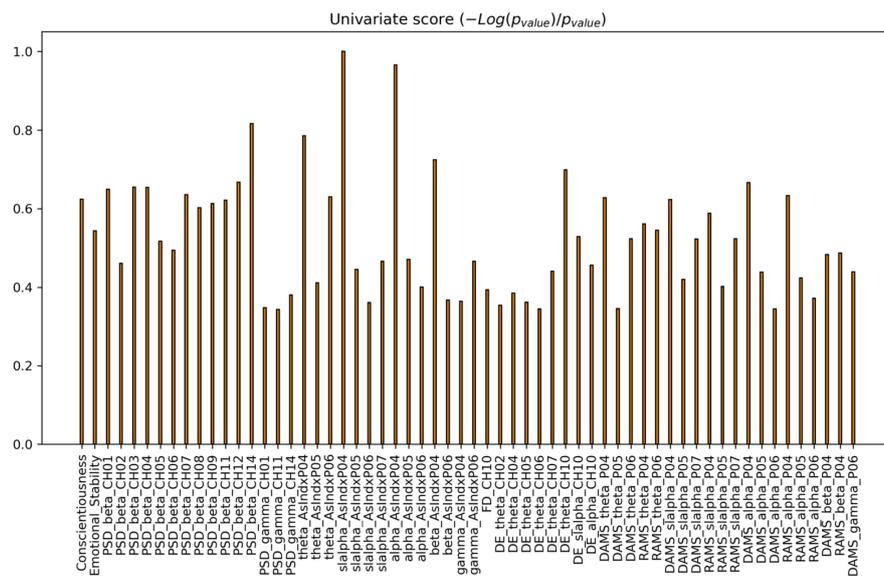


(g) Disgust



(h) Anger

Figure 4. Cont.



(i) Neutral

**Figure 4.** Univariate selection features with first set of features for self-assessed emotional labels classification. On the x-axis, the list of features is represented; on the y-axis, univariate score is shown.

For the arousal label, the EEG traits selected were: PSD in the theta (O2, P8), slow alpha (O2, T8), and alpha (O2, T8); PSA index for FC5/FC6, and T7/T8 in the theta, slow alpha, and alpha bands, for O1/O2 in the beta band, and for P7/P8 and, O1/O2 in the gamma band; DE in theta (O2), and gamma (CH14); and DASM in the theta, slow alpha, and alpha bans for FC5/FC6, in beta for O1/O2, and gamma band for P7/P8, and O1/O2. For valence label, important EEG features selected were: DE for AF3, and F7 in the theta band.

### 3.2.3. RFE with Cross-Validation

In this case EEG traits were selected for the nine different scenarios. No demographic characteristics and personality traits were selected by the algorithm:

- We obtained 16 features for arousal label: PSD in slow alpha (AF3, T8) and gamma band (FC6), PSA index in the theta (FC5/FC6), alpha (T7/T8), and gamma (FC5/FC6) bands; and DE in the theta (F3, T7, O1, O2, F4), slow alpha (P8, AF4), alpha (T7), beta (FC6), and gamma (AF3) bands.
- We obtained 40 features for valence label: PSD in theta (P7, T8, AF4), slow alpha (AF3, T8), alpha (O1, T8), beta (T8, FC6), and gamma (T8, F8, AF4) bands, PSA index in the theta (F7/F8), slow alpha (AF3/AF4, F7/F8, T7/T8, O1/O2) alpha (FC5/FC6), and beta (F7/F8, O1/O2) band; FD in FC5, T7, O2 channels; DE in theta (F7, F3, F4), beta (F3, FC5, P8, F4, AF4), and gamma (AF3) bands; DASM for theta (AF3/AF4), alpha (P7/P8), and beta (F7/F8) bands; and RASM for beta (AF3/AF4, P7/P8), theta (AF3/AF4), slow alpha (O1/O2) and alpha (F7/F8) bands.
- We obtained 8 features for disgust: PSD in theta (AF3, F7, P7, AF4), slow alpha (AF3, FC5, P7, F4), alpha (F7, P7) and gamma (AF3, F3, AF4) band; PSA index in the beta band (P7/P8); and DASM and RASM for slow alpha and beta bands (F3/F4).
- We obtained only one feature for: sadness (PSD in theta band channel O2), fear (PSD in alpha band channel P8), happiness (PSD in gamma band channel F8), neutral (PSD in beta band channel P8), anger (PSD in alpha band channel FC6), and surprise (PSD in alpha band channel P7).

Finally, we built a dataset combining the results from the univariate selection and the RFE feature selection process to determine how the performance of the classifiers behave in contrast to the original sets of traits.

In general, PSD, PSA DE and DASM features were selected for arousal labels. Diverse EEG information were retrieved for valence labels. PSD, PSA and DE at the temporal (T7, T8), and occipital (O1, O2) regions of the scalp were selected for sadness. For happiness PSD and DE features were selected. PSD, PSA, DASM and RASM were the features selected for surprise label. For disgust, PSD, PSA, DE DASM and RASM at the frontal (AF3, AF4, F3, F4, FC5, FC6, F7, F8) and parietal (P7, P8) regions of the scalp were selected.

