Stress assessment with EEG and machine learning in affective VR environments

Neurocomputing 638 (2025) 130185

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Stress assessment with EEG and machine learning in affective VR environments

Federica Marcolin [a,*], Elena Carlotta Olivetti [a], Ivonne Angelica Castiblanco Jimenez [b], Giorgia Passavanti [a], Sandro Moos [a], Enrico Vezzetti [a], Alessia Celeghin [b]

[a] Department of Management and Production Engineering, Politecnico di Torino, Torino, Italy
[b] Department of Psychology, University of Torino, Torino, Italy

ARTICLE INFO

Keywords:
Stress
EEG
Virtual reality
Emotion assessment
Affective computing
Machine learning

ABSTRACT

Stress is a reaction that occurs when a person perceives, with or without awareness, an imbalance between requests and available resources. Relying on this definition, we have carried out an experiment in a Virtual Reality environment to elicit (light) stress in the user and analyze the emotional responses with electroencephalography (EEG). The virtual environment is divided in eight parts; in each of them a stressor has been put in action, meaning that in every part the participants perform a task, but a specific resource is missing (time, knowledge, control, salvation, no or too many alternatives, engagement, self-confidence). EEG is used to assess the emotional response with the aid of Valence/Arousal/Dominance/Stress indicators presented in previous literature. Nine indicators calculated for 87 participants, labeled according to self-assessment replies (post-experimental questionnaires), were classified with eXtreme Gradient Boosting, k-Nearest Neighbor, Support Vector Machine and Random Forest classifiers. The lowest results in terms of accuracy were obtained with k-Nearest Neighbor (around 70 %), whilst the highest ones were obtained with eXtreme Gradient Boosting and Random Forest (above 98 %), showing that EEG could be a valuable tool to assess the emotional response in stressful situations, with a particular focus on the Stress indicators.

1. Introduction

Stress is defined as a general adaptation syndrome aimed at establishing a new internal balance following stress factors, called stressors [1]. A person may become stressed if s/he feels threatened, insecure, or challenged by specific events or social interactions. The way stress is experienced is unique, although some factors can be identified as more prone to elicit stress clinical manifestations. The perception of uncertainty and the urge to act in conditions that show missing alternatives, time pressure, limited knowledge, lack of control or (metaphoric) salvation, lead to an overwhelming mental and physical load and, consequently, to a stress response. Among the sources responsible for generating stress and anxiety in individuals, we can list a few which appear to be more significant in a varied audience: competition [2], public speaking [3,4], social relationships [2,5–7], missing entertainment [2], examinations [6], exposure to life threatening events [8,9], lack of knowledge and competence [7], lack of control [10,11], lack of alternatives [12,13] or, contrarily, choice overload [14–16], and time pressure [17–19].

Although stress is frequently described as a challenge to homeostasis, the environmental features that contribute to the production of stress responses remain underexplored due to limitation in recreating integrated ecological, contextual and multiple scenarios as those triggering stress in everyday life [20]. The interrelation of emotions and Virtual Reality (VR) has been recognized thanks to its enhanced ecological validity [21]. VR is a simulated experience in which a person can look around the virtual world, move around in it, and engage with virtual objects or features, through visuals, sounds and other sensations. It is acknowledged that there is a tendency for VR to increase emotional responses [22], as participants experience slightly stronger valence and arousal in VR [23], to increase sense of presence (“the more I feel present the more I am moved” and vice versa) [24] or immersion, and the illusion of body ownership and agency [25]. A few research experiments and related studies investigated VR as emotion eliciting material [26]: basic emotions [27–31], frustration, confusion, boredom, engagement and anxiety [32–36], relaxation, depression, distress [37,38] have been

* Corresponding author.
E-mail address: federica.marcolin@polito.it (F. Marcolin).

https://doi.org/10.1016/j.neucom.2025.130185
Received 16 January 2025; Received in revised form 15 March 2025; Accepted 29 March 2025
Available online 10 April 2025

chosen as targets in some VR-based experimentations. The methodology often conceptualizes affective VR environments, relying on valence, arousal and dominance dimensions and have been tested using an in-world interface for the self-assessment, made through a valence-arousal Self-Assessment Manikin (SAM) slider [39]. With these advantages, VR was recently used as a tool to elicit and treat stress. Among the common paradigms used to induce stress in laboratory settings, the Trier Social Stress Test (TSST) represents one of the widely used, as it emphasizes the environmental features known to be effective stressors. This paradigm, indeed, requires the participants to hold a speech and to do an arithmetic task in front of an audience [40–43]. The study of stress and stressors takes on added importance when placed in the context of everyday life, in which individuals are exposed to stressful stimuli of different kinds, such as domestic and work-related stress [44–47]. The study of the latter is assuming considerable relevance because it is an integral part of the transformation process from the smart, connected and self-controlling production model to the sustainable, resilient and human-centered model [48]. In light of this, the use of VR as eliciting means assume a twofold potentiality; in fact, along with the benefits that immersivity can induce in the effectiveness of the stressors at the level of in-laboratory study, the possibility to create digital twins of real work environment integrated with tools to monitor induced stress levels can support the process of redesigning and reengineering of the workplace [49]. Moreover, the effectiveness of VR in stress-related research is shown by VR-exposure therapy (VRET). VRET has been often used to treat post-traumatic stress disorder, with a reported success rate of 66–90 %, as VR allows to transport the patient into a setting similar to an experience where an anxiety-causing effect occurred, with the purpose to desensitize her/him [50].

As a (more typically) negative valence and high arousal emotion, stress has some specific biochemical and physiological responses in the body, such as an increase of the secretion of cortisol, blood pressure and heart rate. Other neural properties, such as pattern electroencephalographic waves in response to stress, can be traced but less easily framed. The frontal asymmetry, specifically of Alpha waves (and known as frontal Alpha asymmetry - FAA), seems to be the best candidate for a neural correlate of stress and is one of the parameters/features that can be extracted from EEG bands [51]. More generally, the sensitivity of high Alpha to contingent stress has been more widely studied [52,53]. A variation of Beta activity was also observed [52,54,55]. The role of Alpha and Beta during stress has also been acknowledged thanks to development of three Stress Indicators: β/α [56], reprised in the present study, α/β and θ/β [57,58], though, still little literature explored their effectiveness. Brouwer et al. [44] show a link between frontal Alpha asymmetry, heart variability and cortisol on the one hand and stress on the other hand. Tarrant et al. [55] focused on Alpha and Beta sub--bandwidths' activity suggesting an increased Alpha associated with calm and relaxed states, and Beta associated with qualitatively anxious states. Though, further research on EEG and stress is needed to contrast registered methodological and finding discrepancies [59,60].

Recent research has been focused on the use of machine learning (ML) to automatically classify two or three stress levels (corresponding to the classes of the classifier) thanks to EEG-based features, including Alpha frontal asymmetry. Concerning the specific application of ML in the affective analysis, it can be noticed that, both for time and frequency domain, widely adopted algorithms are Support Vector Machine (SVM), k-Nearest Neighbor (kNN), Naive Bayes, and ensemble learning algorithms (e.g., Random Forest - RF, eXtreme Gradient Boosting - XGBoost) [61,62].

In particular, recent literature has applied ML to EEG stress related waves patterns supporting its effectiveness although with a certain variability among the classifier selected. Al-Shargie [63] found a significant reduction in Alpha rhythm power from one stress level to another level, and in 2015 detected mental stress with EEG with an accuracy of 94 %, 85 %, and 80 % (using SVM) at level one, two and three of arithmetic problem difficulty, respectively. Jun and Smitha [64] obtained an accuracy of 75 % with a SVM algorithm used to classify the relative difference of Beta and Alpha power, used as EEG-based features, into three levels of stress. Smitha et al. [65] used SVM to effectively differentiate good and bad stress from EEG band power features. Arsalan et al. [66] obtained a high accuracy (92.85 %) for the binary classification with Theta band features. Frontal asymmetry was used as a real-time EEG feature for stress assessment by Arpaia et al. [67]. Four standard ML classifiers reached more than 90 % accuracy in classifying each 2-s epoch of EEG acquired from 10 stressed subjects. Wen et al. [58] obtained with SVM a validation accuracy of 98 %; labeling was not performed using self-assessment questionnaires as ground truth, but k-means clustering was adopted.

The valence-arousal-dominance dimensions were considered in ML classifications in several studies. The adopted approaches and main findings of each study have been summed up in Table 1.

With the aim of contributing to the studies on stress-related conditions with a human-computer interaction and a human-centered perspective, this work focuses on analyzing EEG-related emotional indicators to assess the stressed state in participants navigating a 'stressful' Virtual Reality (VR) environment, created with eight different tasks and stressors. Relying on previous literature, we have selected eight micro-stressors and creted a twin 'stressless' VR environment. Two experiments were conducted: one where participants were asked to navigate the virtual environments using a monitor (in the following, called non-Head-Mounted Display experiment) and one with a Head-Mounted Display (HMD). Participants were asked to fill-in a questionnaire after the navigation of the virtual environments to self-assess their levels of arousal, valence, dominance and stress. During the experiment, participants were asked to wear an EEG headset. EEG-based indicators such as Valence, Arousal, Dominance and Stress indicators were computed. Their values, labeled according to the relative self-assessments, were classified with four classifiers (XGBoost, kNN, SVM, RF) to investigate their validity as emotional assessment indicators for stressful states.

This work integrates several aspects of innovation, with potential implications for both research and practical applications. First, the diversity of stressors. The study employs eight distinct micro-stressors presented in VR, providing a comprehensive evaluation of stress induction mechanisms in virtual environments. Second, a within-subject paired comparison was adopted, leveraging a stressful environment and its stressless counterpart. This approach enhances the reliability of stress induction validation and allows to assess the effectiveness of the selected stressors through VR, enabling future modulation of stressors' intensity for the development of novel VR-based stressors. Third, the use of affective indicators as ML features. Rather than relying on raw EEG data or generic statistical features (e.g., signal dispersion, entropy-based measures, or power spectral density), this study employs well-established affective indicators (stress, valence, arousal, dominance) as ML features. This ensures interpretability of the extracted features in relation to emotional states. Fourth, a five-class ML classification was performed, with a direct correlation with SAM rates, while previous approaches only adopted two-three classes (low level, neutral level, high level) or combination of low/high levels of the affective dimensions. This makes the study more nuanced and specific in terms of the emotional assessment and more relevant to the original SAM formulation, which involved five classes. Lastly, an evaluation of multiple affective indicator formulations is performed. Different computational formulations of affective indicators from the literature were tested to determine their effectiveness in discriminating stress-related emotional states, providing insights into the most reliable formulations for EEG-based emotional assessment. To the best of our knowledge, literature lacks studies integrating these aspects.

2. Materials and methods

100 gender-balanced participants aged between 18 and 71 years were involved in two experimental sessions. The first session, later on

Table 1
Main literature findings related to ML classification of valence-arousal-dominance dimensions extracted from EEG data.

authors, year and reference	ML approach	dataset	dimensions	classification	accuracy
Li et al. 2018 [68]	Hierarchical CNN kNN SVM	DEAP [69]	Valence / Arousal	Binary (low/high)	86–88 % 79 % 74 %
Verma et al. 2017 [70]	kNN SVM MLP (multi layer perceptron)	DEAP	Valence / Arousal / Dominance	Three classes (low/medium/ high)	65 % 70 % 68 %
Gaertner et al. 2021 [71]	RF	DEAP	Valence / Arousal / Dominance	Eight classes	99 %
Kumar et al. 2021 [72]	kNN SVM	DEAP	Valence / Arousal / Dominance	Two classes	85–92 %
Liu et al. 2023 [73]	XGBoost	DEAP	Valence / Arousal	Binary (low/high) Multiclass	80–95 % 91–95 %
Zong et al., 2023 [74]	XGBoost SVM	DEAP / DREAMER [74]	Valence / Arousal	Binary (low/high)	94 % 92 %
Abdel-Hamid 2023 [75]	kNN SVM	DEAP	Valence	Binary/Multiclass	96.33 % 97.42 %
Pei et al. 2024 [76]	kNN RF	DEAP	Valence	Three classes	72 % 95 %
Miah et al. 2022 [77]	MLR (multinomial logistic regression) SVM	GAMEEMO [78]	Valence / Arousal	Four classes	99.80 % 99 %
Yu et al. 2022 [79]	SVM	VREED [79]	Valence	Binary	70 %

called ***non-HMD experimental session***, involving 39 participants (19 females, aged between 20 and 71), concerns the use of a desk monitor for navigating the VR environments; the second session, called ***HMD experimental session***, involving 61 participants who did not take part in the first session (37 females, aged between 18 and 60), involves the adoption of the Meta Quest 2 as HMD for a more immersive virtual experience in the same environments. The number of participants was sufficient to meet the required sample size for paired *t*-test given a power of 0.8, an alpha level of 0.05 and an effect size of 0.5. These values are in line with previous literature on EEG based analyses and common practice [80,81]. All experimentations have been conducted according to the Declaration of Helsinki. Written informed consent was obtained from all participants.

