(Article begins on next page)

Electrodermal Activity in the Evaluation of Engagement for Telemedicine Applications

1st Giulia Masi
Università di Torino
Turin, Italy
giulia.masi@unito.it

2nd Gianluca Amprimo
Politecnico di Torino
Turin, Italy
gianluca.amprimo@polito.it

3rd Irene Rechichi
Politecnico di Torino
Turin, Italy
irene.rechichi@polito.it

4th Claudia Ferraris
CNR Italy
Turin, Italy
claudia.ferraris@cnr.ieiit.it

5th Lorenzo Priano
Università di Torino
Turin, Italy
lorenzo.priano@unito.it

***Abstract*—Electrodermal Activity (EDA) is a broadly-investigated physiological signal, whose behaviour is connected to nervous system arousal. Such system, indeed, influences the properties of the skin, producing a measurable electrical signal. Among the possible applications of such measurements, several studies have correlated the signal behaviour to engagement during mental and physical tasks, and the subjects' response to specific multimodal stimuli. Also due to the possibility of performing remote assessment and rehabilitation, telemedicine applications are gaining ground in the healthcare system. However, acceptance and engagement, hence continuity of usage, still remain significant obstacles. Therefore, it would be highly beneficial to verify, through objective measures, if these solutions are actually providing a sufficient stimulation to properly engage subjects while playing. This study investigates the possibility of employing EDA in the automatic recognition of different levels of user engagement, while playing a motor-cognitive exergame specifically designed for this purpose. Preliminary results, obtained on a cohort of 25 healthy subjects, seem to confirm that features extracted from EDA analysis are significant and able to train supervised classifiers, achieving high accuracy and precision in the engagement recognition problem.**

***Index Terms*—EDA, Telemedicine, Exergames, User's Engagement**

I. Introduction

Electrodermal Activity (EDA), also known as Galvanic Skin Response or Skin Conductance is widely used in the field of neuroscience. It is the continuous variation in the skin electrical characteristics due to sympathetic nervous system arousal. Indeed, different emotional states and stimuli can provoke the activation of the sympathetic system that will influence the sweat glands and pores functioning, modifying skin conductance properties. In the literature, the EDA is used in evaluating the arousal level for various purposes: stress assessment, classification of emotional states, engagement recognition and assessment of the subjects' response to various stimuli (relaxing music, images, standardized tasks) [1]–[5]. Nowadays, many sectors are converting to smart, automatic solutions which require advanced human-computer interaction approaches. As a result, there is an increasing interest in technologies that could understand the needs of customers or potential customers. For instance, automatically recognising the emotional effect due to human-technology interaction could allow for optimisation in the learning process in education or for an increase in engagement in the entertainment industry, particularly video-gaming [6], [7]. Also in the medical field, telemedicine is rapidly gaining ground [8], especially in the wake of the COVID-19 pandemic. Patient-support facilities, diagnostic procedures and follow-up protocols are being converted into remote solutions, making extensive use of technological aids of various kinds. Consequently, monitoring and regulating the arousal state in order to optimise the design and customisation of telemedicine applications can be useful in this framework. Indeed, many remote protocols involve contact between the caregiver and the patient via a device and/or envisage replacing or complementing the in-person health assessment with alternative measures to be carried out at home or in the outpatient clinic. For instance, several attempts to implement e-Health platforms for motor and cognitive assessment and rehabilitation can be found in the literature [9], [10]. Among those, an approach of increasing interest consists in the use of serious games and exergames paradigms to gamify traditional exercises for assessment and rehabilitation purposes [11]–[15]. These telemedicine solutions are often targeting the elderly population, that is constantly growing and commonly presents a complex clinical picture. Indeed, telemedicine solutions would be a great economic advantage and could guarantee continuity of care if applied to this group of subjects. Unfortunately, this segment of the population is usually also the most wary of technology adoption [16], which could make these telemedicine solutions substantially useless or less performing than those in person. Being able to determine *a-priori* and in an objective and quantitative manner whether an interface, a protocol or a device is able to stimulate the desired emotional response in the user (e.g., engagement, attention, fear) can be pivotal in this situation. Notably, this is crucial in the case of exergames and rehabilitation protocols, where entertainment and motivation are essential to get the user/patient to perform the assigned tasks and guarantee continuity of use. Attempts to evaluate the arousal state, in order to optimise the platforms and games, have already been carried out in this field [17]–[20], employing EDA as well [21]. The novelty presented in this paper is set within this last framework: the aim is to establish if EDA could be used to identify distinct levels of engagement while playing a custom exergame, designed for remote cognitive and motor rehabilitation. The developed exergame has four levels in which different elements are progressively introduced to solicit the users' response, providing four levels of user engagement to be measured. In this pilot study, twenty-five young subjects were enrolled to play the exergame while wearing the Empatica E4 sensor for EDA recording. The measured EDA was analysed through different approaches: 26 features were computed in each level. An automatic recognition of the engagement reached in each game level was attempted using these features. The analysis of the obtained results allowed preliminary conclusions to be drawn on the use of EDA for this purpose.