### 3.3. Classifiers

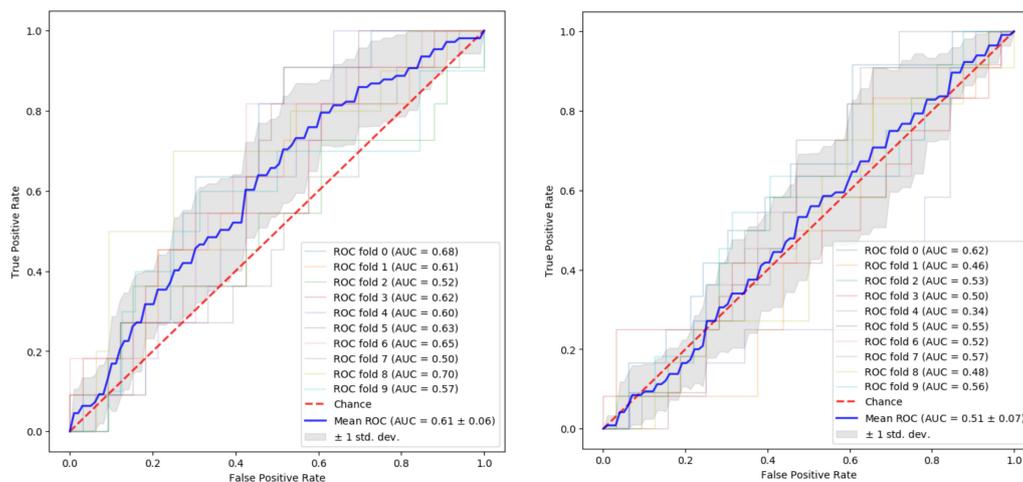
#### 3.3.1. EEG Data, Sex, Age, and Personality Traits to Predict Video Emotional Labels

We tested the different machine learning classification models with a 10-fold cross-validation for the two features sets defined as follow: a first set of features with EEG data, sex, age, and personality traits; and a second set of features with EEG data, sex, age, and personality traits reduced using feature selection. In Table 3, the mean accuracy, mean F1, and mean AUC scores are shown for the different sets, classifiers, and scenario labels. For valence–arousal scenario, the first set of features outperformed when we used SVM with linear kernel for the HAHV (accuracy 0.61, F1 0.14, AUC 0.61), LAHV (accuracy 0.64, F1 0.15, AUC 0.54), and LALV (accuracy 0.61, F1 0.15, AUC 0.55) labels. When we used the second set of features, HALV (accuracy 0.74, F1 0.00, AUC 0.51) using SVM with linear kernel had a good performance—higher than change for accuracy and AUC scores. For arousal and valence labels, using the second set of features for arousal, the best classifier was SVM with linear kernel (accuracy 0.52, F1 0.49, AUC 0.51); for valence the best classifier was ANN (accuracy 0.51, F1 0.67, AUC 0.56). For arousal labels, the worst performance was obtained with random forest and the first set of features. For valence labels, the worst performance was obtained with Naïve Bayes and the first set of features.

**Table 3.** Classifiers performance for each of the scenarios with the different set of traits: accuracies, F1, and AUC scores.

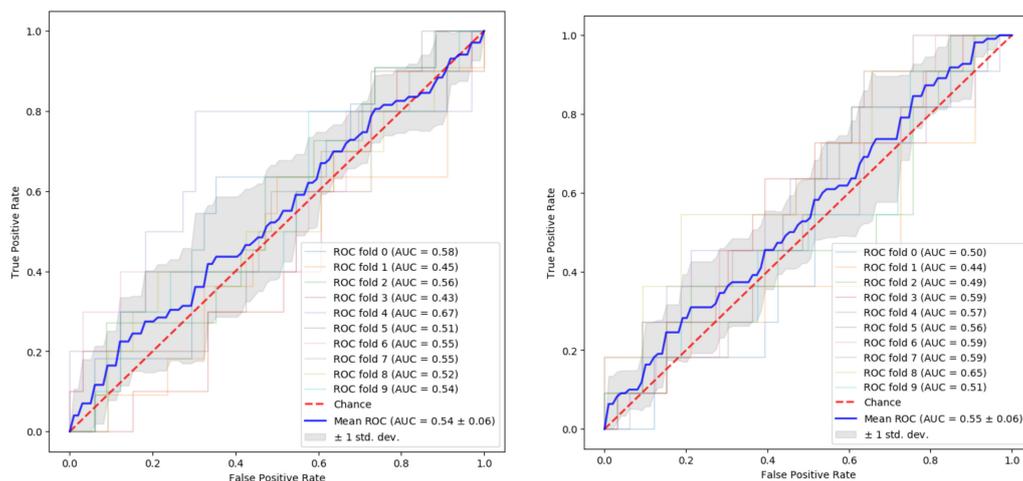
Scenario	Classifiers	Label	EEG Data, Demographic Characteristics, and Personality Traits (First Set of Features)			EEG Data, Demographic Characteristics, and Personality Traits (Reduction) (Second Set of Features)		
			Mean Accuracy	Mean F1	Mean AUC	Mean Accuracy	Mean F1	Mean AUC
Valence–arousal	SVM linear	HAHV	0.61	0.14	0.61	0.75	0.00	0.50
		HALV	0.61	0.23	0.49	0.74	0.00	0.51
		LAHV	0.64	0.15	0.54	0.76	0.00	0.48
		LALV	0.61	0.15	0.55	0.75	0.00	0.49
	SVM RBF	HAHV	0.60	0.20	0.49	0.75	0.00	0.56
		HALV	0.63	0.28	0.46	0.74	0.00	0.51
		LAHV	0.66	0.24	0.48	0.76	0.00	0.48
		LALV	0.60	0.19	0.49	0.75	0.00	0.48
Arousal	SVM linear		0.48	0.46	0.53	0.52	0.49	0.51
	SVM RBF		0.50	0.49	0.51	0.50	0.45	0.43
	Naïve Bayes		0.44	0.51	0.45	0.51	0.24	0.51
	Random Forest		0.44	0.39	0.39	0.44	0.40	0.43
	ANN		0.51	0.13	0.47	0.52	0.21	0.54
Valence	SVM linear		0.49	0.50	0.55	0.51	0.55	0.45
	SVM RBF		0.47	0.49	0.52	0.52	0.59	0.44
	Naïve Bayes		0.42	0.31	0.43	0.54	0.43	0.55
	Random Forest		0.44	0.48	0.39	0.49	0.51	0.50
	ANN		0.50	0.34	0.47	0.51	0.67	0.56