Before starting a session, participants were asked to read an informative document and to sign a consent form. A 15-seconds EEG baseline was acquired both with open fixed eyes and then with closed eyes. Then, participants were asked to navigate the two VR environments. After each environment, a questionnaire was administered. The experimental workflow is shown in Fig. 1.

In the non-HMD experimental session, participants are then asked to sit on a chair without wheels in front of a desk equipped with a 27-inch monitor computer, a keyboard (WASD keys to move), a mouse (to control the viewpoint) and a RGB-D camera. In the HMD experimental session, participants wear the Meta Quest 2 and train in a neutral VR environment to learn how to move and to interact with virtual objects with the controllers.

2.1. Virtual Reality environments

Two VR environments have been used with the aim of studying the stressful condition from an EEG perspective. The developed VR environments, briefly explained in the following section, have been in-depth presented and validated both with direct (questionnaires) and indirect (electrodermal activity) measures in another work (Nonis et al., under review).

The first VR environment (called in the following 'stressful environment') was conceived with the aim to elicit stress in the participants; in the second (called 'stressless environment'), the stressors were alleviated to provide a less stressful (or stressless) experience. In both cases, the environment was an urban office building in which the participant had to attend a job interview. The majority of the tasks had to be accomplished in the elevator of this building. Unity (version 2021.3.18f1) was used for creating the environments and Blender19 (version 3.4) was used to design the objects.

Eight stressors were incorporated into the VR environment thanks to specific tasks and situations, one at a time, in which the stressor could manifest. Every task lasts 30 seconds and represents the lack of a condition or of an element; the stressor will be mitigated in the stressless environment (e.g., the 'limited time' stressor in the stressful environment is suppressed in the stressless environment by giving an unlimited time for the same task). The selection and the number of the stressors relied on i) the literature most effective choices of stressors, ii) the practical possibility to implement a related task in VR, iii) the possibility to alleviate the stressor in the stressless environment and, iv) the time

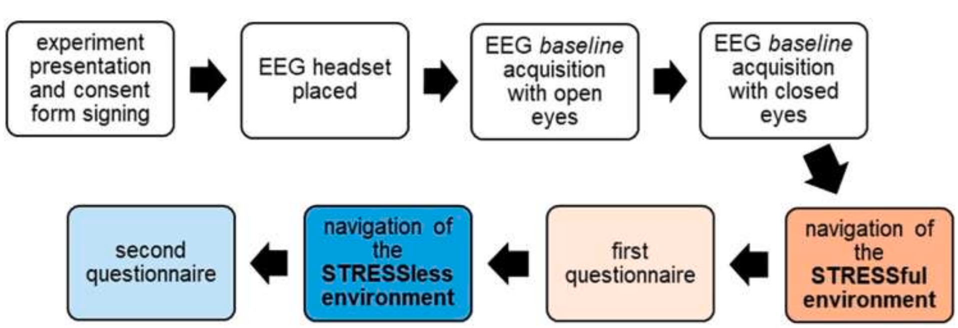

Fig. 1. Experimental workflow.

availability for the whole experimentation.

The eight missing elements and related tasks in the stressful environment are:

1. **lack of time**: solve a mathematical equation or copy numerical code in a limited time frame;
2. **lack of salvation**: fire propagation;
3. **lack of engagement**: blocked elevator;
4. **lack of control**: alarm activation with buttons working wrong;
5. **lack of knowledge**: mount a breadboard;
6. **lack of self-confidence**: job interview;
7. **lack of alternatives**: return to floor 0 and restart;
8. **overchoice**: office floor or contract choice.

A selection of screenshots from the environments is shown in Fig. 2.

2.2. Emotional self-assessment

After every VR experience, a questionnaire is administered to the participant with an online form. The Self-Assessment Manikin (SAM) is used to ask participants their levels of arousal, valence and dominance for every task. Their level of felt stress in a Likert scale 1–5 is also asked for every task. The last two questions in the questionnaire concern 1) the other emotion(s) felt during the experiment (a list was provided plus an open field for other replies) and 2) the resource that the participant felt s/he lacked the most during the task (proper knowledge, enough time, control of the situation, involvement, ability to choose, having an alternative, ease/confidence/energy/courage, safety, other…). The questionnaire was used here for labeling data for the ML classification.

2.2.1. EEG

Two devices were used to collect EEG data. The use of two different devices allowed to assess whether variations in the spatial resolution of the recorded signals could affect the calculation of affective indicators and, consequently, their effectiveness in capturing emotional changes. For the non-HMD experimental session, a 14-channel wireless Emotiv Epoc X headset was used. The electrodes are AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4; the reference electrodes Common Mode Sense (CMS) and Driven Right Leg (DRL) were placed at P3 and P4 (Fig. 3). An even number indicates the electrodes placed in the right brain hemisphere, whilst an odd number indicates a placement in the left hemisphere. For the HMD experimental session, a configurable 32-channel wireless Emotiv Flex 2 saline head cap was used. The electrodes are AF3, AF4, F7, F3, Fz, F4, F8, FT9, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, TP9, P7, P3, Pz, P4, P8, O1, Oz and O2; CMS/DRL were placed at TP9 and TP10 (Fig. 4). The international 10–20 system was adopted to place the electrodes in both cases. For both experimental sessions, the sampling rate was 128 Hz. A band-pass filter was applied, and the Fast Fourier Transform (FFT) was computed to obtain the band powers ($\mu V^2/Hz$); to reduce noise, a high-pass filter was applied before FFT. For data analysis and feature extraction, a Hanning window size of 2 seconds epoch was used consisting of 256 EEG samples; this window was slided by 16 samples to create the new window. Then, the squared magnitude of the complex FFT value was averaged in each frequency band (θ: 4–8 Hz, α: 8–12 Hz, β: 12–25 Hz, γ: 25–45 Hz), and only artifact-free signals were considered. Thus, the EEG dataset of 100 (39 + 61) participants was reduced to 87 (33 for the non-HMD session and 54 for the HMD session).

Nine affective indicators, namely three for Valence, two for Arousal, one for Dominance and three for Stress, detailed in Table 2, were calculated at every instant from EEG waves, relying on formulas

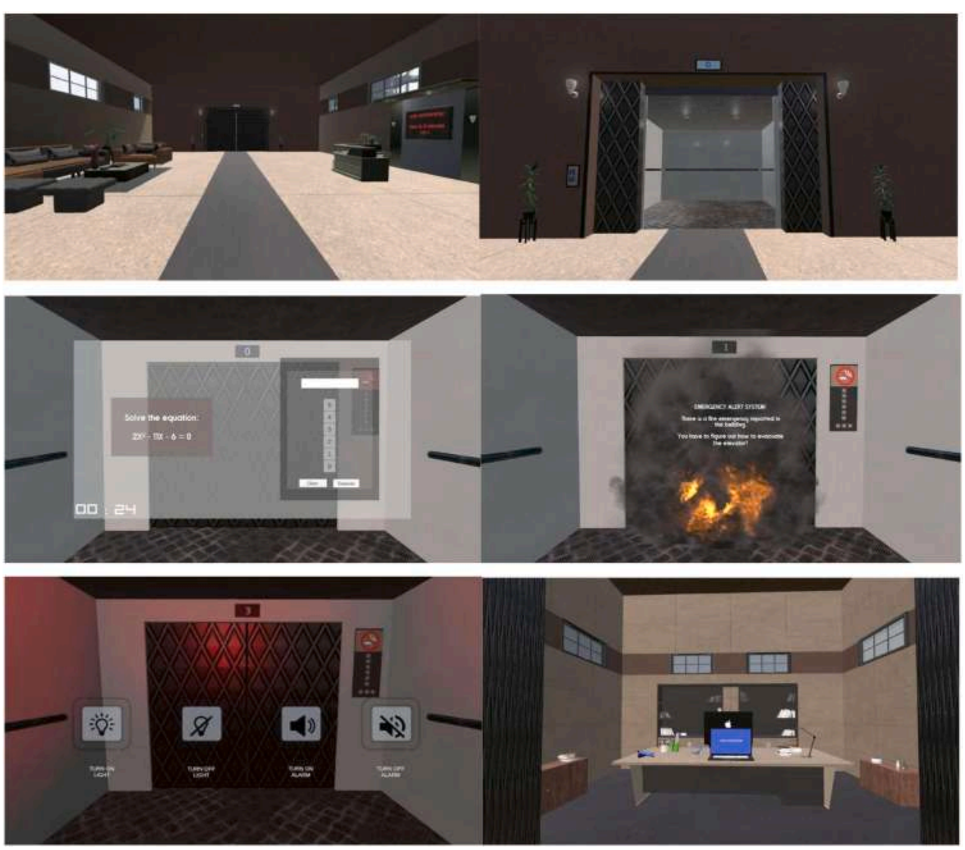

Fig. 2. Selected screenshots of the VR environments. From top to bottom, from left to right: the hall of the building, elevator viewed from outside, the equation task (no time), the fire (no salvation), the alarm activation with wrong working buttons (no control), the interview office (no self-confidence). Colors and general background design choices were made according to previous literature on affective VR [27].

Fig. 3. 14-channel EEG headset (above, left) and its electrode position (above, right); 32-channel EEG head cap (below, left) and its electrode position (below, right).

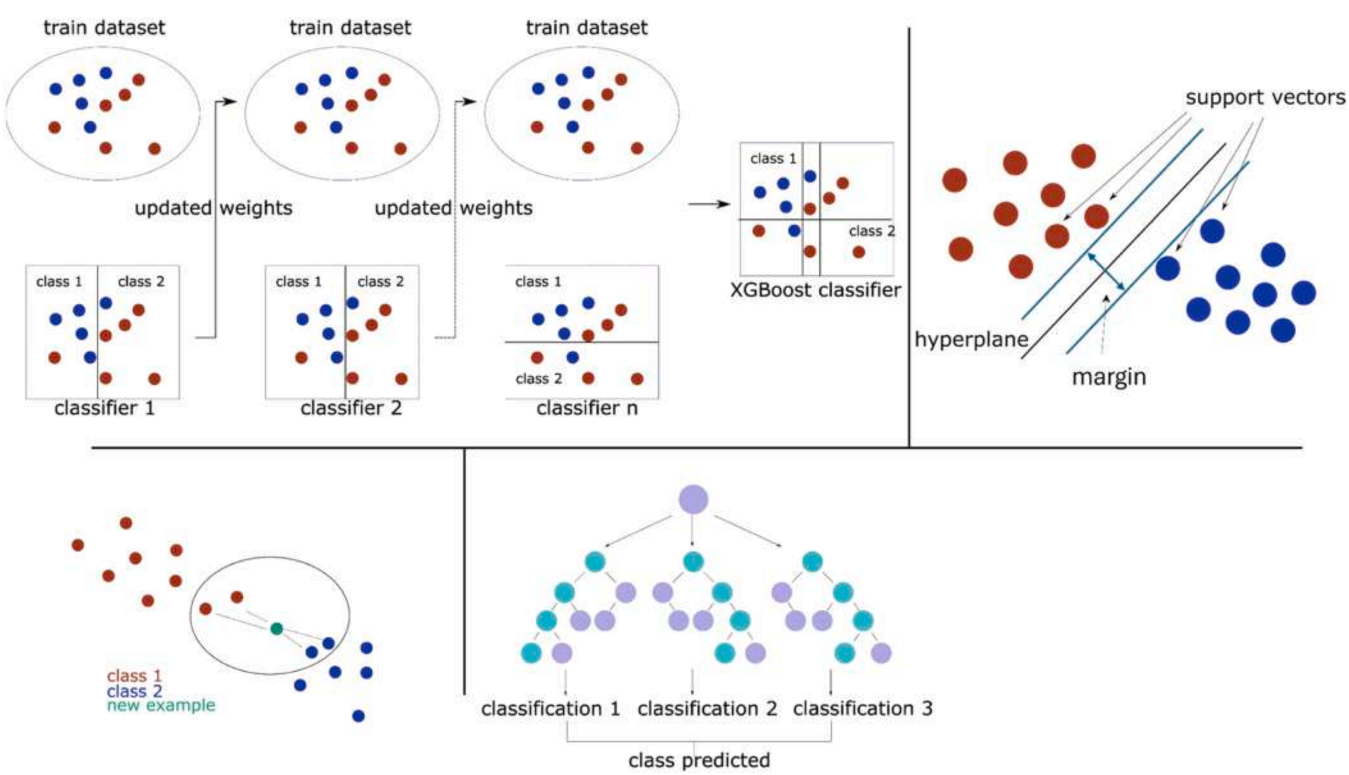

Fig. 4. Theoretical overview of classification algorithms: XGBoost (above, left), SVM (above, right), kNN (below, left), RF (below, right). For the sake of simplicity, a binary classification is shown as an example.