II. Materials and Methods

A. Setup and experimental protocol

This preliminary study involved 25 healthy young subjects (27±7 years old; 11 females, 14 males; no mental or physical disorders, no medications). Each subject played a custom-developed exergame while seated in front of a RGB-Depth camera (i.e., Microsoft Azure Kinect). The game involves the use of the dominant arm and hand to complete each level. The detailed description of the exergame can be found in the following dedicated paragraph. During the exergame play, the CE-medical certified Empatica E4 device was employed for recording EDA. The wristband was applied on the not-used arm in order to minimise the artefact due to movement and it was worn at least ten minutes before the game beginning. The Empatica device provides the EDA as a Comma Separated Value (CSV) file containing 4Hz-sampled EDA values in μS and also including the initial time of the recording stored as an unique timestamp in Coordinated Universal Time (UTC). This last information is used to synchronize the EDA samples with the starting time of each game level.

B. Exergame

The exergame has been developed in Unity and uses Microsoft Azure Kinect to capture real-time movements of arm and hand. The game was originally designed for high motor and cognitive involvement, by stimulating goal-oriented movements of the upper limb. It was created for rehabilitation/evaluation purposes and appropriately adapted for this preliminary investigation on EDA in healthy young subjects, stressing the engagement modulation aspect. The simplicity and non-invasiveness of the solution allows it to be played also in home environments, thus making it suitable for any Telemedicine application aimed at these purposes. The exergame includes four game levels designed to increase the difficulty of the task and the subject's engagement. Each subject is asked to keep the dominant hand in front of the RGB-Depth camera, at about a 50–60 cm distance from it. On the screen, the virtual hand follows the movements of the real hand, through the GMH-D algorithm [22], which combines Google Mediapipe Hands solution with the depth sensor of the Azure Kinect for high accuracy tracking [23]. The game consists of grasping the correct object, among the four displayed, whose shape and colour are specified through textual on-screen instructions, and carrying it into the box displayed at the bottom of the game scene. This needs to be done for all scheduled objects in each level and thus requires

Fig. 1: Example of a user performing the exergame. The game scene of level 1 is shown here.

the execution of specific hand actions including grab, drag, and release. An example of the game scene is shown in Figure 1. The levels were designed to increase the engagement and difficulty progressively. For this purpose, background music is played at an increasing speed and the time available to complete each movement (i.e., grab, drag, and release) is shortened, thus making the game progressively more pressing [24]. In addition, specific elements are introduced in each level: in the second level, multiple objects displayed on screen have the same colour or shape as the one specified by the on-screen instruction; in the third level, the on-screen instruction is displayed in a discordant color with respect to the color of the required object (Stroop Test style [25]); in the fourth and last level, moving objects are added, i.e., the box in which to release the objects continuously moves along the horizontal direction, making the task very difficult. In this fashion, levels 1 and 2 – that are conceived to make the user familiar with the game interface and are considered as low engagement levels – are not expected to provoke a strong response in users, whereas levels 3 and 4 should provide a strong response, as they contain high engaging features [26]. In particular, level 4 introduces a sudden change in difficulty and requires a different approach to be completed. As a consequence, the user has to come up with a new strategy in the shortest time possible, inducing a sudden increase in the required cognitive skills (e.g., attention, logic), hence causing additional involvement and challenge.