We used receiver operating characteristic (ROC) curves to describe the performance of the best classifiers obtained from each scenario. In Figure 5, we show the 10-fold cross-validation ROC curves for each of the valence-arousal labels when using the first and second features sets with the best accuracies scores. For HAHV, LAHV and LALV labels, the first set of features containing EEG data, sex, age, and personality traits without feature selection obtained the best accuracy score. For the HALV label, the second set of features containing EEG data, sex, age, and personality traits with feature selection had the best accuracy classification. The curves show that the classification process is higher than chance; however, the F1 scores were low, indicating that the classifiers did not achieve good precision (number of correct positive predictions the model got among all the items identified as positive) nor recall (proportion of predictions correctly identified as positive); i.e., the predictions are not relevant in these cases.



(a) High arousal and high valence (HAHV) Support Vector Machine (SVM) linear kernel (first set of features)

(b) High arousal and low valence (HALV) Support Vector Machine (SVM) linear kernel (second set of features)

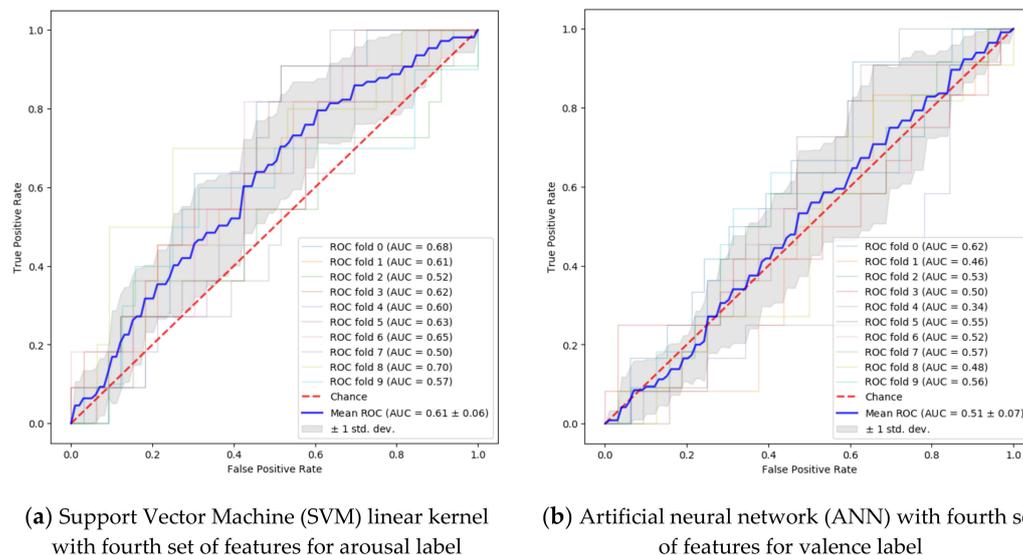


(c) Low arousal and high valence (LAHV) Support Vector Machine (SVM) linear kernel (first set of features)

(d) Low arousal and low valence (LALV) Support Vector Machine (SVM) linear kernel (first set of features)

**Figure 5.** Best receiver operating characteristic curves with 10-fold cross-validation for valence-arousal label.

Figure 6 shows the 10-fold cross-validation ROC curves for the arousal label scenario and the valence label scenario with the best accuracies scores. The curves show the best classification performance for arousal label was obtained using EEG traits with feature reduction and SVM with linear kernel classifier (0.52 accuracy score when AUC score is higher than chance). For the valence scenario, the second set of features and the ANN classifier had the best accuracy; in this case, the curve shows that the classification process was slightly higher than chance.



**Figure 6.** Best receiver operating characteristic curves with 10-fold cross-validation.

### 3.3.2. EEG Data, Sex, Age, and Personality Traits to Predict Self-Assessed Traits Labels

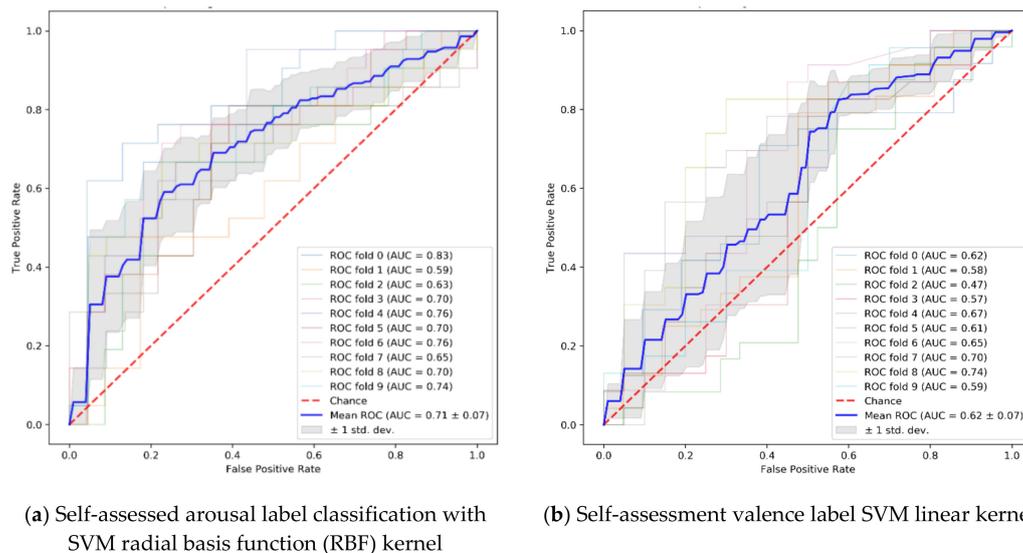
In Table 4, the mean accuracy, mean F1 and mean AUC scores are shown for the different classifiers and scenario labels. For the arousal scenario, the first set of features (EEG data, sex, age, and personality traits without reduction) performed better when we used SVM with RBF kernel (accuracy 0.68, F1 0.67, AUC 0.71). For the valence scenario, the second set of features (EEG data, sex, age, and personality traits with reduction) performed better when we used SVM with linear kernel (accuracy 0.61, F1 0.65, AUC 0.62). In these cases, we noticed that no demographic characteristics nor personality traits were selected in the reduced set of features; i.e., the improvement in the classification accuracies was owing to the EEG traits selected. For arousal labels, the worst performance was obtained with ANN in both sets of features. For valence labels, the worst performance was obtained with Naïve Bayes and ANN in both sets of features.