Table 2
Affective indicators with related names, formulas, (essential) references and abbreviations adopted in this study. The FAA has been used as an indicator both for valence and stress for its theoretical characterization related to the negativity-positivity of the emotion and its potential role of neural correlate of stress reported in previous literature.

affective dimension / emotion	affective indicator	formula	references	we will refer to it as
VALENCE	Frontal Alpha Asymmetry (FAA)	$\ln \frac{\alpha_{F4}}{\alpha_{F3}}$	Ramirez et al., 2018 [82] Aldayel et al., 2020 [83] Castiblanco et al., 2023; 2023 [84,85]	V1
	Alpha/Beta valence indicator	$\frac{\alpha_{F4}}{\beta_{F4}} - \frac{\alpha_{F3}}{\beta_{F3}}$	Ramirez and Vamvakousis, 2012 [86] Giraldo and Ramirez, 2013 [87] Al-Nafjan et al., 2017 [88] Hwang et al., 2018 [89]	V2
	Beta/Alpha valence indicator	$\frac{\beta_{AF3,F3}}{\alpha_{AF3,F3}} - \frac{\beta_{AF4,F4}}{\alpha_{AF4,F4}}$	Aldayel et al., 2020 [83]	V3
AROUSAL	F3-F4 arousal indicator	$\frac{\beta_{F3} + \beta_{F4}}{\alpha_{F3} + \alpha_{F4}}$	Giraldo and Ramirez, 2013 [87] Castiblanco et al., 2023; 2023 [84,85]	A1
	AF3-AF4 arousal indicator	$\frac{\beta_{F3} + \beta_{F4} + \beta_{AF3} + \beta_{AF4}}{\alpha_{F3} + \alpha_{F4} + \alpha_{AF3} + \alpha_{AF4}}$	Al-Nafjan et al., 2017 [88]	A2
DOMINANCE	dominance indicator	$\frac{\beta_{FC6} + \beta_{F8}}{\alpha_{FC6} + \alpha_{F8}}$	Liu and Sourina, 2012 [90]	D
STRESS	Beta/Alpha stress indicator	$\frac{\beta}{\alpha}$	Yusuf et al., 2019 [56]	S1
	Alpha/Beta stress indicator	$\frac{\alpha}{\beta}$	Wen et al., 2020 [58] Altaf et al., 2021 [57]	S2
	Theta/Beta stress indicator	$\frac{\theta}{\beta}$		S3
	Frontal Alpha Asymmetry (FAA)	$\ln \frac{\alpha_{F4}}{\alpha_{F3}}$	Ramirez et al., 2018 [82] Aldayel et al., 2020 [83] Castiblanco et al., 2023; 2023 [84,85] Seo et al., 2010 [51]	S4

available in previous literature studies. Statistical paired *t*-test (level of significance $\alpha = 0.05$) was employed to ascertain a significant difference in the results of the EEG indicators used to evaluate the stressful and stressless VR environments.

2.3. ML-based analysis

With the aim of supporting the thesis that EEG-based affective indicators (Valence, Arousal, Dominance and Stress) could be valuable tools to assess the emotional response in stressful situations, four different ML classifiers were adopted, relying on previous literature.

ML is well-suited for classifying EEG signals [58]. This is due to several factors: 1) it can effectively handle the non-linearity of brain signals, 2) it can effectively capture the intricate patterns that constitute the brain signals, 3) it can handle the high-dimensional nature of brain signals better than traditional statistics, and, moreover, 4) ML supervised algorithms can be trained on different datasets that are representative of the variation of the brain signals over the time, allowing to adapt the pattern identification to individual variabilities. In the proposed methodological approach, additional factors influenced the choice of classical ML algorithms over other methods such as deep learning models, even if literature is growing in this direction. First, the dataset size was not sufficient to train neural networks effectively avoiding the risk of overfitting. Second, our approach involved a rigorous preprocessing of raw EEG data to maintain a strict control over quality of data before feeding them into the ML pipeline. Third, feature extraction, specifically the computation of valence, arousal, dominance and stress indicators following different literature-based formulations, was performed prior to classification; therefore, the autonomous feature learning typical of neural networks was not required.

A brief theoretical overview of the adopted classifiers is given below (Fig. 4).

2.3.1. Adopted classifiers

XGBoost is a ML algorithm based on the gradient boosting framework. It sequentially builds a series of weak learners, here decision trees, to correct the errors of the previous one, gradually improving the overall model; the final model is a robust learner [91]. It can efficiently handle missing data by automatically learning how to deal with missing values during the training process, reducing the need for pre-processing or imputation. Additionally, XGBoost can provide insights into feature importance, aiding users in understanding which features contribute the most to the model's predictions. This is valuable for feature selection and comprehends the underlying patterns in the data; thus, it seems suitable to handle EEG based features due to their complexity. Unlike other decision tree-based classifiers such as RF, XGBoost constructs new models to rectify the classification errors of the previous ones instead of creating separate decision trees.

SVM is a supervised algorithm for classification and regression in different applications [92,93]. It was originally designed to handle binary problems, but implementations for multiclass classification have been proposed. A hyperplane is used to separate the classes in the 2D/3D feature space, and is defined according to the support vectors, i.e., points in the data that are closest to the decision boundary. The distance from the support vectors of each class and the hyperplane is called margin (Fig. 5). SVM is particularly effective in high-dimensional spaces with linear and non-linear problems. On the other hand, its non-probabilistic nature, together with the strong dependence of performance on the appropriate choice of parameters, are disadvantages in its adoption.

KNN is a non-parametric model that is suitable for binary and multiclass classification problems, as well as regression tasks [94]. It is particularly useful when decision boundaries are complex and cannot be effectively represented by a parametric model; in fact, its sensitivity to local patterns makes it effective when the decision boundaries are not linear and when the decision is determined by the local distribution of instances. KNN is commonly referred to as a 'lazy learner' because it does not learn a model during the training phase. Instead, it memorizes the entire training dataset and uses it for predictions during the testing phase. The parameter 'k' in kNN represents the number of nearest neighbors to consider. The choice of this parameter impacts the performance of the algorithm, and it is often determined through cross-validation.

Random Forest is an ensemble method that builds a collection (i.e., a forest) of decision trees during training and combines their predictions to improve accuracy and generalization. The use of decision trees makes RF robust and capable of capturing complex relationships in the data. At each split in a decision tree, RF considers only a random subset of features instead of all features. This technique helps to decorrelate the trees

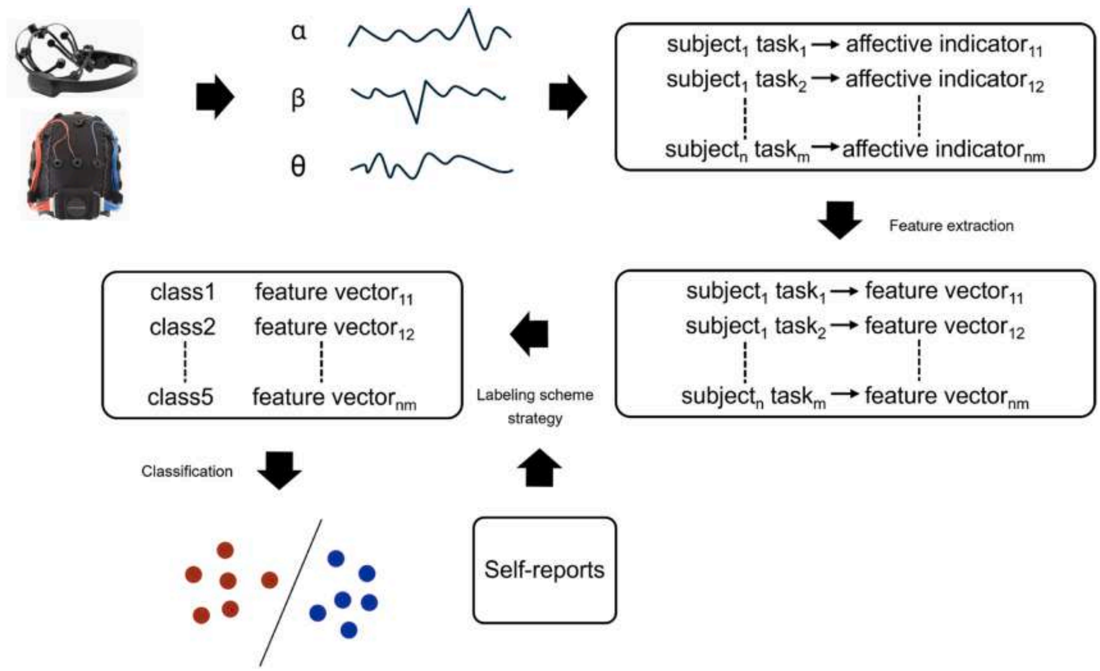

Fig. 5. Schematic representation of the construction of the feature vectors and data labeling. Raw data were preprocessed to obtain band powers (not reported in this scheme); then, affective indicators were computed for each participant in each task. Feature extraction was performed to reduce the feature vector before the classification phase, preserving the separation between subjects and tasks; data labeling was performed according to the SAM responses of each participant for each task (ground truth). Finally, classification was performed with multiple classifiers.

and increase diversity among them, enhancing the ability of this algorithm to generalize. As the XGBoost algorithm, RF can measure feature importance in making predictions. The importance of a feature is determined by its contribution to reducing impurity, such as Gini impurity, in decision trees [95].

2.3.2. Feature extraction

The FFT computed at a frequency resolution of 0.5 Hz with a sliding window of 16 samples (0.125 sliding window step) produced, over each 30-second collection, a total number of 225 windows. To reduce the dimension of the sample given in input to the classifiers and avoid redundancy, a feature vector was constructed to extrapolate significant information from each indicator computed over the complete experiences. To this aim, for each participant and for each task, mean, median and standard deviation were computed on all the values assumed by the indicator over the complete experience, and positive and negative peaks were identified. A peak was considered positive if its value was greater than the mean value plus two times the standard deviation; on the opposite, a peak was considered negative if its value was less than the mean value minus two times the standard deviation. Thus, the feature vector was constructed considering the median, the number of positive peaks, the number of negative peaks and the values of positive and negative peaks. Due to signal loss and variability between subjects, the number of positive and negative peaks could differ. As a result, missing values in the feature vectors were imputed by substituting them with the mean value calculated from the corresponding values in the feature matrix.

According to the adoption for the classification task of supervised algorithms, data labeling was performed based on the self-reports. A labeling scheme strategy was adopted so that each point on the 5-point Likert scale became a label for the feature vectors; in particular, the rate self-reported by a participant in a task was associated with the feature vector corresponding to that participant in that task. The same approach was adopted for all the computed affective indicators (for each participant and for each task). The FAA indicator has been labeled according to both valence [82–85] and stress [51] self-assessment rates and classified separately. The scheme in Fig. 5 summarizes the described process. The same scheme was followed for both non-HMD and HMD experimental sessions.

3. Results

The questionnaires' replies revealed that the participants have mostly experienced high levels of stress (statistical mode in the 1–5 Likert scale equal to 4 or 5, corresponding, respectively, to "very stressed" and "extremely stressed") during the stressful environment during tasks related to lack of time, salvation, control, knowledge and self-confidence. Lower levels of stress were elicited during the other tasks, showing that the lack of engagement and alternatives, and the too-many-alternatives condition were not effective stressors for most participants. Lower values have been reported for all the tasks in the stressless environments, with modes equal to 1 ("not stressed at all") or 2 ("not very stressed"). Therefore, the design of the virtual environments was 'affectively' effective in most cases.

The replies to the SAM questionnaires registered predominant medium-low valence, medium-high arousal and low dominance for the stressful environment; medium-high valence, medium arousal and medium-high dominance for the stressless environment. This shows that the desired SAM values have been achieved both when the stressor is put in action (stressful environment) and when it is lightened (stressless environment), especially for valence and dominance levels. The partially unchanged level of arousal between the two environments, especially in the non-HMD experimentation, may be ascribed to the general tendency for Virtual Reality experiences to increase arousal in participants [22]. The most chosen 'other' felt emotion beside stress was helplessness for the stressful environment in the non-HMD experimentation, anxiety and frustration in the HMD session, and satisfaction for the stressless environment in both experimental setups. A further in-depth analysis of questionnaires' responses is out of the scope of the present study, as self-reports have been mainly used to label data for classification. Though, for the sake of completeness, we have reported the questionnaire results in the Appendix of the present paper (Appendix

A).