C. EDA Analysis and Feature Extraction

The EDA analysis and feature extraction were performed in MATLAB R2021a. There is not a standardised protocol for the EDA analysis; hence, a multi-approach was carried out. Twenty-six features were computed in four main types of analysis, considering: (1) the EDA signal amplitude, computing descriptive statistical features; (2) the EDA signal trend; (3) the frequency domain; (4) the EDA signal decomposition in tonic and phasic components. For the first type of analysis, the EDA signal was firstly low pass filtered with a cut-off frequency of 1.5 Hz, as in [27]. The z-score of the whole signal and the descriptive statistical features of its amplitude in each level were computed. The first derivative of the EDA signal was also calculated and used for feature extraction regarding EDA trend. In addition, Hjorth features (Activity and Mobility) [28] were calculated on the EDA amplitude

TABLE I: Features Description

Features	Description
range_EDA	Range of z-scored EDA (zEDA) signal
kurt_EDA	Kurtosis of zEDA signal
std_EDA	Standard Deviation of zEDA signal
mean_EDA	Mean of zEDA signal
max_der	Absolute maximum of the EDA first derivative
meannegder	Mean of the negative EDA first derivative
Activity	(Hjorth) Variance of EDA signal
Mobility	(Hjorth) Square root of the variance of the first derivative of EDA divided by activity
P_B5	Power in the [0.05 0.5] Hz band of EDA signal
P_90	Frequency corresponding to the 90% of spectral power
P_std	Standard Deviation of the Power Spectrum
cvx_Tmean	Mean of the Tonic component
cvxT_range	Range of the Tonic component
ske_cvxT	Skewness of the Tonic component
kur_cvxT	Kurtosis of the Tonic component
cvxTmedian	Mean of the Tonic component
cvxTstd	Standard Deviation of the Tonic component
cvx_Pmean	Mean of the Phasic (p) Component
cvxP_range	Range of Phasic (p) Component
ske_cvxP	Skewness of the Phasic (p) Component
kur_cvxP	Kurtosis of the Phasic (p) Component
cvxPmedian	Mean of the Phasic (p) Component
cvxPstd	Standard Deviation of the Phasic (p) Component
cvx_Pmax	Maximum value of the Phasic (p) Component
cvx_Prmax	Maximum value of the Phasic (r) Component

in order to explore their usability also in the EDA Analysis framework, as suggested in [29]. Frequency domain analysis of the EDA signal was performed by considering each level separately as well. The power spectrum was calculated on the filtered EDA signal (no z-scored) through the Fast Fourier Transform. A detailed description of these features is reported in Table I. Moreover, the EDA is usually evaluated in controlled experimental protocols, where a known stimulus is administered to the user. Therefore, the signal analysis takes into account specific activities (usually a single peak or a train of peaks) that occur after the stimulus by computing features concerning number of peaks, latency, and decay time. This kind of analysis is very popular and the Ledalab software [30] and EDA Explorer [31] are common tools employed for it. Nevertheless, the protocol used in this work does not involve the use of controlled stimuli because it measures the EDA during the execution of an exergame that involves continuous and differentiated stimulation over a prolonged period of time. This makes it impossible to identify a single stimulus response window and requires taking into account the presence of simultaneous and different stimuli. As a consequence, we employed the EDA analysis decomposition (tonic and phasic components) through convex optimisation (cvxEDA), as proposed by [32]. In [32], indeed, the implementation of the model under the cvxEDA algorithm is physiologically inspired and directly deals with the not controlled interstimulus interval (ISI) aspect, as it is in real life conditions. This is why this approach was chosen from the rich landscape of EDA analysis tools available in the literature. On the other hand, the algorithm cvxEDA is prone to typical recording artefacts such as detachment. Indeed, wearable wristbands, such as the Empatica E4, rely on two metal sensors leaning against the skin. During motion or in the case of pressure applied, the sensors can be slightly moved and cause loose contact with the skin (detachment). In this case, the signal will present sudden changes in amplitude. Physiologically, the EDA signal cannot have sudden droppings; thus, the cvxEDA algorithm used cannot deal with them. Hence, all signals were preliminarily visually checked for the presence of such artefacts. The set-up chosen for this study guaranteed a robust signal (the wrist for the data recording was not involved in the game and the participants were asked to keep it still while playing); indeed, no detachment artefacts were found. Moreover, visual inspection of the accelerometer signal, also available from Empatica E4, confirmed the absence of wrist motion during the game execution. From the decomposition found from the cvxEDA algorithm, the tonic component (t), the phasic component (r) and the sparse SMNA driver of phasic component (p) were identified. From them, the last part of the EDA analysis was performed, and 14 features were further estimated in each level (for a total of 26 features per level obtained from the four analysis performed). The mean, range, median, standard deviation, skewness and kurtosis of t and p were estimated for each level, as well as the maximum value of p and r, as shown in Table I.