When we compared the classifier performance for the discrete emotions, we identified that the accuracy and the AUC scores yielded good results for some of the cases; however, the F1 scores were low, indicating that the classifiers did not achieve good precision (number of correct positive predictions the model got among all the items identified as positive) nor recall (proportion of predictions correctly identified as positive), indicating that the predictions were not relevant in these cases.

**Table 4.** Classifiers performance for each of the scenarios with the defined set of traits accuracies, F1, and area under the curve (AUC) scores.

Scenario	Classifiers	AMIGOS	EEG Data, Demographic Characteristics, and Personality Traits (Third Set of Features)			EEG Data, Demographic Characteristics, and Personality Traits (Reduction) (Fourth Set of Features)		
		F1	Mean Accuracy	Mean F1	Mean AUC	Mean Accuracy	Mean F1	Mean AUC
Arousal	SVM linear	0.592	0.63	0.60	0.66	0.62	0.58	0.65
	SVM RBF		0.68	0.67	0.71	0.64	0.63	0.67
	Naïve Bayes		0.54	0.60	0.57	0.59	0.60	0.62
	Random Forest		0.64	0.61	0.69	0.63	0.61	0.69
	ANN		0.52	0.20	0.54	0.52	0.04	0.62
Valence	SVM linear	0.576	0.53	0.56	0.47	0.61	0.65	0.62
	SVM RBF		0.52	0.56	0.46	0.59	0.64	0.62
	Naïve Bayes		0.50	0.59	0.47	0.52	0.67	0.49
	Random Forest		0.52	0.60	0.50	0.53	0.60	0.55
	ANN		0.51	0.49	0.47	0.53	0.63	0.53
Sadness	SVM linear		0.59	0.29	0.47	0.71	0.00	0.52
	SVM RBF		0.62	0.35	0.57	0.70	0.18	0.60
	Naïve Bayes		0.52	0.32	0.49	0.67	0.30	0.62
	Random Forest		0.67	0.09	0.61	0.67	0.16	0.57
	ANN		0.71	0.00	0.53	0.71	0.00	0.55
Fear	SVM linear		0.64	0.16	0.53	0.79	0.00	0.46
	SVM RBF		0.71	0.20	0.48	0.79	0.00	0.49
	Naïve Bayes		0.30	0.32	0.43	0.79	0.00	0.53
	Random Forest		0.78	0.00	0.47	0.66	0.16	0.50
	ANN		0.79	0.00	0.47	0.79	0.00	0.48
Happiness	SVM linear		0.76	0.11	0.50	0.88	0.00	0.49
	SVM RBF		0.80	0.08	0.50	0.88	0.00	0.56
	Naïve Bayes		0.42	0.19	0.45	0.85	0.12	0.59
	Random Forest		0.87	0.00	0.50	0.87	0.03	0.56
	ANN		0.88	0.00	0.46	0.88	0.00	0.41
Neutral	SVM linear		0.59	0.34	0.55	0.70	0.19	0.61
	SVM RBF		0.6	0.44	0.63	0.69	0.37	0.62
	Naïve Bayes		0.50	0.49	0.60	0.68	0.15	0.61
	Random Forest		0.70	0.23	0.60	0.67	0.25	0.61
	ANN		0.70	0.00	0.49	0.70	0.00	0.50
Disgust	SVM linear		0.80	0.20	0.52	0.89	0.00	0.52
	SVM RBF		0.85	0.25	0.64	0.88	0.10	0.56
	Naïve Bayes		0.33	0.21	0.53	0.43	0.20	0.56
	Random Forest		0.88	0.00	0.63	0.88	0.03	0.63
	ANN		0.89	0.00	0.59	0.89	0.00	0.49
Anger	SVM linear		0.54	0.33	0.50	0.61	0.01	0.49
	SVM RBF		0.53	0.36	0.43	0.63	0.09	0.55
	Naïve Bayes		0.39	0.52	0.52	0.53	0.40	0.54
	Random Forest		0.58	0.18	0.50	0.60	0.27	0.49
	ANN		0.60	0.05	0.47	0.63	0.00	0.48
Surprise	SVM linear		0.76	0.18	0.41	0.86	0.00	0.57
	SVM RBF		0.78	0.18	0.39	0.85	0.00	0.56
	Naïve Bayes		0.32	0.27	0.54	0.46	0.24	0.56
	Random Forest		0.85	0.03	0.51	0.84	0.02	0.54
	ANN		0.86	0.00	0.53	0.86	0.00	0.61

We use ROC curves to describe the performance of the best classifiers obtained. Figure 7 shows the 10-fold cross-validation ROC curves for arousal and valence labels. The best accuracy scores were obtained for the arousal with 0.68 and valence with 0.61. For the discrete emotions, we decided not to show the ROC curves owing to the low F1 scores obtained in each case.