The results of the paired t-tests indicate that the majority of the EEG indicators adopted show significant differences between stressful and stressless conditions suggesting that EEG signals can be effectively used to discriminate between the two states (and the two environments). Specifically, the indicators of arousal, valence (with the exception of V3), dominance and stress all show p-values inferior to 0.05. V3 in the HMD experimental session is the only indicator that is not statistically

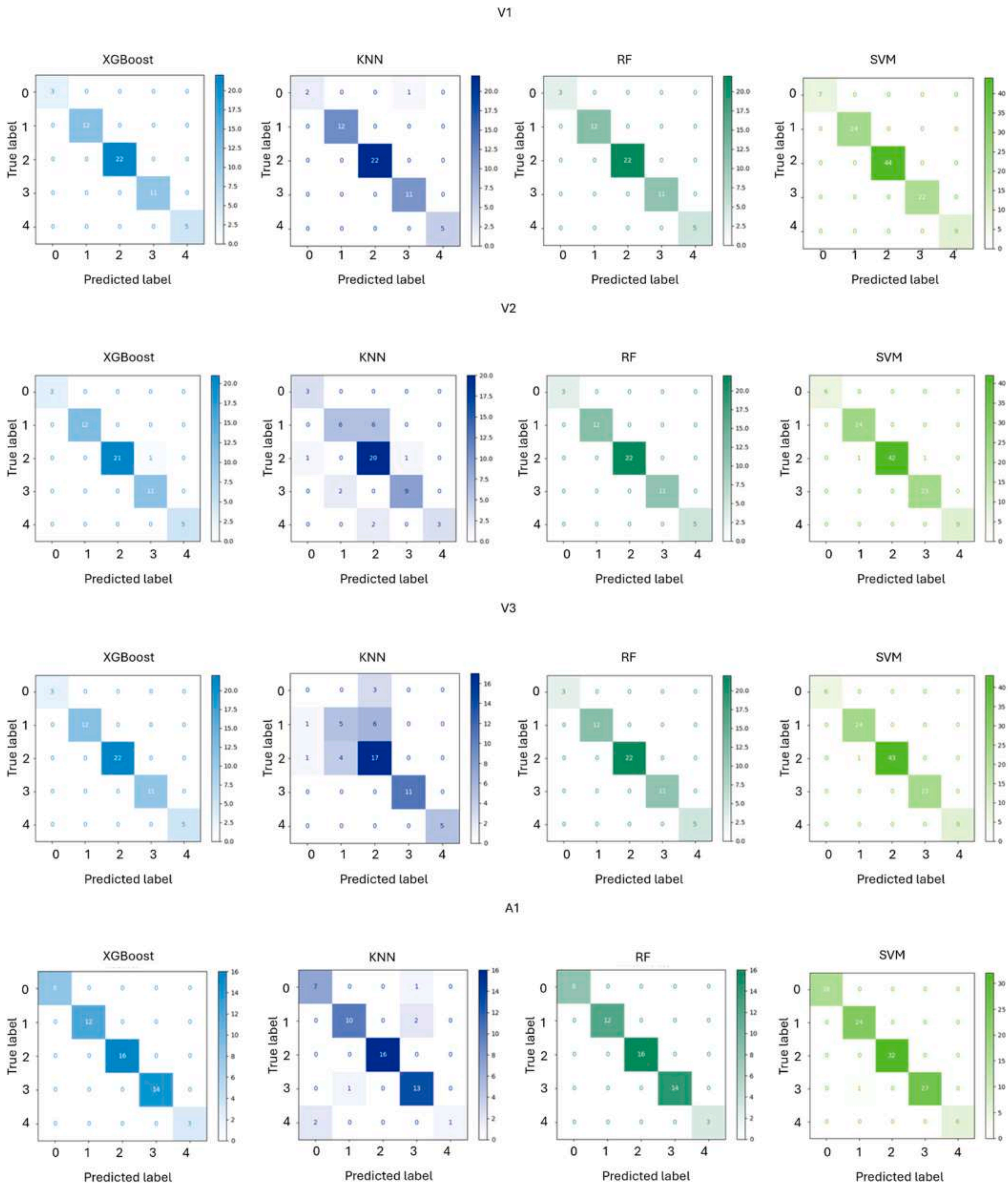

Fig. 6. Confusion matrices for the 5-class classification on the affective indicators from the *non-HMD experimental session*. Rows: adopted formulas for the affective indicators (V1, V2,…). Columns: adopted classifiers (eXtreme Gradient Boosting – XGBoost, k-Nearest Neighbor – kNN, Random Forest – RF, Support Vector Machine – SVM).

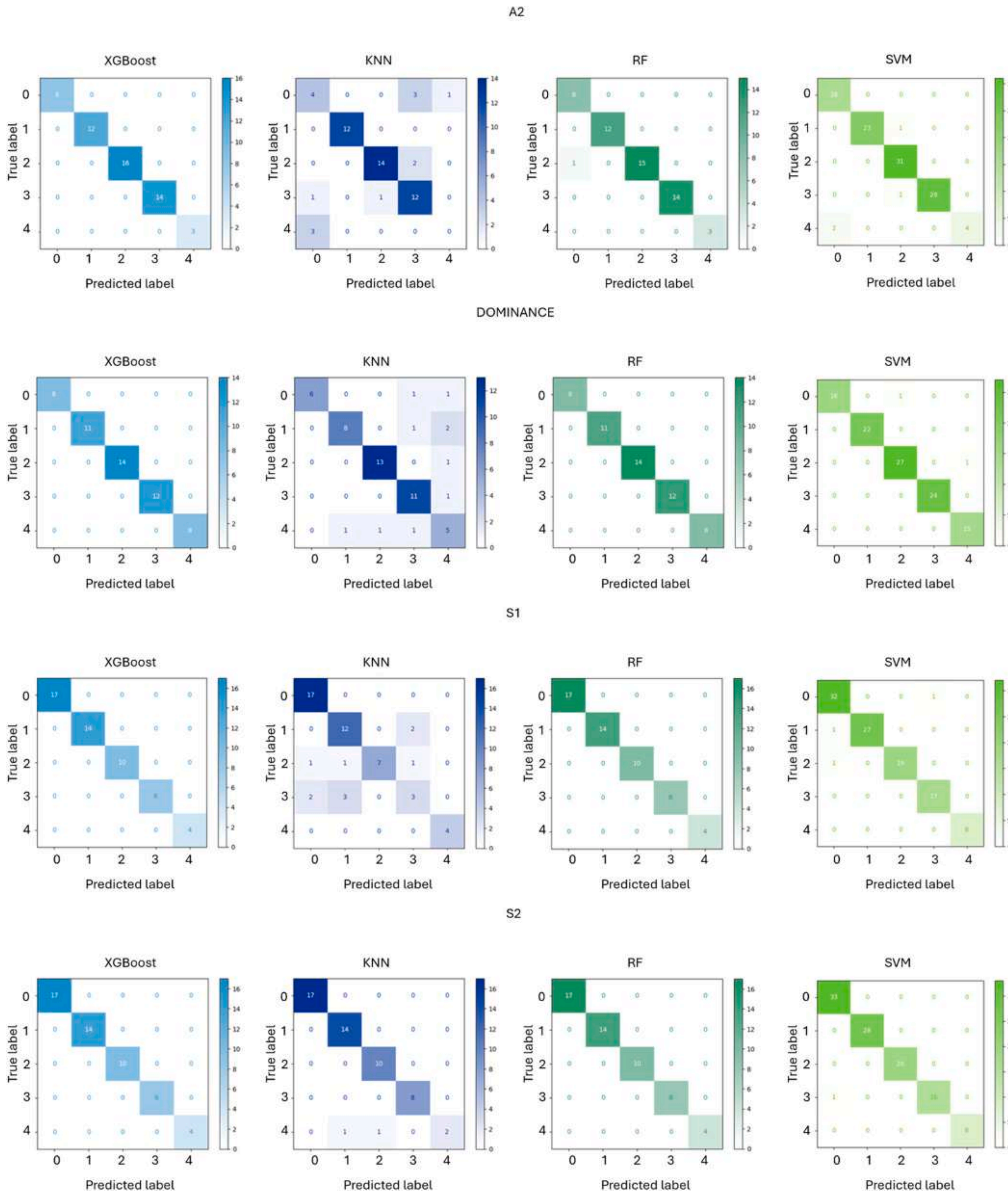

Fig. 6. (*continued*).

significant (p = .39).

The ML classification was conducted using the four classifiers on a test set, which comprised 30 % of the total dataset, while 70 % was allocated for training and validation. The classification was performed for all nine affective indicators, and confusion matrices for non-HMD and HMD experimental sessions are presented in Fig. 6 and Fig. 7, respectively.

Non-HMD experimental session. The best results in terms of accuracy have been obtained with XGBoost and RF classifiers on all the considered affective indicators computed with the nine formulas. RF obtained the best results on the three formulations of valence (V1, V2, V3), with f1_scores equal to 1.00 for all the five classes. XGBoost

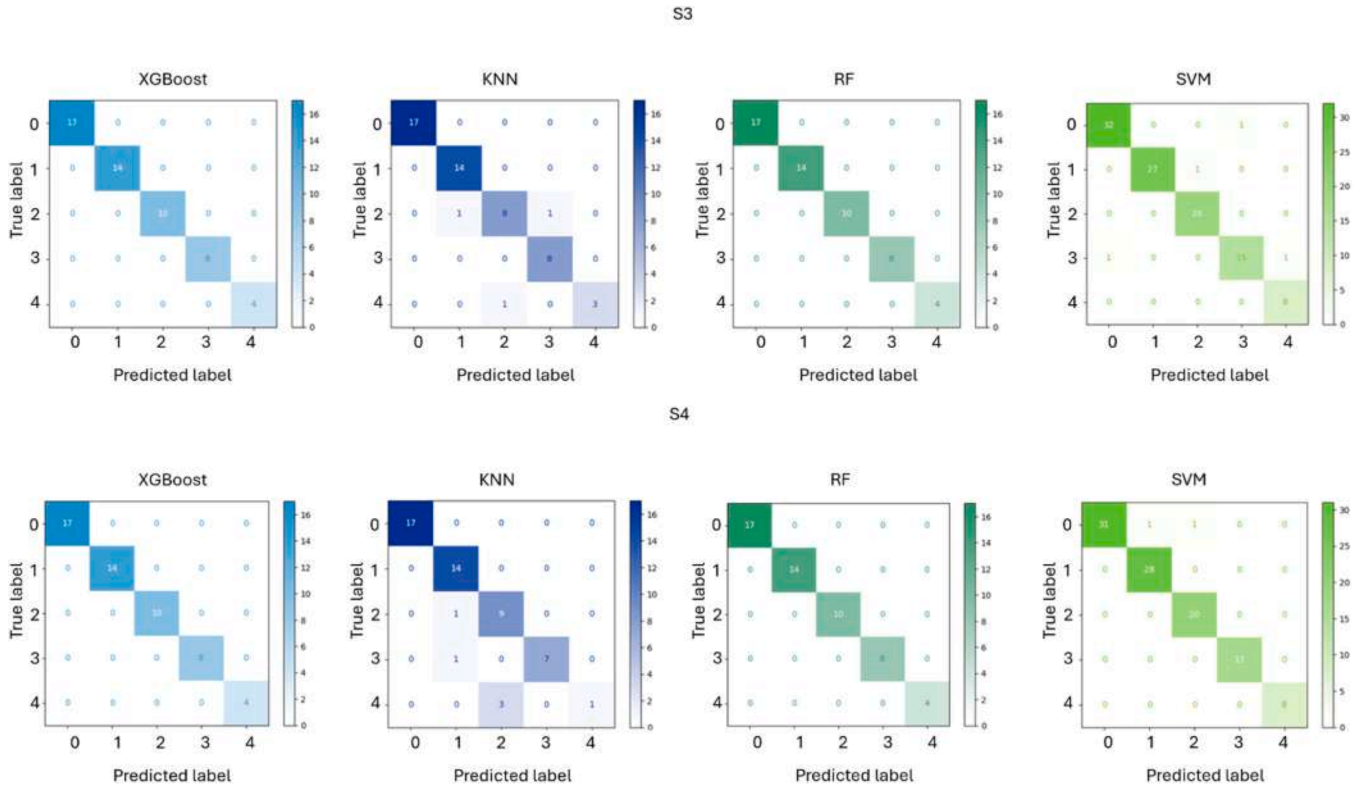

Fig. 6. (*continued*).

obtained f1_scores equal to 1.00 for V1 and V3 on all the 5 classes, and f1_scores of 0.98 and 0.96 for class 3 and class 4, respectively, for V2. For both classifiers, the overall accuracy was outstanding (1.00 and 0.98 for RF and XGBoost, respectively). On the other hand, RF performed slightly worse than XGBoost on the arousal, particularly on A2 (f1_score equal to 0.94 for class 1 and f1_score equal to 0.97 for class 3). On the A1, both decision tree-based classifiers obtained an overall accuracy of 1.00. The dominance and the four formulations of the stress (S1, S2, S3, S4) were correctly classified in the five classes on the complete test set with both RF and XGBoost.

SVM gave notable results, with an overall performance slightly lower than RF and XGBoost. The obtained accuracy on the valence formulations were 1.00 for V1, 0.98 for V2, 0.99 for V3, in line with the results of RF and XGBoost. Slightly lower metrics were obtained also for the stress, with accuracy of 0.97 (S1), 0.99 (S2), 0.96 (S3), and 0.98 (S4), and f1_score equal to 0.91 on class 4 in S3. The worst result obtained with SVM was for A2, with an overall accuracy of 0.96 and a f1_score of 0.80 for class 5, whilst for A1 f1_scores of 0.98 for class 2 and class 4 determined an overall accuracy of 0.99. The accuracy on the dominance was lower than that obtained with RF and XGBoost, but a remarkable value of 0.98 was obtained.

The worst performances have been obtained with kNN on all the affective indicators and all their different formulations. The worst result was obtained on V3 (accuracy equal to 0.72), with an f1_score of 0.48 for class 2, followed by V2 (accuracy equal to 0.77) with an f1_score of 0.60 on class 2. On the contrary, an overall accuracy of 0.98 was obtained on V1 (f1_score equal to 1.00 on class 1, class 2 and class 5). In terms of f1_score, the worst result was given by class 4 on S2 (f1_score equal to 0.43). KNN performed well on the stress, with accuracy of 0.81, 0.96, 0.94 and 0.91 for S1, S2, S3 and S4, respectively.