D. Features Analysis and Classification

In order to understand the behaviour and significance of the extracted features and to create graphical representation, Python (Plotly and Seaborn libraries) and Jamovi tools were exploited [33]. The four game levels were also rearranged in 2 classes: levels 1 and 2 were considered the Low Engagement class while levels 3 and 4 were considered the High Engagement one. This was done not only to simplify the visualization, but also to better understand the discriminating power of the features in a binary or multi-class classification problem. On top of that, this arrangement follows the original structure designed for the game, as explained in section II-B. To empirically understand the discriminating power of the features, radar plots of the averaged values (among subjects) of the features for each level were observed. Furthermore, the correlation matrix of the 26 features plus the game level label was computed (Spearman's correlation coefficients ρ). The game level label, indeed, corresponds to the progressively increasing level of engagement (game level 1 = engagement level 1, game level 2 = engagement level 2 and so on), thus virtually representing the true class of belonging. Consequently, the most label-related features will likely provide good discriminatory power of engagement level. The correlation matrix was used to identify the redundant features and determine a threshold of correlation to the class label (ρ_L). In this way, the resulting sub-group of features can be fed to the classification algorithms. In particular, to determine whether the EDA signal is able to distinguish different levels of engagement using this setup, automatic classification was exploited. For the classification different supervised models were tested; namely, the Support Vector Machine (SVM), Random Forest (RF) and k-Nearest Neighbour (k-NN) classifiers. The classifiers were evaluated through nested cross-validation (10 outer train-test splits, 3 internal splits). Regarding SVM, kernel, cost of misclassification (C), γ and degree were optimised through a grid search approach in the internal cross-validation. As for RF, several values of the maximum depth of the trees were explored as well as several k values for k-NN. For a better insight into the model performance, the confusion matrices, reporting precision of classification for each label, were computed averaging the single confusion matrices obtained in the 10 train-test splits of outer cross-validation. Precision is deemed as an important metric, considering that our goal is to precisely identify the different engagement level reducing as much as possible the number of false positives while doing so. After excluding the redundant features (cross-correlation coefficient > 0.9), the algorithms were tested with three different feature sets for both binary and multi-class problems: using all the features, using only the features with $\rho_L > 0.4$ and, lastly, using only the features with $\rho_L > 0.6$.

III. Results and Discussion

A. Features Relevance

The majority of the features do not present with a normal distribution, following the p-value of the Shapiro-Wilk test ($\rho - value < 0.001$), $Pcvx\_B5$ and ske_cvxT make an exception. For them it is not possible to infer the distribution characteristics. The distribution of the values taken by the selected features in the low-engagement levels versus high-engagement ones was preliminary analysed using violin plots and descriptive statistics. These plots were not reported, to limit the length of the paper. However, from them it was observed that distributions changed shape among levels and that many features presented with promising average differences and different distributions, in particular cvx_Tmean, $cvxTmedian$, $mean\_EDA$, max_der, $cvxP\_range$ and cvx_Pmax. This result has been confirmed by the visualization of the radar plot of the averaged values through all the subjects, comparing the behaviour in the four levels. The radar plot is shown in Figure 2 and it suggests that many of the features have discriminant power between the four levels. cvx_Tmean, $cvxTmedian$ and max_der stand out, while $mean\_EDA$ seems to differ significantly among level 1, 2 and 3 but not between level 3 and 4, finally cvx_Pmean seems to differ significantly among level 2 and 3. Moreover, the correlation matrix was computed. First, to spot redundancy between pairs of features, cross correlation values greater than 0.9 were considered. Between the highly cross-correlated features, only the ones with higher ρ_L where kept in the subsequent analysis. The resulting subgroup of features in shown in Table II with ρ_L from the correlation matrix. The sub-group of the remaining features such that $\rho_L > 0.4$ and the sub-group of features $\rho_L > 0.6$ (in violet in Table II) are created to use them as feature sets with different selection criteria. It is possible to observe from Table II that the trend and values of the EDA amplitude seem to be highly correlated with the game level, especially after the decomposition in tonic and phasic components through the cvxEDA optimisation. It is interesting to notice that the leading features are from 3 out of 4 main types of EDA analysis