**Figure 7.** Best receiver operating characteristic curves with 10-fold cross-validation with for Valence label.

#### 4. Discussion

The results obtained in this work revealed that none age, sex, or personality had a correlation with arousal and valence labels from the emotional stimuli. However, compared with self-assessed emotional labels, some demographic characteristics and personality traits were chosen by the feature selection for arousal; for some of the discrete emotions, this might be because the self-assessed responses relied on participants' subjective emotion assessment. If so, demographic characteristics and personality traits would correlate more with the self-assessed emotion responses than with the emotional labels from the stimuli videos. If we analyze the classification performance, only relevant results were obtained for arousal and valence labels from self-assessed answers (owing to low values of F1 scores for discrete emotions). Feature selection showed only an improvement in the classification scores for the valence label; neither demographic characteristics nor personality traits were selected by the feature selection process, which shows that age, sex, and personality traits did not foster classification performance improvement for the selected labels.

It is known from previous works that sex and age can be correlated with these emotional labels and can improve emotion recognition process [2]; however, it is still unclear how personality can be used to obtain better emotion recognition models. We believe that one of the reasons why sex, age, and personality were not chosen by the feature selection algorithms was because the nature of the data. If we adjust the values to a categorical and binary codification, the feature selection algorithms could select these kind of features (as age was selected in the Rukavina and colleagues' work [2]). We decided to work with the continuous data owing to the real description of the population. Other possible limitation is related to the distribution of personalities in the participants, because the sample is relative small to obtain a vast distribution in the five personality traits assessed, and the reported scores are close to each other, implying that the participants exhibit the same type of personality among the group [51], it is difficult to obtain data that describe all the possible outcomes to design classifiers.

Works like [69], intended to create more complex deep learning models which personality information can increase in 10% accuracy of the emotion recognition. Although the literature shows a strong relationship between emotion, mood affective states, and personality, the papers presented here and the information from the AMIGOS dataset analysis still does not provide conclusive information about whether there is a strong correlation between emotion stimuli, emotional states, and personality. We believe this is owing to how the information from the personality questionnaires is being fit as a feature for the classifiers and how the classifiers are being designed for emotion recognition. Utilizing

new, deep learning techniques could possibly integrate this kind of information in a more suitable way to achieve personalized emotion recognition models. There is also a need for a behavioral metric that can identify differences between how people perceive and manifest emotions. Behavior changes and emotional reaction can vary from person to person owing to past experiences, memories, and context.

Comparing the results, it is still difficult, using traditional machine learning models or basic deep learning models, to obtain higher classification accuracies using EEG traits when different variables need to be considered (the number of participants in an experiment, number of EEG channels, EEG signals and traits, etc.). Furthermore, it is important to consider the dynamics of the emotional stimuli and how the participants perceive these stimuli; pictures, videos, interactive interfaces, and virtual environments come with different variables. There is still the need to analyze how time, familiarity, interaction, and so on affect individuals' emotion recognition processes, and how the EEG features are correlated with these variables to describe individuals' emotional behavior when interacting with stimuli material. Classification accuracies vary between the different EEG traits used in the classification process and the number of participants in the experiment [11,14]. How to obtain good classification accuracies in cross-participants experiments, which can allow researchers to have more freedom in using different stimuli methods and degrees of interaction with systems to identify emotional states from EEG signals, remains unknown. The literature provides hints about how behavioral cues can be described as digital data to use in emotion recognition when there is an interaction between a person and a machine.

One of the most important physiological signals used in emotion recognition is EEG, owing to the number of features that can identify emotional behavior in the brain and the idea of integrating emotion to BCI systems; however, it is still difficult to achieve higher accuracies for emotion classification using only EEG signals. To face this challenge, multimodal approaches are being implemented because they are robust and increase the accuracy for emotion classification, in contrast to systems that only rely on one information source. Signals like ECG, EMG, and GSR are also considered in these studies because they provide relevant information about individuals' emotional and behavioral state.

Perceived emotions may be owing to exposure to the emotional stimuli (video in this case); however, the chosen dataset did not have information about arousal–valence scores related to the video time traces. In the scope of this analysis, we did not try to trace the changes in emotional response related to the emotional stimuli over time; instead, we wanted to determine how EEG data, age, sex, and personality traits performed while classifying emotions compared with the AMIGOS dataset, in which the classification was made by averaging the time window. Consequently, we used different machine learning techniques and compared the results with the AMIGOS dataset. To analyze emotion recognition over time, techniques such as RNN or learning and teaching support material are recommended, which are beyond the scope of this study.

In future research, it is important to address specific challenges like: the access to a wider and diverse population where participant exhibit different demographic characteristics, personality traits, and behavioral cues; the nature of the emotional stimuli, whether they are passive or active; the data gathered and its evaluation during stimuli exposure time, and the interaction type that the participant can experience while using HCI systems. For personalized HCI, it is important to analyze, not only intrinsic characteristics as demographic or personality traits, but also behavioral cues that manifest when using HCI systems and its context. For future works we would like to focus our approach on capturing and analyzing behavioral cues, together with physiological signals, related to the use of a specific technology or the task being performed.

**Author Contributions:** All authors contributed in the conceptualization methodology of the manuscript. L.A.M.-T. and Y.M. performed the data preparation and formal analysis and writing, N.Y. and Y.K. reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by Tateishi Science and Technology Foundation (grant number 2188001), JST PRESTO (Precursory Research for Embryonic Science and Technology) (grant number JPMJPR17JA), JST MIRAI grant number JY300171).