HMD experimental session. The results obtained reflect the trend highlighted by the classification metrics for the non-HMD experimental session. In fact, XGBoost and RF gave the best results, with overall accuracies ranging from 0.98 to 1.00. RF obtained f1_score equal to 1.00 for each class for V1, V2, and V3; the same result was obtained by XGBoost on V1 and V3, whilst f1_score of 0.99 was obtained for V2. A1 was correctly classified in the five classes on the complete test set with both decision tree-based classifiers; A2 obtained a slightly lower result (accuracy of 0.99 with RF and 0.98 with XGBoost) with the lowest f1_score (0.97) obtained with XGBoost on class 2 and class 4. The dominance obtained outstanding results (accuracy equal to 1.00, f1_score equal to 1.00 on all the five classes) with both RF and XGBoost. Performances on stress confirmed the excellent performances of decision tree-based classifiers; S1 and S4 were classified with accuracy of 1.00 with both RF and XGBoost (f_1 score equal to 1.00 for the five classes with RF and for class 1, class 2 and class 4 with XGBoost; f1_score equal to 0.99 for class 3 and class 5 with XGBoost). XGBoost performed slightly better than RF on S2 obtaining f1_score of 1.00 on class 1, class 2, class 3 and class 5, and f1_score of 0.99 on class 4, whilst RF obtained f1_score of 1.00 only on class 1; f1_score of 0.99 was obtained on the other classes with the exception of class 5 (0.98). XGBoost obtained its worst performance on S3 (f1_score of 0.94 on class 5), but reached values from 0.99 to 1.00 on the other classes and an overall accuracy of 0.99; RF reached the same accuracy, with f1_score of 0.97 (class 4), 0.98 (class 1 and class 5), 1.00 (class 2 and class 3).

While achieving notable results, performances of SVM were lower than those of RF and XGBoost. For the nine formulations of affective indicators, accuracy of 1.00 was only achieved for S2 (f1_score of 1.00 for all classes except class 3); the lowest overall accuracy was obtained for S4 (0.96), with f1_score ranging from 0.95 (class 2) to 0.99 (class 3). The best performance was achieved on S2, for which f1_score of 1.00 was obtained on all classes except class 3 (0.99). On S3, the overall accuracy was comparable to those of RF and XGBoost; SVM outperformed XGBoost only on class 5 (f1_score equal to 0.96). On S1, accuracy of 0.98 was reached, with the lowest f1_score on class 3 (0.96). On A1, SVM performances were slightly lower than those of RF and XGBoost, with an overall accuracy of 0.99; on A2, accuracy was in line with the decision tree-based classifiers (0.98). SVM obtained a lower f1_score on the dominance, particularly on class 2 (0.95), leading anywhere to a remarkable overall accuracy (0.98). Concerning the valence,

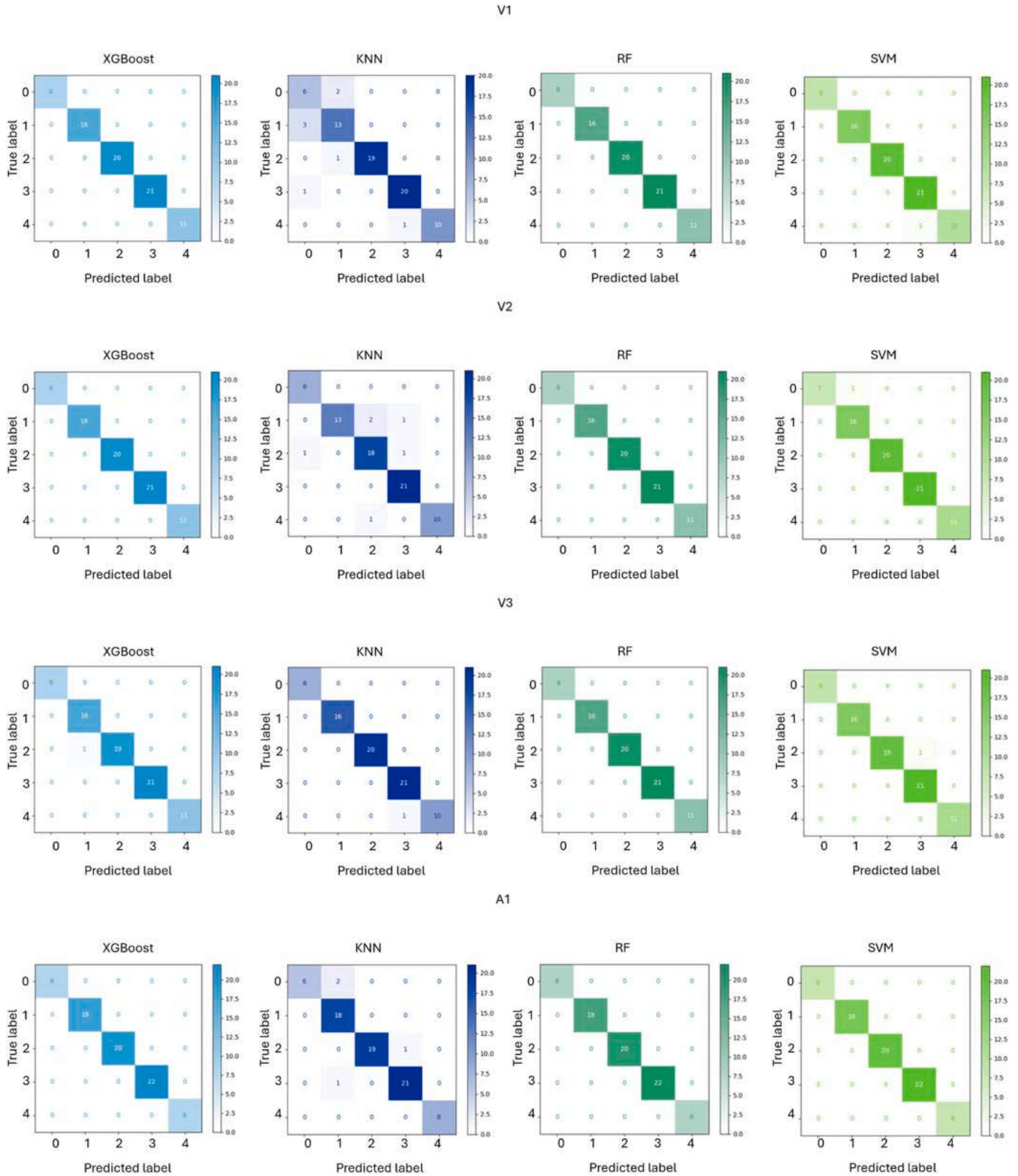

Fig. 7. Confusion matrices for the 5-class classification on the affective indicators from the *HMD experimental session*. Rows: adopted formulas for the affective indicators (V1, V2,…). Columns: adopted classifiers (eXtreme Gradient Boosting – XGBoost, k-Nearest Neighbor – kNN, Random Forest – RF, Support Vector Machine – SVM).

lowest metrics were obtained, with lowest performances on class 3, class 4 and class 5 of V1 (0.98, 0.99 and 0.99, respectively), class 1, class 2, class 3 and class 4 on V2 (0.98, 0.99, 0.97 and 0.97, respectively), and class 1, class 3, class 4 and class 5 on V3 (0.95, 0.98, 0.99 and 0.99, respectively).

As for the non-HMD experimental session, kNN classifier reached the

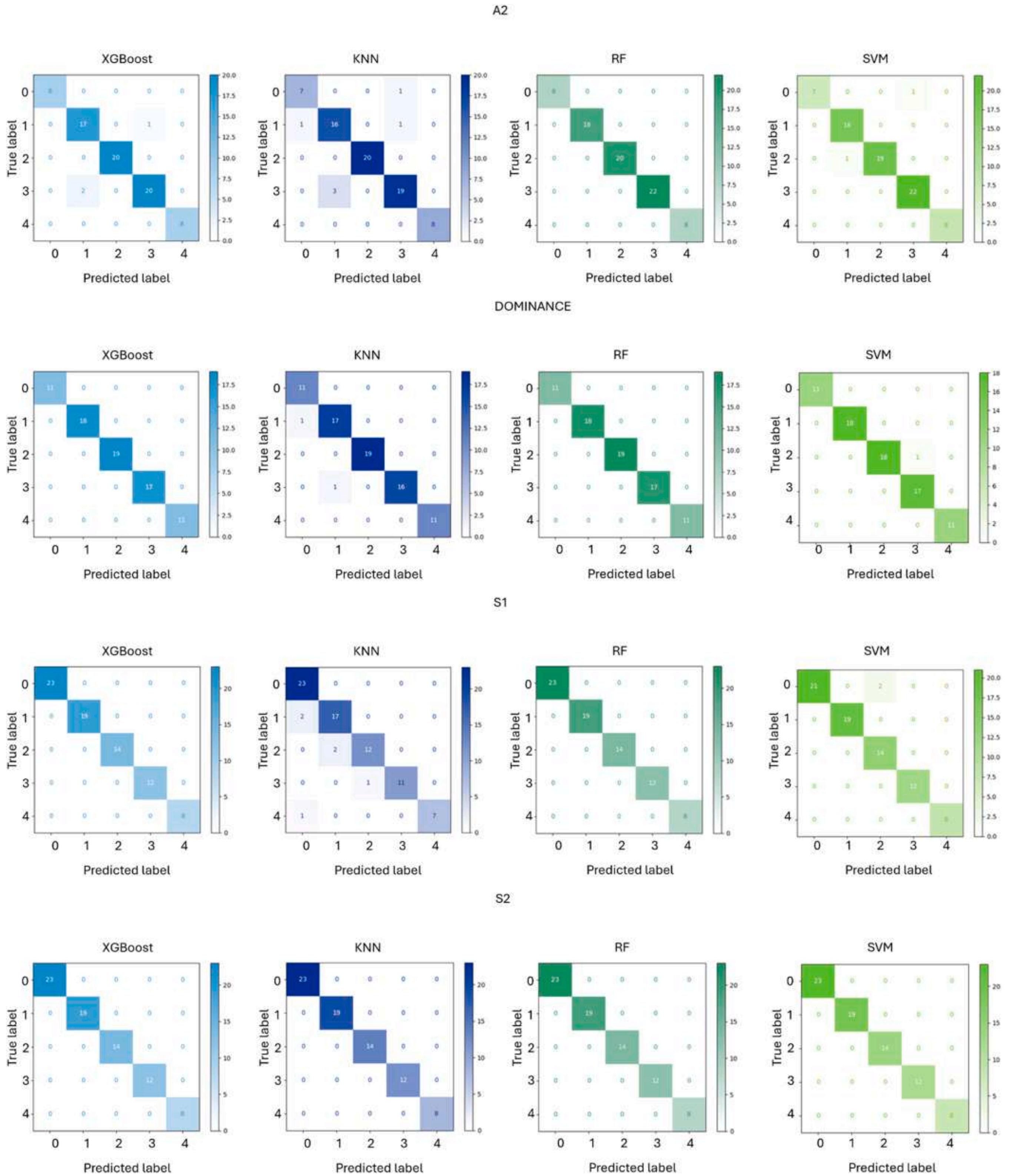

Fig. 7. (*continued*).

worst results. Particularly, for A2, the f1_score on class 1 reached only 0.77. The overall accuracy on A1 (0.93) was lower than that of decision tree-based classifiers of 0.07, and lower than that of SVM of 0.06. The same accuracy was obtained on the dominance, with the lowest f1_score obtained on class 2 (0.89). Class 1 was those with the worst metrics also for V1, where an f1_score of 0.78 was achieved; on the remaining classes, f1_score varies from 0.91 (class 2) to 0.96 (class 5), to 0.98 (class 3 and class 4). For V2, the overall accuracy was lower than that of the other classifiers of around 0.10, with the lowest f1_score on class 1 (0.80). On the contrary, the accuracy on V3 (0.98) was comparable to those of RF, XGBoost and SVM, with the lowest f1_score on class 4 (0.96). Metrics computed for the stress showed that kNN performances were lower than