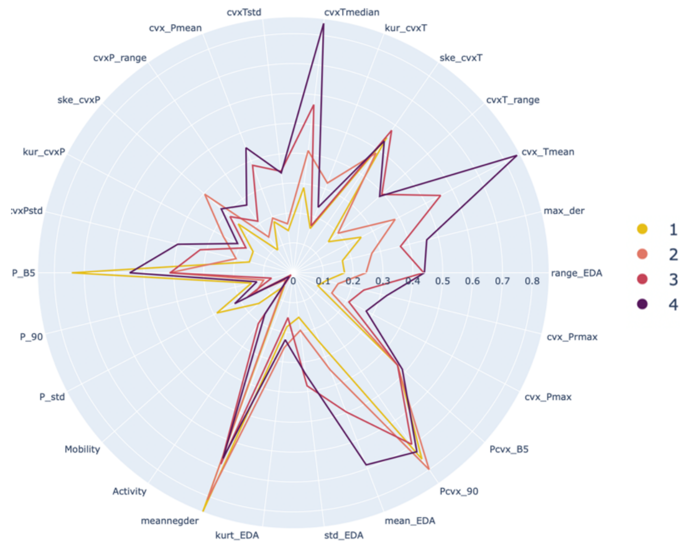

Fig. 2: Radar plot of the averaged features in the four levels.

done in this work. Indeed $mean\_EDA$ comes from the EDA amplitude statistical descriptive features, max_der comes from the EDA-trend analysis, specifically from the first derivative and, at last, $cvxTmean$ is produced by the statistical analysis of the amplitude of the tonic component alone as extracted from cvxEDA. Features from the frequency domain do not seem to be relevant observing radar (or violin) plot and ρ_L, as well as the Hjorth parameters. As a consequence of the previous observation, a 3D scatter plot in the plane $cvxTmean$, max_der and $mean\_EDA$ of all the subjects in all levels has been employed, as shown in Figure 3. The four sets are clearly visible and the centroids well distinguished, except for some outliers. Levels 2 and 3 appear partially superimposed, while levels 1 and 4 are well separated, except for a single outlier subject. The corresponding dots are highlighted with black arrows in Figure 3 and they belong, in fact, to the same subject trial, that actually shows a different EDA response, with descending signal until level 4. The signal was not affected by artefacts due to detachment or arm movement, but the subject encountered problems with audio playback at the beginning of the first level. The game was re-started, therefore it is not easy to explain this behaviour. Indeed, it could be due to a different subject response to the exergame stimulus or, as an alternative, such response could have been modified from the previous setback.

B. Automatic Classification of Engagement

To assess the automatic classification power of EDA between levels of engagement, SVM, RF and k-NN classifiers are exploited using different feature set configurations. The classifiers' accuracy is reported in Table III. All the classifiers perform very well, proving the significance of the extracted features. The SVM and RF show a similar behaviour and both outperform the k-NN in accuracy and stability, especially in the multi-class problem. In the binary problem, an averaged

TABLE II: Sperman's correlation coefficient of the features with respect to LABEL (level of the game = level of engagement). A high correlation coefficient suggests good predictive performance. Values of $|\rho_L| > 0.6$ are shown in violet.