**Acknowledgments:** We thank Queen Mary University of London and the University of Trento for the permission to use the AMIGOS dataset, which is a good source of information to develop new approaches in the emotion recognition context.

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

## References

1. Jeon, M. Chapter 1 - Emotions and Affect in Human Factors and Human-Computer Interaction: Taxonomy, Theories, Approaches, and Methods. In *Emotions and Affect in Human Factors and Human-Computer Interaction*; Elsevier: Amsterdam, The Netherlands, 2017. [[CrossRef](#)]
2. Rukavina, S.; Gruss, S.; Hoffmann, H.; Tan, J.-W.; Walter, S.; Traue, H.C. Affective computing and the impact of gender and age. *PLoS ONE* **2016**, *11*, e0150584. [[CrossRef](#)]
3. Okon-Singer, H.; Hendler, T.; Pessoa, L.; Shackman, A.J. The neurobiology of emotion-cognition interactions: Fundamental questions and strategies for future research. *Front. Hum. Neurosci.* **2015**, *9*, 1–14.
4. Laborde, S. Bridging the Gap between Emotion and Cognition: An Overview. *Perform. Psychol. Percept. Act. Cognit. Emot.* **2016**, 275–289. [[CrossRef](#)]
5. Lench, H.C.; Flores, S.A.; Bench, S.W. Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: A meta-analysis of experimental emotion elicitation. *Psychol. Bull.* **2011**, *137*, 834–855. [[CrossRef](#)] [[PubMed](#)]
6. Vinciarelli, A.; Mohammadi, G. A Survey of Personality Computing. *IEEE Trans. Affect. Comput.* **2014**, *5*, 273–291. [[CrossRef](#)]
7. Pocius, K.E. Personality factors in human-computer interaction: A review of the literature. *Comput. Human Behav.* **1991**, *7*, 103–135. [[CrossRef](#)]
8. Kim, K.H.; Bang, S.W.; Kim, S.R. Emotion recognition system using short-term monitoring of physiological signals. *Med. Biol. Eng. Comput.* **2004**, *42*, 419–427. [[CrossRef](#)]
9. Russell, J. A circumplex model of affect. *J. Pers. Soc. Psychol.* **1980**, *39*, 1161–1178. [[CrossRef](#)]
10. Tao, J.; Tan, T. Affective computing: A review. *Affect. Comput. Intell. Interact.* **2005**, *3784*, 981–995.
11. Al-Nafjan, A.; Hosny, M.; Al-Ohali, Y.; Al-Wabil, A. Review and Classification of Emotion Recognition Based on EEG Brain-Computer Interface System Research: A Systematic Review. *Appl. Sci.* **2017**, *7*, 1239. [[CrossRef](#)]
12. Coogan, C.G.; He, B. Brain-Computer Interface Control in a Virtual Reality Environment and Applications for the Internet of Things. *IEEE Access* **2018**, *6*, 10840–10849. [[CrossRef](#)] [[PubMed](#)]
13. Kim, M.-K.; Kim, M.; Oh, E.; Kim, S.-P. A Review on the Computational Methods for Emotional State Estimation from the Human EEG. *Comput. Math. Methods Med.* **2013**, *2013*, 1–13.
14. Alarcao, S.M.; Fonseca, M.J. Emotions Recognition Using EEG Signals: A Survey. *IEEE Trans. Affect. Comput.* **2017**, *3045*, 1–20. [[CrossRef](#)]
15. Liu, Y.; Sourina, O.; Nguyen, M.K. Real-time EEG-based human emotion recognition and visualization. In Proceedings of the 2010 International Conference on Cyberworlds, Singapore, 20–22 October 2010; pp. 262–269.
16. Jatupaiboon, N.; Pan-Ngum, S.; Israsena, P. Emotion classification using minimal EEG channels and frequency bands. In Proceedings of the 2013 10th International Joint Conference on Computer Science and Software Engineering, Maha Sarakham, Thailand, 29–31 May 2013; pp. 21–24.
17. Balconi, M.; Mazza, G. Brain oscillations and BIS/BAS (behavioral inhibition/activation system) effects on processing masked emotional cues. ERS/ERD and coherence measures of alpha band. *Int. J. Psychophysiol.* **2009**, *74*, 158–165. [[CrossRef](#)]
18. Vecchiato, G.; Toppi, J.; Astolfi, L.; Fallani, F.D.V.; Cincotti, F.; Mattia, D.; Bez, F.; Babiloni, F. Spectral EEG frontal asymmetries correlate with the experienced pleasantness of TV commercial advertisements. *Med. Biol. Eng. Comput.* **2011**, *49*, 579–583. [[CrossRef](#)]
19. Davidson, R.J. Anterior cerebral asymmetry and the nature of emotion. *Brain Cogn.* **1992**, *20*, 125–151. [[CrossRef](#)]
20. Li, M.; Lu, B.L. Emotion classification based on gamma-band EEG. In Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, Minneapolis, MN, USA, 3–6 September 2009; pp. 1323–1326.