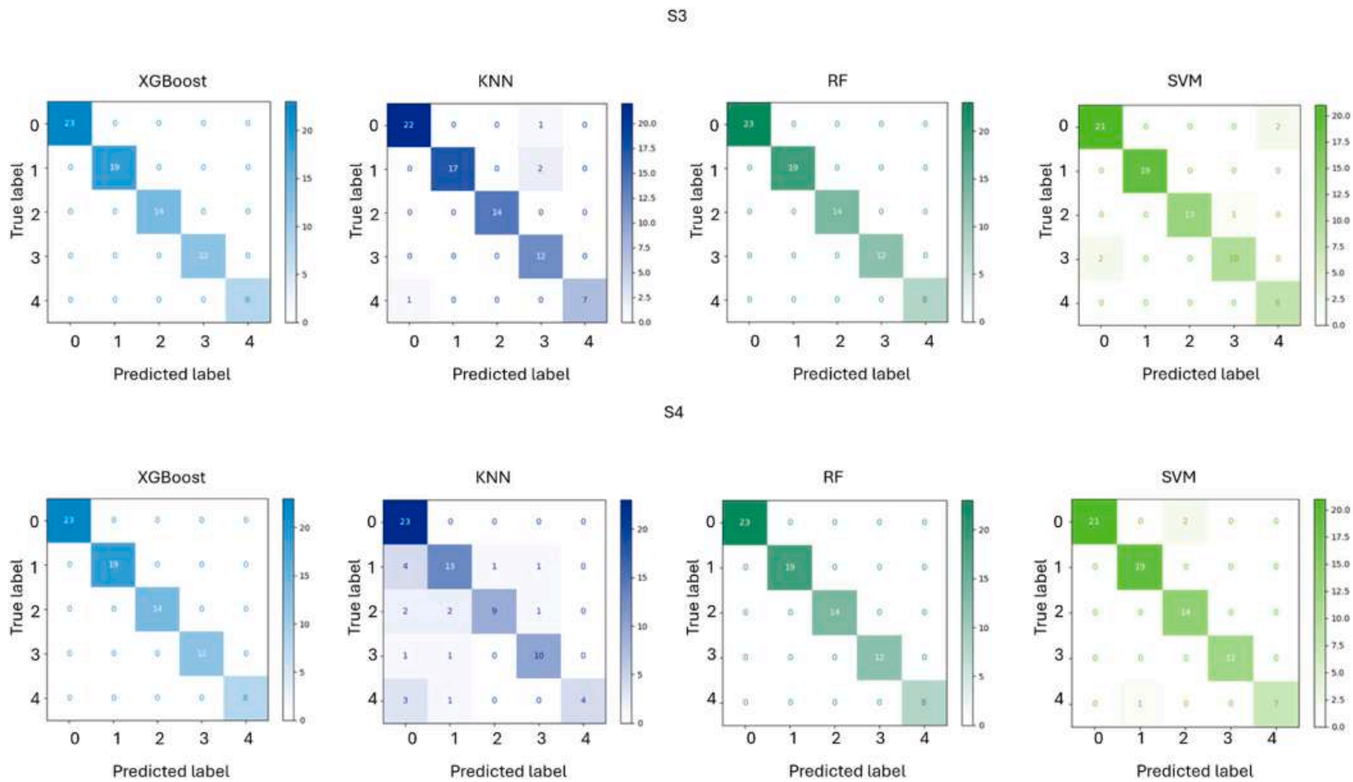

Fig. 7. (*continued*).

the other classifiers; particularly, an overall accuracy of 0.8 was obtained for S4, with f1_score of 0.82, 0.73, 0.80, 0.86 and 0.78 on class 1, class 2, class 3, class 4 and class 5, respectively. Higher accuracy was reached on the other formulations of stress, with values of 0.93 (S1) and 0.94 (S2 and S3).

4. Discussions

The results of the present study show that EEG is an effective tool to quantify and characterize the stress level, at least in experiments where VR is used as affective stimuli. In particular, the nine affective indicators adopted in previous literature served in this study as quantitative descriptors of emotion in two ways. First, the paired *t*-test revealed that the EEG-based indicators are significantly different in the stressful and stressless environments, thus validating the effectiveness of the affective content of our virtual environments and differentiating the two emotional states. Secondarily, ML results show that the nine indicators related to valence, arousal, dominance and stress correlate with the related self-assessed responses, demonstrating the efficacy of the indicators themselves. Although deep learning methods have shown promising results in capturing complex patterns in EEG signals, we chose to investigate features for which interpretability was strictly tied to well-established neuroscientific principles. Our approach focuses on features derived from frequency band power values, including ratios and logarithmic transformations, which capture variations in band activation and their relative interactions. This feature extraction is performed prior to the ML phase, eliminating the need for automated feature learning. By explicitly defining affective indicators our methodology enhances interpretability and ensures greater control over the data, providing valuable insights for a more specific and informed feature engineering process within deep learning architectures. A robust understanding of the features that best represent stress-related emotional states enhance the aggregation of higher-level features and capturing distinct and intricate EEG patterns. To the best of our knowledge, the study offers for the first time a comprehensive list of EEG-based emotional indicators with related formulas, and tests them on validated affective stimuli.

In our study, the ML classifiers that achieved the best performances are XGBoost and RF, endorsing higher accuracy on EEG data obtained using non-emotional indicators [96–98]. Similarly, the effectiveness of SVM is confirmed especially in EEG-based classification of affective states [75,99]. The preferred adoption of SVM classifiers in the domain of EEG-based affective classification and a less frequent adoption of RF and, particularly, XGBoost [73,100], despite the better performances obtained by the second ones, suggest a higher suitability of decision tree-based classifiers in handling feature vectors constructed upon the affective indicators formulas. This consideration seems to be supported by the fact that the majority of the previous studies involving affective indicators tend to label EEG-based feature (time and frequency domain) basing on ratings given to the perceived levels of the affective indicators (e.g., valence, arousal, dominance, like/dislike), but do not associate these ratings to values of affective indicators computed with the formulas provided by the literature. Additionally, this prevalent approach typically involves a classification in only two-three classes (low level, neutral level, high level) or combination of low/high levels of the affective dimensions, reducing the complexity of the emotional state investigated. In this study, a five-class classification was performed, with a direct correlation with SAM rates, further highlighting the efficacy of the proposed EEG-based feature vectors.

The values of the metrics obtained with kNN were the lowest, with accuracies between 0.70 and 0.94, even if in line with previous literature adopting this non-parametric model [68,70,72,75-77]; this seems to be an additional point in favor of adopting the proposed approach when dealing with VR elicitation, particularly with HMD, since the performances of the worst classifier have significantly increased in the HMD experimental session. Interestingly, the classifiers applied to the FAA indicator reported high accuracies both with valence (in the non-HMD experimental session kNN: 0.98, RF: 1.00, XGBoost: 1.00, SVM: 1.00; in the HMD one kNN: 0.94, RF: 1.00, XGBoost: 1.00, SVM: 0.99) and stress labeling (in the non-HMD experimental session kNN: 0.91, RF: 1.00, XGBoost: 1.00, SVM: 0.98; in the HMD one kNN: 0.8, RF: 1.00, XGBoost: 1.00, SVM: 0.96), showing a significant positive correlation [51,82,83].

A further consideration regards the consistency of the results across the different VR approaches and EEG headsets. Indeed, for stress elicitation using classical stressors, a 14-channel setup (used in the non-HMD session) results sufficient to capture variations in EEG-based affective indicators. This finding has practical implications, as the 14-channel system requires less preparation time while still providing reliable data.

The primary next step of research is the development of a real-time deep/machine-learning-based system for stress assessment using EEG and other physiological data, allowing early intervention systems in crucial applications such as wellbeing in workplaces, enhanced rehabilitation treatments, telemedicine and telecare. Also, the adoption of Cave Automatic Virtual Environment (CAVE) as a more immersive VR technology or the investigation of other EEG-based indicators and other stressors could extend the comprehension of stressful situations and mental states. These new developments together with an increase in participants' number will favor a new generalizability and comprehension of stress phenomena.

5. Conclusion

The study presents an analysis of the stressful condition via EEG and VR. Two VR environments have been designed to elicit, respectively, stress and non-stress with the aid of eight micro-stressors inspired by previous literature. 100 volunteers (then reduced to 87) participated in two experiments, one with and one without HMD, where they were asked to navigate the two scenarios.

Relying on literature-acknowledged formulas, nine indicators have been computed from the EEG waves including three related to Valence dimension, two related to Arousal, one to Dominance and three to Stress. A statistically significant difference was registered between the indicators calculated for the participants navigating the stressful environment and those of the same participants navigating the stressless one, showing that our environments are effective in eliciting the desired emotions and, more generally, that VR could be an effective tool to elicit emotions, stress specifically, in individuals.

Also, instant data of these indicators have been labeled according to self-assessments (questionnaires' replies) and classified with Support Vector Machine, XGBoost, k-Nearest Neighbor and Random Forest ML classifiers. The results show that EEG-based affective indicators are reliable to assess the ground truth of valence, arousal, dominance and stress levels. These results, comparable for RF, XGBoost and SVM between the two experimental sessions, significantly increase for kNN in the HMD experimental session, supporting the elicitation effectiveness of this technology. RF, XGBoost and SVM, with slightly better metrics for RF and XGBoost, achieved the best results.

Therefore, the present study demonstrates that the emotional indicators extrapolated from EEG are valid descriptors of emotions, as they correlate with the self-assessment. Moreover, it demonstrates the effectiveness of classical ML approaches in the emotional analysis of EEG signals. Indeed, identifying the most effective features can contribute to the creation of well-structured and robust labeled datasets, which could support transfer learning approaches, reducing the need for large-scale training data and improving model adaptability, thus offering additional advantages for a future deep learning approach.

CRediT authorship contribution statement

Moos Sandro: Project administration, Funding acquisition. **Passavanti Giorgia:** Validation, Data curation. **Celeghin Alessia:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Vezzetti Enrico:** Funding acquisition. **Marcolin Federica:** Writing – original draft, Supervision, Project administration, Methodology, Investigation, Conceptualization. **Castiblanco Jimenez Ivonne Angelica:** Writing – review & editing, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Olivetti Elena Carlotta:** Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Ethical approval

This study was approved by the "Comitato di Bioetica" of the University of Turin on November 30, 2023, with protocol number 06 31 279.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests. Federica Marcolin reports financial support was provided by European Union. Elena Carlotta Olivetti reports a relationship with European Union that includes: funding grants. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was carried out within the MICS (Made in Italy - Circular and Sustainable) Extended Partnership and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) - MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.3 - D.D. 1551.11-10-2022, PE00000004). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

We thank Alessio Lachello, Martina Mancini, Benedetta Mazzetti and Martina Martinengo for contributing to this project. We also thank the participants of the experiment for letting their EEG data be acquired on a voluntary basis.

Appendix

Questionnaire Results

Tables A.1 and A.2 report the mode values (most frequent responses) from questionnaires administered to participants (N = 100; 39 in the non-HMD experimental session and 61 in the HMD experimental session) after their experiences in stressful and stressless virtual environments. The questionnaire evaluated four dimensions on a 5-point Likert scale: valence (1 =very negative, 5 =very positive), arousal (1 =very calm, 5 =very excited), dominance (1 =no control, 5 =complete control), and stress (1 =not stressed, 5 =very stressed).

Table A.1
Mode Values from non-HMD experimental session

	Stressful non-HMD experimental session				Stressless non-HMD experimental session			
	Valence	Arousal	Dominance	Stress	Valence	Arousal	Dominance	Stress
Lack of time	2	4	2	4	3	3	4	2
Lack of salvation	3	3	2	4	3	4	3	1
Lack of engagement	3	3	3	2	4	1	4	1
Lack of control	2	4	2	4	3	3	4	1
Lack of knowledge	2	4	1	4	3	4	3	3
Lack of self-confidence	2	4	3	4	4	4	4	2
Lack of alternatives	3	3	3	1	3	1	3	1
Overchoice	3	3	4	2	4	1	4	1

Table A.2
Mode Values from HMD experimental session

	Stressful HMD experimental session				Stressless HMD experimental session			
	Valence	Arousal	Dominance	Stress	Valence	Arousal	Dominance	Stress
Lack of time	2	4	3	4	4	3	4	2
Lack of salvation	2	4	2	5	4	2	3	2
Lack of engagement	3	3	2	2	3	3	5	1
Lack of control	2	4	1	3	5	2	4	1
Lack of knowledge	2	4	1	4	4	2	4	1
Lack of self-confidence	2	4	3	5	4	4	4	1
Overchoice	3	3	4	1	4	2	4	1

Data Availability

The authors do not have permission to share data.