Spearman's correlation coefficient to the level LABEL ρ_L ($\|\rho_L\|$<0.4)									
P_90	kurt_EDA	ske_cvxT	Pcvx_B5	kur_cvxT	Pcvx_90	ske_cvxP	cvxPmedian	P_B5	Mobility
0.008	-0.018	0.032	0.046	0.052	-0.064	0.072	0.111	-0.238	-0.375

Spearman's correlation coefficient to the level LABEL ρ_L ($\|\rho_L\|$>0.4)							
meannegder	cvxT_range	cvxTstd	range_EDA	cvxPstd	max_der	mean_EDA	cvx_Tmean
-0.422	0.466	0.496	0.506	0.514	0.519	0.646	0.900

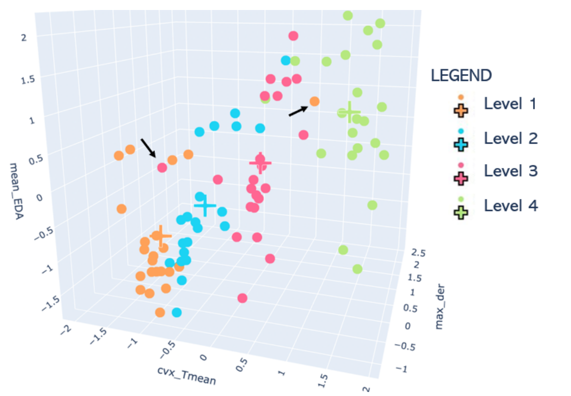

Fig. 3: 3D scatter plot of all the subjects in all levels represented in the plane $cvxTmean$, max_der and $mean\_EDA$. The centroid of each game level set is shown with a cross.

accuracy above 90% is reached across all tested conditions; the classifiers performances are comparable, and RF presents the best accuracy (94.3 ± 5.7%) when taking into account all features. An averaged accuracy above 80% is reached across all tested conditions in the multi-class problem; RF still presents with the best accuracy (93.3 ± 7.4 %) in the all-in feature set. The feature set reduction based on ρ_L ($\rho_L > 0.4$) provokes a sensible reduction in accuracy and higher variability in the SVM and RF classifiers, suggesting that ρ_L alone is not sufficient to describe the significance of the excluded features. The k-NN presents with the opposite trend, i.e., the smaller feature set gives the best accuracy; this is probably due to the fact that the k-NN is affected by the so-called *dimensionality curse*. As for SVM, the loss in accuracy with features reduction is partially restored with the smaller feature set ($\rho_L > 0.6$). This, along with the k-NN behaviour and the overall good accuracy, suggests that the most correlated features to the label ($mean\_EDA$ and cvx_Tmean) are actually the strongest. However, it will be necessary to explore other methods of feature selection to better understand the significance of the remaining features in this classification problem, as they actually bolster the classification, as it was also suggested by the previously shown data such as the radar plot in Figure 2. Looking at the averaged confusion matrix (normalised to report precision) shown in Figure 4, the classification seem to be robust to the noisy features. Indeed, mis-classification

TABLE III: Accuracy of the SVM, RF and k-NN Classifiers in the binary and multi-class problem.

	Multi-class (%)		
	SVM	**RF**	**k-NN**
All-in	90.8± 7.2	93.3 ± 7.4	75.0 ± 11.1
$\rho_L > \mathbf{0.4}$	83.9 ± 12.7	81.8 ± 12.4	71.7 ± 13.2
$\rho_L > \mathbf{0.6}$	86.3 ± 8.6	78.3 ± 7.9	87.5 ± 9.2
	Binary (%)		
	SVM	**RF**	**k-NN**
All-in	93.2 ± 5.6	94.3 ± 5.7	91.9 ± 7.4
$\rho_L > \mathbf{0.4}$	91.9 ± 7.3	93.3 ± 7.4	91.9 ± 10.4
$\rho_L > \mathbf{0.6}$	93.3 ± 5.4	89.9 ± 6.0	93.9 ± 5.7

only occurs between neighbouring classes, and the level 4, that is strongly demanding, is well identified. The trend is similar also when observing the confusion matrices obtained with the other two feature sets. The precision in the level 4 recognition increases and has very high values: the errors seem to occur more often between level 1 and 2, while level 3 and 4 are very well separated. This seems to confirm that level 1 and 2 have a low level of engagement and level 4 a very high one, as expected, and that the EDA analysis is suitable to distinguish this kind of response. Indeed, the binary classification performances are higher in all tested conditions, confirming that, even though there is a partial superposition between level 2 and 3, the two groups (low-engagement and high-engagement) appear distinguished.

Fig. 4: Confusion Matrix of RF in the multi-class problem, feature set: all-in.