21. Park, K.S.; Choi, H.; Lee, K.J.; Lee, J.Y.; An, K.O.; Kim, E.J. Emotion recognition based on the asymmetric left and right activation. *Int. J. Med. Med. Sci.* **2011**, *3*, 201–209.
22. Kandel, E.R.; Schwartz, J.H.; Jessell, T.M. *Principles of Neural Science*, 5th ed.; McGraw-Hill: New York, NY, USA, 2013.
23. American Psychological Association. “Personality,” APA. 2015. Available online: <https://www.apa.org/topics/personality/> (accessed on 13 April 2020).
24. Eysenck, H.J.; Eysenck, S.B.G. *Manual of the Eysenck Personality Questionnaire: (EPQ-R Adult)*; EdITS/Educational and Industrial Testing Service: San Diego, CA, USA, 1994.
25. McCrae, R.R.; Costa, P.T., Jr. A Five-Factor theory of personality. In *Handbook of Personality: Theory and Research*, 2nd ed.; Guilford Press: New York, NY, USA, 1999; pp. 139–153.
26. Gray, J.A. A Critique of Eysenck’s Theory of Personality. In *A Model for Personality*; Springer: Berlin/Heidelberg, Germany, 1981; pp. 246–276.
27. Ashton, M.C.; Lee, K.; Perugini, M.; Szarota, P.; de Vries, R.E.; Di Blas, L.; Boies, K.; De Raad, B. A Six-Factor Structure of Personality-Descriptive Adjectives: Solutions from Psycholexical Studies in Seven Languages. *J. Pers. Soc. Psychol.* **2004**, *86*, 356–366. [[CrossRef](#)]
28. Li, H.; Pang, N.; Guo, S.; Wang, H. Research on textual emotion recognition incorporating personality factor. In Proceedings of the 2007 IEEE International Conference on Robotics and Biomimetics (ROBIO), Sanya, China, 15–18 December 2007; pp. 2222–2227.
29. Omheni, N.; Kalboussi, A.; Mazhoud, O.; Kacem, A.H. Annotation-Based Learner’S Personality Modeling in Distance Learning Context. *Turkish Online J. Distance Educ.* **2016**, *17*, 46–62. [[CrossRef](#)]
30. Wei, W.L.; Wu, C.H.; Lin, J.C.; Li, H. Interaction style detection based on Fused Cross-Correlation Model in spoken conversation. In Proceedings of the ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 8495–8499.
31. Fallahnezhad, M.; Vali, M.; Khalili, M. Automatic Personality Recognition from reading text speech. In Proceedings of the 2017 Iranian Conference on Electrical Engineering (ICEE), Tehran, Iran, 2–4 May 2017; pp. 18–23.
32. Batrinca, L.; Mana, N.; Lepri, B.; Sebe, N.; Pianesi, F. Multimodal Personality Recognition in Collaborative Goal-Oriented Tasks. *IEEE Trans. Multimed.* **2016**, *18*, 659–673. [[CrossRef](#)]
33. Alam, F.; Riccardi, G. Predicting personality traits using multimodal information. In Proceedings of the 2014 Workshop on Computational Personality Recognition, Workshop of MM 2014, WCPR 2014, Orlando, FL, USA, 7 November 2014; pp. 15–18.
34. Batrinca, L.; Lepri, B.; Pianesi, F. Multimodal recognition of personality during short self-presentations. In Proceedings of the 2011 ACM Multimedia Conference and Co-Located Workshops - JHGBU 2011 Workshop, J-HGBU’11, MM’11, Scottsdale, AZ, USA, 1 December 2011; pp. 27–28.
35. Guo, A.; Ma, J. Archetype-based modeling of persona for comprehensive personality computing from personal big data. *Sensors* **2018**, *18*, 684. [[CrossRef](#)] [[PubMed](#)]
36. Celli, F.; Lepri, B. Is Big Five better than MBTI ? A personality computing challenge using Twitter data. In Proceedings of the CEUR Workshop, Torino, Italy, 10–12 December 2018; Volume 2253.
37. Callejas-Cuervo, M.; Martínez-Tejada, L.A.; Botero-Fagua, J.A. Architecture of an emotion recognition and video games system to identify personality traits. In Proceedings of the VII Latin American Congress on Biomedical Engineering CLAIB 2016, Bucaramanga, Santander, Colombia, 26–28 October 2017; pp. 42–45.
38. Hu, K.; Guo, S.; Pang, N.; Wang, H. An intensity-based personalized affective model. In Proceedings of the 2007 IEEE International Conference on Robotics and Biomimetics, ROBIO, Sanya, China, 15–18 December 2007; pp. 2212–2215.
39. Wei, X.S.; Zhang, C.L.; Zhang, H.; Wu, J. Deep Bimodal Regression of Apparent Personality Traits from Short Video Sequences. *IEEE Trans. Affect. Comput.* **2018**, *9*, 303–315. [[CrossRef](#)]
40. Nasoz, F.; Lisetti, C.L.; Vasilakos, A.V. Affectively intelligent and adaptive car interfaces. *Inf. Sci. NY* **2010**, *180*, 3817–3836. [[CrossRef](#)]
41. Anzalone, S.M.; Varni, G.; Ivaldi, S.; Chetouani, M. Automated Prediction of Extraversion During Human–Humanoid Interaction. *Int. J. Soc. Robot.* **2017**, *9*, 385–399. [[CrossRef](#)]
42. Bhin, H.; Lim, Y.; Park, S.; Choi, J. Automated psychophysical personality data acquisition system for human-robot interaction. In Proceedings of the 2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence, URAI 2017, Jeju, Korea, 28 June–1 July 2017; pp. 159–160.