References

[1] H. Selye, Stress and the general adaptation syndrome, Br. Med J. 1 (1950) 1383–1392.
[2] A. Tully, Stress, sources of stress and ways of coping among psychiatric nursing students, J. Psychiatr. Ment. Health Nurs. 11 (2004) 43–47, https://doi.org/10.1111/j.1365-2850.2004.00682.x.
[3] M. Al'absi, S. Bongard, T. Buchanan, et al., Cardiovascular and neuroendocrine adjustment to public speaking and mental arithmetic stressors, Psychophysiology 34 (1997) 266–275, https://doi.org/10.1111/j.1469-8986.1997.tb02397.x.
[4] P.J. Feldman, S. Cohen, N. Hamrick, S.J. Lepore, Psychological stress, appraisal, emotion and Cardiovascular response in a public speaking task, Psychol. Health 19 (2004) 353–368, https://doi.org/10.1080/0887044042000193497.
[5] H. Oner Altiok, B. Ustun, The Stress Sources of Nursing Students, Educ. Sci. Theory Pr. 13 (2013) 760–766.
[6] M. Shah, S. Hasan, S. Malik, C.T. Sreeramareddy, Perceived Stress, Sources and Severity of Stress among medical undergraduates in a Pakistani Medical School, BMC Med Educ. 10 (2010) 2, https://doi.org/10.1186/1472-6920-10-2.
[7] X. Zupiria Gorostidi, X. Huitzi Egilegor, M. Jose Alberdi Erice, et al., Stress sources in nursing practice. Evolution during nursing training, Nurse Educ. Today 27 (2007) 777–787, https://doi.org/10.1016/j.nedt.2006.10.017.
[8] V.M.E. Bridgland, E.K. Moeck, D.M. Green, et al., Why the COVID-19 pandemic is a traumatic stressor, PLOS ONE 16 (2021) e0240146, https://doi.org/10.1371/journal.pone.0240146.
[9] M. Hoelterhoff, M.C. Chung, Death anxiety and well-being; coping with life-threatening events, Traumatology 19 (2013) 280–291, https://doi.org/10.1177/1534765613477499.
[10] J.R. Averill, Personal control over aversive stimuli and its relationship to stress, Psychol. Bull. 80 (1973) 286–303, https://doi.org/10.1037/h0034845.
[11] S. Mineka, K.A. Kelly, The relationship between anxiety, lack of control and loss of control. In: Stress, personal control and health, John Wiley & Sons, Oxford, England, 1989, pp. 163–191.
[12] S. Glazer, B. Kruse, The role of organizational commitment in occupational stress models, Int J. Stress Manag 15 (2008) 329–344, https://doi.org/10.1037/a0013135.
[13] A. Panaccio, C. Vandenberghe, Perceived organizational support, organizational commitment and psychological well-being: a longitudinal study, J. Vocat. Behav. 75 (2009) 224–236, https://doi.org/10.1016/j.jvb.2009.06.002.
[14] G.A. Haynes, Testing the boundaries of the choice overload phenomenon: The effect of number of options and time pressure on decision difficulty and satisfaction, Psychol. Mark. 26 (2009) 204–212, https://doi.org/10.1002/mar.20269.
[15] T.L. Saltsman, M.D. Seery, C.L. Kondrak, et al., Too many fish in the sea: A motivational examination of the choice overload experience, Biol. Psychol. 145 (2019) 17–30, https://doi.org/10.1016/j.biopsycho.2019.03.010.
[16] B. Schwartz, The Paradox of Choice. In: Positive Psychology in Practice, John Wiley & Sons, Ltd, 2015, pp. 121–138.
[17] A. Baethge, T. Vahle-Hinz, J. Schulte-Braucks, R. van Dick, A matter of time? Challenging and hindering effects of time pressure on work engagement, Work Stress 32 (2018) 228–247, https://doi.org/10.1080/02678373.2017.1415998.
[18] G. Phillips-Wren, M. Adya, Decision making under stress: the role of information overload, time pressure, complexity, and uncertainty, J. Decis. Syst. 29 (2020) 213–225, https://doi.org/10.1080/12460125.2020.1768680.
[19] R. Prem, M. Paškvan, B. Kubicek, C. Korunka, Exploring the ambivalence of time pressure in daily working life, Int J. Stress Manag 25 (2018) 35–43, https://doi.org/10.1037/str0000044.
[20] A.O. de Berker, R.B. Rutledge, C. Mathys, et al., Computations of uncertainty mediate acute stress responses in humans, Nat. Commun. 7 (2016) 10996, https://doi.org/10.1038/ncomms10996.
[21] T.D. Parsons, Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences, Front Hum. Neurosci. 9 (2015), https://doi.org/10.3389/fnhum.2015.00660.
[22] S. Estupiñán, F. Rebelo, P. Noriega, et al., Can virtual reality increase emotional responses (arousal and valence)? a pilot study, in: A. Marcus (Ed.), Design, User Experience, and Usability. User Experience Design for Diverse Interaction Platforms and Environments. Springer International Publishing, Cham, 2014, pp. 541–549.
[23] R. Rivu, R. Jiang, V. Mäkelä, et al., Emotion elicitation techniques in virtual reality, in: C. Ardito, R. Lanzilotti, A. Malizia, et al. (Eds.), Human-Computer Interaction – INTERACT 2021. Springer International Publishing, Cham, 2021, pp. 93–114.
[24] G. Riva, F. Mantovani, C.S. Capideville, et al., Affective interactions using virtual reality: the link between presence and emotions, Cyber Behav. 10 (2007) 45–56, https://doi.org/10.1089/cpb.2006.9993.
[25] B. Barbot, J.C. Kaufman, What makes immersive virtual reality the ultimate empathy machine? Discerning the underlying mechanisms of change, Comput. Hum. Behav. 111 (2020) 106431, https://doi.org/10.1016/j.chb.2020.106431.
[26] M. Li, J. Pan, Y. Gao, et al., Neurophysiological and subjective analysis of VR emotion induction paradigm, IEEE Trans. Vis. Comput. Graph 28 (2022) 3832–3842.
[27] N. Dozio, F. Marcolin, G.W. Scurati, et al., A design methodology for affective Virtual Reality, Int J. Hum. -Comput. Stud. 162 (2022) 102791, https://doi.org/10.1016/j.ijhcs.2022.102791.
[28] N. Dozio, F. Marcolin, G.W. Scurati, et al., Development of an affective database made of interactive virtual environments, Sci. Rep. 11 (2021) 24108, https://doi.org/10.1038/s41598-021-03380-y.

[29] A. Felnhofer, O.D. Kothgassner, M. Schmidt, et al., Is virtual reality emotionally arousing? Investigating five emotion inducing virtual park scenarios, Int J. Hum. -Comput. Stud. 82 (2015) 48–56, https://doi.org/10.1016/j.ijhcs.2015.05.004.

[30] F. Marcolin, G. Wally Scurati, L. Ulrich, et al., Affective virtual reality: how to design artificial experiences impacting human emotions, IEEE Comput. Graph Appl. 41 (2021) 171–178, https://doi.org/10.1109/MCG.2021.3115015.

[31] S. Susindar, M. Sadeghi, L. Huntington, et al., The feeling is real: emotion elicitation in virtual reality, Proc. Hum. Factors Erg. Soc. Annu Meet. 63 (2019) 252–256, https://doi.org/10.1177/1071181319631509.

[32] P. Diniz Bernardo, A. Bains, S. Westwood, D.C. Mograbi, Mood induction using virtual reality: a systematic review of recent findings, J. Technol. Behav. Sci. 6 (2021) 3–24, https://doi.org/10.1007/s41347-020-00152-9.

[33] T. Luong, A. Lecuyer, N. Martin, F. Argelaguet, A survey on affective and cognitive VR, IEEE Trans. Vis. Comput. Graph 28 (2022) 5154–5171, https://doi.org/10.1109/TVCG.2021.3110459.

[34] F. Pallavicini, A. Ferrari, A. Pepe, et al., Effectiveness of Virtual Reality Survival Horror Games for the Emotional Elicitation: Preliminary Insights Using Resident Evil 7: Biohazard, in: M. Antona, C. Stephanidis (Eds.), Universal Access in Human-Computer Interaction. Virtual, Augmented, and Intelligent Environments. Springer International Publishing, Cham, 2018, pp. 87–101.

[35] R. Somarathna, T. Bednarz, G. Mohammadi, Virtual reality for emotion elicitation – a review, IEEE Trans. Affect Comput. 14 (2023) 2626–2645, https://doi.org/10.1109/TAFFC.2022.3181053.

[36] J.W. Woodworth, C.W. Borst, Designing immersive affective tasks for emotion elicitation in virtual reality, 2nd Moment Emot. elicitation capture Workshop (2021).

[37] P. Bilgin, K. Agres, N. Robinson, et al., A comparative study of mental states in 2D and 3D virtual environments Using EEG, 2019 IEEE Int. Conf. Syst., Man Cybern. (SMC) (2019) 2833–2838.

[38] K. Hidaka, H. Qin, J. Kobayashi, Preliminary test of affective virtual reality scenes with head mount display for emotion elicitation experiment, : 2017 17th Int. Conf. Control, Autom. Syst. (ICCAS) (2017) 325–329.

[39] M.M. Bradley, P.J. Lang, Measuring emotion: the self-assessment manikin and the semantic differential, J. Behav. Ther. Exp. Psychiatry 25 (1994) 49–59, https://doi.org/10.1016/0005-7916(94)90063-9.

[40] M. Annerstedt, P. Jönsson, M. Wallergård, et al., Inducing physiological stress recovery with sounds of nature in a virtual reality forest — results from a pilot study, Physiol. Behav. 118 (2013) 240–250, https://doi.org/10.1016/j.physbeh.2013.05.023.

[41] P. Jönsson, M. Wallergård, K. Österberg, et al., Cardiovascular and cortisol reactivity and habituation to a virtual reality version of the trier social stress test: a pilot study, Psychoneuroendocrinology 35 (2010) 1397–1403, https://doi.org/10.1016/j.psyneuen.2010.04.003.

[42] C. Kirschbaum, K.-M. Pirke, D.H. Hellhammer, The 'trier social stress test' – a tool for investigating psychobiological stress responses in a laboratory setting, Neuropsychobiology 28 (2008) 76–81, https://doi.org/10.1159/000119004.

[43] M. Wallergård, P. Jönsson, G. Johansson, B. Karlson, A virtual reality version of the trier social stress test: a pilot study, Presence 20 (2011) 325–336, https://doi.org/10.1162/PRES_a_00052.

[44] A.-M. Brouwer, M.A. Neerincx, V. Kallen, et al., EEG alpha asymmetry, heart rate variability and cortisol in response to virtual reality induced stress, J. Cyber Rehabil. 4 (2011) 21–34.

[45] M. Alghamdi, H. Regenbrecht, S. Hoermann, N. Swain, Mild stress stimuli built into a non-immersive virtual environment can elicit actual stress responses, Behav. Inf. Technol. 36 (2017) 913–934, https://doi.org/10.1080/0144929X.2017.1311374.

[46] M.H. Martaindale, W.L. Sandel, A. Duron, M.J. McAllister, Can a virtual reality training scenario elicit similar stress response as a realistic scenario-based training scenario? Police Q 27 (2024) 109–129, https://doi.org/10.1177/10986111231182729.

[47] C. Hanshans, T. Amler, J. Zauner, L. Bröll, Inducing and measuring acute stress in virtual reality: Evaluation of canonical physiological stress markers and measuring methods, J. Environ. Psychol. 94 (2024) 102107, https://doi.org/10.1016/j.jenvp.2023.102107.

[48] X. Xu, Y. Lu, B. Vogel-Heuser, L. Wang, Industry 4.0 and Industry 5.0—Inception, conception and perception, J. Manuf. Syst. 61 (2021) 530–535, https://doi.org/10.1016/j.jmsy.2021.10.006.

[49] X. Li, A. Nassehi, B. Wang, et al., Human-centric manufacturing for human-system coevolution in Industry 5.0, CIRP Ann. 72 (2023) 393–396, https://doi.org/10.1016/j.cirp.2023.04.039.

[50] B.K. Wiederhold, S. Bouchard, Virtual reality for posttraumatic stress disorder, in: B.K. Wiederhold, S. Bouchard (Eds.), Advances in Virtual Reality and Anxiety Disorders, Springer US, Boston, MA, 2014, pp. 211–233.

[51] S.-H. Seo, J.-T. Lee, M. Crisan, Stress and EEG, Converg. Hybrid. Inf. Technol. 27 (2010).

[52] J.F. Alonso, S. Romero, M.R. Ballester, et al., Stress assessment based on EEG univariate features and functional connectivity measures, Physiol. Meas. 36 (2015) 1351, https://doi.org/10.1088/0967-3334/36/7/1351.

[53] P.D. Tyson, Task-related stress and EEG alpha biofeedback, Biofeedback Self-Regul. 12 (1987) 105–119, https://doi.org/10.1007/BF01000012.

[54] T. Hayashi, E. Okamoto, H. Nishimura, et al., Beta activities in EEG associated with emotional stress, Int J. Intell. Comput. Med Sci. Image Process 3 (2009) 57–68, https://doi.org/10.1080/1931308X.2009.10644171.

[55] J. Tarrant, J. Viczko, H. Cope, Virtual reality for anxiety reduction demonstrated by quantitative EEG: A Pilot Study, Front Psychol. 9 (2018), https://doi.org/10.3389/fpsyg.2018.01280.

[56] Yusuf MdSU, M. Islam, S. Roy, Stress identification during sustained mental task and brain relaxation modeling with \beta/\alpha band power ratio, : 2019 4th Int. Conf. Electr. Inf. Commun. Technol. (EICT (2019) 1–5.

[57] H. Altaf, S.N. Ibrahim, N.F.M. Azmin, et al., Machine learning approach for stress detection based on alpha-beta and theta-beta ratios of EEG Signals, : 2021 13th Int. Conf. Inf. Commun. Technol. Syst. (ICTS (2021) 201–206.

[58] T.Y. Wen, N.A. Bani, F. Muhammad-Sukki, S.A.M. Aris, Electroencephalogram (eeg) human stress level classification based on theta/beta ratio, Int J. Integr. Eng. 12 (2020) 174–180.

[59] R. Katmah, F. Al-Shargie, U. Tariq, et al., A review on mental stress assessment methods using EEG Signals, Sensors 21 (2021) 5043, https://doi.org/10.3390/s21155043.