IV. Conclusions and Future Work

This work aimed to explore the use of EDA in the evaluation of the engagement provided by the execution of an exergame consisting of several levels. Indeed, emotional involvement during the performance of an exercise can bring great learning and motivation benefits, which are important in a medical protocol such as a rehabilitation one. In this preliminary work, the level of engagement is identified with the level of the game, which was specifically designed for progressive difficulty and mental/physical demand. From the qualitative statistical analysis and the automatic classification results obtained, the features extracted from the EDA seem to be able to describe the trend of engagement in the various levels. The results show that the strongest features were cvx_Tmean and $mean\_EDA$. However, all the features resulted to be relevant in providing the best accuracy for RF and SVM classifiers. Therefore, a deeper analysis of features selection methods should be explored in the future. The employed RF classifier achieved the best performance, with a maximum accuracy of 94.3 ± 5.7 % in the binary problem and 93.3 ± 7.4 % in the multi-class one. In general, the classifiers did not seem to show difficulty in correctly distinguishing distant classes nor discriminating between the four game levels, thus demonstrating a great sensitivity to the engagement degree. These preliminary results should be enriched considering the many factors that could intervene in this setup. In particular, the number of subjects should be enlarged to verify the robustness of this approach, involving pathological subjects who could provide different indications of relevant clinical interest. Moreover, the complexity of the emotional state and its influence on the EDA response should be disentagled, exploiting for example other approaches, like the use of additional physiological signals of interest. Lastly, the presented setup could be easily translated in a real time approach, using a suitable EDA sensor and data log. However, it is very encouraging that the classification was possible with such good results in this EDA protocol, as normally this signal is only used to assess the response to a single, often standardised, stimulus. In this case, instead, it was measured non-invasively during the execution of an exergame that presented with several simultaneous stimuli (e.g., visual, cognitive, auditory, stressors), similarly to what would happen in a real-world application. In fact, the interaction with the technology of a Telemedicine application generally involves psychological and multi-sensory stimuli, hence the good results obtained in this study bode well for the adoption of this approach in such field.

REFERENCES

[1] J. Chen *et al.*, "Pain and stress detection using wearable sensors and devices-a review," *Sensors (Basel)*, vol. 21, no. 4, p. 1030, 2021.

[2] A. Greco, G. Valenza, L. Citi, and E. P. Scilingo, "Arousal and valence recognition of affective sounds based on electrodermal activity," *IEEE Sens. J.*, vol. 17, no. 3, pp. 716–725, 2017.

[3] Y. Liu and S. Du, "Psychological stress level detection based on electrodermal activity," *Behav. Brain Res.*, vol. 341, pp. 50–53, 2018.

[4] K. S. McNeal *et al.*, "Biosensors show promise as a measure of student engagement in a large introductory biology course," *CBE Life Sci. Educ.*, vol. 19, no. 4, p. ar50, 2020.

[5] V. Sharma *et al.*, "Audio-video emotional response mapping based upon electrodermal activity," *Biomed. Signal Process. Control*, vol. 47, pp. 324–333, 2019.

[6] A. Dzedzickis *et al.*, "Human emotion recognition: Review of sensors and methods," *Sensors (Basel)*, vol. 20, no. 3, p. 592, 2020.

[7] E. N. Wiebe *et al.*, "Measuring engagement in video game-based environments: Investigation of the user engagement scale," *Comput. Human Behav.*, vol. 32, pp. 123–132, 2014.

[8] Grandviewresearch, "Europe digital health market size & share report, 2030." https://www.grandviewresearch.com/industry-analysis/europe-digital-health-market-report. Accessed: 2022-8-8.

[9] D. M. Brennan *et al.*, "Telerehabilitation: enabling the remote delivery of healthcare, rehabilitation, and self management," *Stud. Health Technol. Inform.*, vol. 145, pp. 231–248, 2009.

[10] A. Peretti *et al.*, "Telerehabilitation: Review of the state-of-the-art and areas of application," *JMIR Rehabil. Assist. Technol.*, vol. 4, no. 2, 2017.

[11] G. Barry, B. Galna, and L. Rochester, "The role of exergaming in parkinson's disease rehabilitation: a systematic review of the evidence," *J. Neuroeng. Rehabil.*, vol. 11, no. 1, p. 33, 2014.

[12] C. Ferraris *et al.*, "Telerehabilitation of cognitive, motor and sleep disorders in neurological pathologies: the rehome project," in *2022 IEEE Symposium on Computers and Communications*, 2022.