43. Cai, R.; Guo, A.; Ma, J.; Huang, R.; Yu, R.; Yang, C. Correlation Analyses Between Personality Traits and Personal Behaviors Under Specific Emotion States Using Physiological Data from Wearable Devices. In Proceedings of the 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech), Athens, Greece, 12–15 August 2018; pp. 46–53.
44. Miranda-Correa, J.A.; Patras, I. A Multi-Task Cascaded Network for Prediction of Affect, Personality, Mood and Social Context Using EEG Signals. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018; pp. 373–380.
45. Mittermeier, V.; Leicht, G.; Karch, S.; Hegerl, U.; Möller, H.-J.; Pogarell, O.; Mulert, C. Attention to emotion: Auditory-evoked potentials in an emotional choice reaction task and personality traits as assessed by the NEO FFI. *Eur. Arch. Psychiatry Clin. Neurosci.* **2011**, *261*, 111–120. [[CrossRef](#)]
46. Subramanian, R.; Wache, J.; Abadi, M.; Vieriu, R.; Winkler, S.; Sebe, N. Ascertain: Emotion and personality recognition using commercial sensors. *IEEE Trans. Affect. Comput.* **2018**, *9*, 147–160. [[CrossRef](#)]
47. Mueller, C.J.; Kuchinke, L. Individual differences in emotion word processing: A diffusion model analysis. *Cogn. Affect. Behav. Neurosci.* **2016**, *16*, 489–501. [[CrossRef](#)]
48. Carver, C.S.; Sutton, S.K.; Scheier, M.F. Action, Emotion, and Personality: Emerging Conceptual Integration. *Personal. Soc. Psychol. Bull.* **2000**, *26*, 741–751. [[CrossRef](#)]
49. Allers, R. *Emotion and Personality*; Columbia University Press: New York, NY, USA, 1961; Volume 35.
50. John, O.P.; Gross, J.J. Healthy and Unhealthy Emotion Regulation: Personality Processes, Individual Differences, and Life Span Development. *J. Pers.* **2004**, *72*, 1301–1334. [[CrossRef](#)]
51. Miranda Correa, J.A.; Abadi, M.K.; Sebe, N.; Patras, I. AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups. *IEEE Trans. Affective Comput.* **2018**, *1*. [[CrossRef](#)]
52. Zheng, W.L.; Lu, B.L. Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. *IEEE Trans. Auton. Ment. Dev.* **2015**, *7*, 162–175. [[CrossRef](#)]
53. Wu, D.; Courtney, C.G.; Lance, B.J.; Narayanan, S.S.; Dawson, M.E.; Oie, K.S.; Parsons, T.D. Optimal arousal identification and classification for affective computing using physiological signals: Virtual reality stroop task. *IEEE Trans. Affect. Comput.* **2010**, *1*, 109–118. [[CrossRef](#)]
54. Koelstra, S.; Muhl, C.; Soleymani, M.; Lee, J.-S.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I. DEAP: A Database for Emotion Analysis using Physiological Signals. *IEEE Trans. Affective Comput.* **2012**, *3*, 18–31. [[CrossRef](#)]
55. Zheng, W.-L.; Liu, W.; Lu, Y.; Lu, B.-L.; Cichocki, A. EmotionMeter: A Multimodal Framework for Recognizing Human Emotions. *IEEE Trans. Cybern.* **2019**, *49*, 1110–1122. [[CrossRef](#)] [[PubMed](#)]
56. Raad, B.D.; Perugini, M. *Big Five Assessment*; Hogrefe & Huber Publishers: Ashland, OH, USA, 2002.
57. Li, X.; Song, D.; Zhang, P.; Zhang, Y.; Hou, Y.; Hu, B. Exploring EEG Features in Cross-Subject Emotion Recognition. *Front. Neurosci.* **2018**, *12*, 162. [[CrossRef](#)] [[PubMed](#)]
58. Jenke, R.; Peer, A.; Buss, M. Feature Extraction and Selection for Emotion Recognition from EEG. *IEEE Trans. Affect. Comput.* **2014**, *5*, 327–339. [[CrossRef](#)]
59. Becker, H.; Fleureau, J.; Guillotel, P.; Wendling, F.; Merlet, I.; Albera, L. Emotion recognition based on high-resolution EEG recordings and reconstructed brain sources. *IEEE Trans. Affect. Comput.* **2017**. [[CrossRef](#)]
60. Sourina, O.; Liu, Y. A Fractal-based Algorithm of Emotion Recognition from EEG using Arousal-Valence Model. In Proceedings of the BIOSIGNALS International Conference on Bio-Inspired Systems and Signal, Rome, Italy, 26–29 January 2011; pp. 209–214.
61. Chen, D.-W.; Miao, R.; Yang, W.-Q.; Liang, Y.; Chen, H.-H.; Huang, L.; Deng, C.-J.; Han, N. A Feature Extraction Method Based on Differential Entropy and Linear Discriminant Analysis for Emotion Recognition. *Sensors* **2019**, *19*, 1631. [[CrossRef](#)]
62. Duan, R.-N.; Zhu, J.-Y.; Lu, B.-L. Differential entropy feature for EEG-based emotion classification. In Proceedings of the 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER), San Diego, CA, USA, 6–8 November 2013; pp. 81–84.
63. Bradley, M.M.; Lang, P.J. Measuring emotion: The self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* **1994**, *25*, 49–59. [[CrossRef](#)]

64. Guyon, I. *Feature Extraction Foundations and Applications*; Springer: Berlin/Heidelberg, Germany, 2006.
65. Saeys, Y.; Inza, I.; Larranaga, P. A review of feature selection techniques in bioinformatics. *Bioinformatics* **2007**, *23*, 2507–2517. [[CrossRef](#)]
66. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
67. Huan, L.; Hiroshi, M. *Computational Methods of Feature Selection*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2007.
68. Boschetti, A.; Massaron, L. *Python Data Science Essentials*, 2nd ed.; Packt Publishing: Birmingham, UK, 2016.
69. Zhao, S.; Ding, G.; Han, J.; Gao, Y. Personality-aware personalized emotion recognition from physiological signals. In Proceedings of the IJCAI International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 1660–1667.



© 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/>).