[60] G. Vanhollebeke, S. De Smet, R. De Raedt, et al., The neural correlates of psychosocial stress: a systematic review and meta-analysis of spectral analysis EEG studies, Neurobiol. Stress 18 (2022) 100452, https://doi.org/10.1016/j.ynstr.2022.100452.

[61] X.-W. Wang, D. Nie, B.-L. Lu, Emotional state classification from EEG data using machine learning approach, Neurocomputing 129 (2014) 94–106, https://doi.org/10.1016/j.neucom.2013.06.046.

[62] E.S. Pane, A.D. Wibawa, M.H. Purnomo, Improving the accuracy of EEG emotion recognition by combining valence lateralization and ensemble learning with tuning parameters, Cogn. Process 20 (2019) 405–417, https://doi.org/10.1007/s10339-019-00924-z.

[63] Al-Shargie F, Multilevel Assess. Ment. Stress Using SVM ECOC: EEG Approach (2019).

[64] G. Jun, K.G. Smitha, EEG based stress level identification, : 2016 IEEE Int. Conf. Syst., Man, Cybern. (SMC) (2016) 003270–003274.

[65] K.G. Smitha, N.Y. Xin, S.S. Lian, N. Robinson, Classifying subjective emotional stress response evoked by multitasking using EEG, : 2017 IEEE Int. Conf. Syst., Man, Cybern. (SMC) (2017) 3036–3041.

[66] A. Arsalan, M. Majid, A.R. Butt, S.M. Anwar, Classification of perceived mental stress using a commercially available EEG headband, IEEE J. Biomed. Health Inf. 23 (2019) 2257–2264, https://doi.org/10.1109/JBHI.2019.2926407.

[67] P. Arpaia, N. Moccaldi, R. Prevete, et al., A wearable EEG instrument for real-time frontal asymmetry monitoring in worker stress analysis, IEEE Trans. Instrum. Meas. 69 (2020) 8335–8343, https://doi.org/10.1109/TIM.2020.2988744.

[68] J. Li, Z. Zhang, H. He, Hierarchical convolutional neural networks for EEG-based emotion recognition, Cogn. Comput. 10 (2018) 368–380, https://doi.org/10.1007/s12559-017-9533-x.

[69] S. Koelstra, C. Muhl, M. Soleymani, et al., DEAP: a database for emotion analysis; using physiological signals, IEEE Trans. Affect Comput. 3 (2012) 18–31, https://doi.org/10.1109/T-AFFC.2011.15.

[70] G.K. Verma, U.S. Tiwary, Affect representation and recognition in 3D continuous valence–arousal–dominance space, Multimed. Tools Appl. 76 (2017) 2159–2183, https://doi.org/10.1007/s11042-015-3119-y.

[71] M. Gaertner, R. Buettner, D. Sauter, et al., Multi--class emotion recognition within the valence-arousal-dominance, Space Using EEG (2021).

[72] J.L.M. Kumar, M. Rashid, R.M. Musa, et al., The classification of EEG-based winking signals: a transfer learning and random forest pipeline, PeerJ 9 (2021) e11182, https://doi.org/10.7717/peerj.11182.

[73] Z.-T. Liu, S.-J. Hu, J. She, et al., Electroencephalogram emotion recognition using combined features in variational mode decomposition domain, IEEE Trans. Cogn. Dev. Syst. 1–1 (2023), https://doi.org/10.1109/TCDS.2022.3233858.

[74] S. Katsigiannis, N. Ramzan, DREAMER: a database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices, IEEE J. Biomed. Health Inf. 22 (2018) 98–107, https://doi.org/10.1109/JBHI.2017.2688239.

[75] L. Abdel-Hamid, An Efficient machine learning-based emotional valence recognition approach towards wearable EEG, Sensors 23 (2023) 1255, https://doi.org/10.3390/s23031255.

[76] G. Pei, Q. Shang, S. Hua, et al., EEG-based affective computing in virtual reality with a balancing of the computational efficiency and recognition accuracy, Comput. Hum. Behav. 152 (2024) 108085, https://doi.org/10.1016/j.chb.2023.108085.

[77] A.S.M. Miah, J. Shin, Md.M. Islam, et al., Natural human emotion recognition based on various mixed reality(mr) games and electroencephalography (EEG) Signals, : 2022 IEEE 5th Eurasia Conf. Educ. Innov. (ECEI) (2022) 408–411.

[78] T.B. Alakus, M. Gonen, I. Turkoglu, Database for an emotion recognition system based on EEG signals and various computer games – GAMEEMO, Biomed. Signal Process Control 60 (2020) 101951, https://doi.org/10.1016/j.bspc.2020.101951.

[79] M. Yu, S. Xiao, M. Hua, et al., EEG-based emotion recognition in an immersive virtual reality environment: From local activity to brain network features, Biomed. Signal Process Control 72 (2022) 103349, https://doi.org/10.1016/j.bspc.2021.103349.

[80] M. Bakker, C.H.J. Hartgerink, J.M. Wicherts, H.L.J. van der Maas, Researchers' intuitions about power in psychological research, Psychol. Sci. 27 (2016) 1069–1077, https://doi.org/10.1177/0956797616647519.

[81] K. Suresh, S. Chandrashekara, Sample size estimation and power analysis for clinical research studies, J. Hum. Reprod. Sci. 5 (2012) 7–13, https://doi.org/10.4103/0974-1208.97779.

[82] R. Ramirez, J. Planas, N. Escude, et al., EEG-based analysis of the emotional effect of music therapy on palliative care cancer patients, Front Psychol. 9 (254) (2018) 1–7.

[83] M. Aldayel, M. Ykhlef, A. Al-Nafjan, Deep Learning for EEG-based preference classification in neuromarketing, Appl. Sci. 10 (2020) 1525, https://doi.org/10.3390/app10041525.

[84] I.A. Castiblanco Jimenez, F. Marcolin, L. Ulrich, et al., Interpreting emotions with EEG: an experimental study with chromatic variation in VR, in: S. Gerbino, A. Lanzotti, M. Martorelli, et al. (Eds.), Advances on Mechanics, Design Engineering and Manufacturing IV. Springer International Publishing, Cham, 2023, pp. 318–329.
[85] I.A. Castiblanco Jimenez, F. Nonis, E.C. Olivetti, et al., Exploring User Engagement in Museum Scenario with EEG—A Case Study in MAV Craftsmanship Museum in Valle d'Aosta Region, Italy, Electronics 12 (2023) 3810, https://doi.org/10.3390/electronics12183810.
[86] R. Ramirez, Z. Vamvakousis, in: F.M. Zanzotto, S. Tsumoto, N. Taatgen, Y. Yao (Eds.), Detecting Emotion from EEG Signals Using the Emotive Epoc Device, Brain Informatics, Springer, Berlin, Heidelberg, 2012, pp. 175–184.
[87] Giraldo S., Ramirez R. (2013) Brain-activity-driven real-time music emotive control. In: The 3rd International Conference on Music & Emotion, Jyväskylä, Finland, June 11-15, 2013. University of Jyväskylä, Department of Music.
[88] A. Al-Nafjan, M. Hosny, Y. Al-Ohali, A. Al-Wabil, Review and classification of emotion recognition based on EEG brain-computer interface system research: a systematic review, Appl. Sci. 7 (2017) 1239, https://doi.org/10.3390/app7121239.
[89] S. Hwang, H. Jebelli, B. Choi, et al., Measuring workers' emotional state during construction tasks using wearable EEG, J. Constr. Eng. Manag 144 (2018) 04018050, https://doi.org/10.1061/(ASCE)CO.1943-7862.0001506.
[90] Y. Liu, O. Sourina, EEG-based dominance level recognition for emotion-enabled interaction, : 2012 IEEE Int. Conf. Multimed. Expo. (2012) 1039–1044.
[91] T. Chen, C. Guestrin, XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, NY, USA, 2016, pp. 785–794.
[92] M.A. Chandra, S.S. Bedi, Survey on SVM and their application in imageclassification, Int J. Inf. Technol. 13 (2021) 1–11, https://doi.org/10.1007/s41870-017-0080-1.
[93] M.L.R. Menezes, A. Samara, L. Galway, et al., Towards emotion recognition for virtual environments: an evaluation of eeg features on benchmark dataset, Pers. Ubiquitous Comput. 21 (2017) 1003–1013, https://doi.org/10.1007/s00779-017-1072-7.
[94] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Trans. Inf. Theory 13 (1967) 21–27, https://doi.org/10.1109/TIT.1967.1053964.
[95] G. Biau, E. Scornet, A random forest guided tour, TEST 25 (2016) 197–227, https://doi.org/10.1007/s11749-016-0481-7.
[96] S.R. Kulkarni, P. Patil, EEG-Based emotion recognition using an ensemble learning algorithm, in: S.M. Thampi, J. Mukhopadhyay, M. Paprzycki, K.-C. Li (Eds.), International Symposium on Intelligent Informatics, Springer Nature, Singapore, 2023, pp. 271–279.
[97] S. Parui, A.K. Roshan Bajiya, D. Samanta, N. Chakravorty, Emotion recognition from eeg signal using xgboost algorithm, : 2019 IEEE 16th India Counc. Int. Conf. (INDICON) (2019) 1–4.
[98] N. Pusarla, A. Singh, S. Tripathi, Ensemble algorithms for EEG based emotion recognition, : 2020 Natl. Conf. Commun. (NCC) (2020) 1–4.
[99] T.Y. Wen, S.A. Mohd Aris, Hybrid approach of EEG stress level classification using k-means clustering and support vector machine, IEEE Access 10 (2022) 18370–18379, https://doi.org/10.1109/ACCESS.2022.3148380.
[100] J. Zong, X. Xiong, J. Zhou, et al., FCAN–XGBoost: a novel hybrid model for EEG Emotion Recognition, Sensors 23 (2023) 5680, https://doi.org/10.3390/s23125680.

Federica Marcolin is Associate Professor at the Department of Production and Management Engineering of the Politecnico di Torino, Italy. After a Master Degree in Mathematics in Engineering in 2010, she discussed her PhD in Production Systems & Industrial Design on January 2015 with a dissertation on Differential Geometry applied to 3D Face Analysis. In 2014 she was at the i·Bug group of Imperial College London to work on the same topics and now is dealing with 3D Face Analysis, Face Expression Recognition, Human-Computer Interaction, Affective Computing, emotional elicitation, monitoring and categorization.

Elena Carlotta Olivetti is Assistant Professor in the Department of Management and Production Engineering of the Politecnico di Torino, Italy. In 2018, she received a Master Degree in Biomedical Engineering at the Politecnico di Torino. In 2022, she received her PhD in Management, Production and Design from the same university. Her dissertation concerned the study of three-dimensional human face for the maxillofacial surgeries' planning. She worked on 3D Face Analysis for recognition and medical purposes, Medical Imaging, Machine Learning for emotional categorization with facial data and psychophysiological signals (electroencephalographic, electrodermal, cardiac).

Ivonne Angelica Castiblanco Jimenez is currently a Post Doc at the Department of Psychology of the University of Turin, Italy, but was working at the Department of Management and Production Engineering of Politecnico di Torino while she conducted the research presented in this study. She received her PhD in Management, Production and Design in 2023 from Politecnico di Torino, Italy. In her dissertation she discussed the role of engagement in product design. Her research topics include the study of electroencephalographic signals for emotional monitoring, user experience, emotional design.

Giorgia Passavanti received her master's degree in Biomedical Engineering in 2024 from Politecnico di Torino, Italy. Her thesis focused on the development of stressful Virtual Reality environments and on the analysis of EEG signals.

Sandro Moos received the master's degree in Mechanical Engineering in 2000. He is currently Full Professor of Design Tools and Methods at the Department of Management and Production Engineering of the Politecnico di Torino, Italy. His current research interests concern Computer-Aided Design applied to Medical Applications, 3D Scanners, Virtual and Augmented Reality, 3D Technologies and Tools for Surgery.

Enrico Vezzetti is Full Professor of Technical Drawing and Design Tools for Industrial Engineering at the Department of Production and Management Engineering of the Politecnico di Torino, Italy. He received the Master Degree in Management Engineering from the Politecnico di Torino in 1997, and the PhD in Manufacturing Engineering from the University of Parma in 2001. In 1998 he was in the 3D Printing Laboratory of the Massachusetts Institute of Technology, USA. His research interests include Computer-Aided Design, 3D Scanners, Virtual and Augmented Reality, Face Recognition, 3D Technologies and Tools for Surgery, Human-Computer Interaction and Artificial Intelligence.

Alessia Celeghin is Associate Professor at the Department of Psychology of the University of Turin, Italy. After a master's degree in clinical psychology in 2009, she discussed her PhD in Psychology, Psychiatry and Neuroscience in May 2014 with a dissertation on the conscious and unconscious dissociations in visual perception. In 2013 she was at the Maastricht Brain Imaging Center, working on neuroimaging techniques applied to affective neuroscience under the supervision of prof. DeGelder, while in 2015 she was at the Laboratory of Cognition and Emotion leaded by prof. Pessoa at Maryland University working on computational analysis of neuroimaging data. She now is dealing with Affective, Cognitive and Visual Neuroscience from psycophysiological, neural and computational perspectives.