[13] M. Morando, S. Ponte, E. Ferrara, and S. Dellepiane, "Definition of motion and biophysical indicators for home-based rehabilitation through serious games," *Information (Basel)*, vol. 9, no. 5, p. 105, 2018.

[14] M. Pirovano *et al.*, "Exergaming and rehabilitation: A methodology for the design of effective and safe therapeutic exergames," *Entertain. Comput.*, vol. 14, pp. 55–65, 2016.

[15] M. Trombetta *et al.*, "Motion rehab AVE 3d: A VR-based exergame for post-stroke rehabilitation," *Comput. Methods Programs Biomed.*, vol. 151, pp. 15–20, 2017.

[16] A. J. E. de Veer *et al.*, "Determinants of the intention to use e-health by community dwelling older people," *BMC Health Serv. Res.*, vol. 15, no. 1, p. 103, 2015.

[17] P. E. Antoniou *et al.*, "Biosensor real-time affective analytics in virtual and mixed reality medical education serious games: Cohort study," *JMIR Serious Games*, vol. 8, no. 3, p. e17823, 2020.

[18] F. Bruns and F. Wallhoff, "Estimating workload from hearth rate and game precision in exergames," in *Proceedings of the 10th IEEE International Conference on Serious Games and Applications for Health*, 2022.

[19] S. Lee, W. Kim, T. Park, and W. Peng, "The psychological effects of playing exergames: A systematic review," *Cyberpsychol. Behav. Soc. Netw.*, vol. 20, no. 9, pp. 513–532, 2017.

[20] R. Mellecker *et al.*, "Disentangling fun and enjoyment in exergames using an expanded design, play, experience framework: A narrative review," *Games Health*, vol. 2, no. 3, pp. 142–149, 2013.

[21] M. Cardona *et al.*, "Modulation of physiological responses and activity levels during exergame experiences," in *8th International Conference on Games and Virtual Worlds for Serious Applications*, IEEE, 2016.

[22] G.Amprimo *et al.*, "Gmh-d: Combining google mediapipe and rgb-depth cameras for hand motor skills remote assessment.," in *IEEE International Conference on Digital Health (ICDH)*, 2022.

[23] Mediapipe, "Hands." https://google.github.io/mediapipe/solutions/hands.html.

[24] S. Hébert, R. Béland, O. Dionne-Fournelle, M. Crête, and S. J. Lupien, "Physiological stress response to video-game playing: the contribution of built-in music," *Life Sciences*, vol. 76, no. 20, pp. 2371–2380, 2005.

[25] J. R. Stroop, "Studies of interference in serial verbal reactions," *J. Exp. Psychol. Gen.*, vol. 121, no. 1, pp. 15–23, 1992.

[26] E. J. Lyons, "Cultivating engagement and enjoyment in exergames using feedback, challenge, and rewards," *Games Health*, vol. 4, no. 1, pp. 12–18, 2015.

[27] F. Pietroni *et al.*, "Identification of users' well-being related to external stimuli: A preliminary investigation," in *Lecture Notes in Electrical Engineering*, pp. 579–590, Cham: Springer International Publishing, 2019.

[28] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalogr. Clin. Neurophysiol.*, vol. 29, no. 3, pp. 306–310, 1970.

[29] J. Shukla, M. Barreda-Angeles, J. Oliver, G. C. Nandi, and D. Puig, "Feature extraction and selection for emotion recognition from electrodermal activity," *IEEE Trans. Affect. Comput.*, vol. 12, no. 4, pp. 857–869, 2021.

[30] M. Benedek and C. Kaernbach, "A continuous measure of phasic electrodermal activity," *J. Neurosci. Methods*, vol. 190, no. 1, pp. 80–91, 2010.

[31] S. Taylor *et al.*, "Automatic identification of artifacts in electrodermal activity data," *Annu Int Conf IEEE Eng Med Biol Soc*, pp. 1934–1937, 2015.

[32] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi, "CvxEDA: A convex optimization approach to electrodermal activity processing," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, pp. 797–804, 2016.

[33] Jamovi, "Jamovi - stats. open. now." http://www.jamovi.org. Accessed: 2022-8-6.