electronics

Article

Analysis of Physiological Signals for Stress Recognition with Different Car Handling Setups

Pamela Zontone [1,†], Antonio Affanni [1,†], Riccardo Bernardini [1,†], Leonida Del Linz [2,†], Alessandro Piras [1,†] and Roberto Rinaldo [1,*,†]

[1] Polytechnic Department of Engineering and Architecture, University of Udine, Via Delle Scienze 206, 33100 Udine, Italy; pamela.zontone@uniud.it (P.Z.); antonio.affanni@uniud.it (A.A.); riccardo.bernardini@uniud.it (R.B.); piras.alessandro@spes.uniud.it (A.P.)
[2] VI-Grade S.r.l., Via G. Galilei 42, 33010 Tavagnacco, Italy; diego.minen@vi-grade.com
* Correspondence: roberto.rinaldo@uniud.it
† These authors contributed equally to this work.

Abstract: When designing a car, the vehicle dynamics and handling are important aspects, as they can satisfy a purpose in professional racing, as well as contributing to driving pleasure and safety, real and perceived, in regular drivers. In this paper, we focus on the assessment of the emotional response in drivers while they are driving on a track with different car handling setups. The experiments were performed using a dynamic professional simulator prearranged with different car setups. We recorded various physiological signals, allowing us to analyze the response of the drivers and analyze which car setup is more influential in terms of stress arising in the subjects. We logged two skin potential responses (SPRs), the electrocardiogram (ECG) signal, and eye tracking information. In the experiments, three car setups were used (neutral, understeering, and oversteering). To evaluate how these affect the drivers, we analyzed their physiological signals using two statistical tests (*t*-test and Wilcoxon test) and various machine learning (ML) algorithms. The results of the Wilcoxon test show that SPR signals provide higher statistical significance when evaluating stress among different drivers, compared to the ECG and eye tracking signals. As for the ML classifiers, we count the number of positive or "stress" labels of 15 s SPR time intervals for each subject and each particular car setup. With the support vector machine classifier, the mean value of the number of positive labels for the four subjects is equal to 13.13% for the base setup, 44.16% for the oversteering setup, and 39.60% for the understeering setup. In the end, our findings show that the base car setup appears to be the least stressful, and that our system enables us to effectively recognize stress while the subjects are driving in the different car configurations.

Keywords: stress recognition in drivers; electrodermal activity; heart rate variability; motion artifact removal; machine learning

Citation: Zontone, P.; Affanni, A.; Bernardini, R.; Del Linz, L.; Piras, A.; Rinaldo, R. Analysis of Physiological Signals for Stress Recognition with Different Car Handling Setups. *Electronics* **2022**, *11*, 888. https://doi.org/10.3390/electronics11060888

Academic Editors: Dorota Kamińska, Gholamreza Anbarjafari, Frane Urem, Rui Raposo and Mário Mário Vairinhos

Received: 10 January 2022
Accepted: 9 March 2022
Published: 11 March 2022

1. Introduction

Well-being assessment and quantification in domestic or in more general scenarios are research areas which are receiving increasing attention [1]. In particular, vehicle drivers' well-being is of paramount importance, as anxiety and stress, for instance, are linked to bad driving behavior [2]. Other factors, such as drowsiness, are linked to car accidents [3]. Furthermore, the effects of frequent stressing trips can cause long-term health problems and increase the risk of cardiovascular [4] and locomotor [5] diseases. Stress, among other parameters, should be therefore monitored in drivers, and interest in this matter is growing, as also demonstrated by the inclusion of more advanced driver assistance systems in many cars in recent years. Physiological signals can be a source of information about stress in this scenario [6–8] and are widely employed in drivers' mental state research (see also [9]).

In the literature, many techniques for automatic stress recognition based on the analysis of physiological signals have been presented. Additionally, in recent years, machine

learning (ML) techniques [10,11] have been applied to the analysis and classification of these signals, given their ability to identify, after training, non-obvious patterns in the characteristics of the signals themselves. The electrocardiogram (ECG) is one of the most used physiological signals in stress detection systems based on ML algorithms [12–16], while the electroencephalogram (EEG) signal is also increasingly common [17–20]. Moreover, authors in [21,22] found a correlation between electrodermal activity (EDA) [23] and stress in drivers. Considering some of these studies, authors in [15] proposed a deep learning framework, called Deep ECGNet, which allows the recognition of two emotional conditions, i.e., stressful and resting, through the analysis of short-term ECG signals. Stress was induced in subjects according to various experimental protocols, which include different tests to be carried out, such as the Stroop Color–Word test or the Montreal Imaging Stress Task (MIST). The performance results show that Deep ECGNet outperforms conventional ML approaches, achieving an accuracy of 87.39% (in a first experiment with 13 subjects) and 73.96% (in a second experiment with 9 subjects). Various ML classifiers, including an SVM, a Naive Bayes (NB), and a k-nearest neighbors (k-NN) classifier, were instead employed in [17] to classify stress using features extracted from the EEG signal. A stress recognition accuracy up to 85.20% was obtained using the SVM classifier.

With an emphasis on driving scenarios, the performance results of SVMs, decision trees (DTs), NB and general Bayesian classifiers were presented in [21]. One single subject was tested, and different signals were acquired during an experiment carried out in real driving conditions. As for the stress evaluation, a scale of two stress levels (i.e., normal and stress) was used. Considering only the features extracted from the physiological signals, with time windows of five minutes, an accuracy equal to 78% was attained with the SVM classifier. By incorporating additional features from a video recording of the driver's face and environment parameters, the SVM accuracy increased up to 86%. The same physiological signals were considered in [22], but analyzing a 10 s window at a time. An NB classifier was used, achieving a stress event detection accuracy equal to 82% by considering only physiological signals logged in real driving conditions. The inclusion of additional data, such as information from the current driving environment and vehicle data, as well as the driving behavior history, allows the system to achieve a higher stress event detection accuracy (96%). In [24], an SVM-based method was proposed to monitor the conditions of a driver, using several physiological data, such as EDA, photoplethysmography (PPG), skin temperature, and wrist movements collected by a wearable glove. Experiments were carried out using an indoor driving simulator, and 28 subjects were tested. A classification accuracy of 98.43% was obtained considering four classes (i.e., stress, fatigue, drowsiness, and normal) and applying a k-fold (with k = 5) cross-validation procedure on the data. In a different test denoted as SIT (subject-independent test), which uses a leave-one-subject-out cross-validation procedure, accuracy values of 68.31% and of 84.46% were obtained with four and three classes (normal, stressed, and fatigued or drowsy states), respectively. EEG was employed in [25], where three ML algorithms were compared and analyzed to recognize stress and fatigue based on EEG patterns. Among the different classifiers employed in the study, the SVM model performed better than the others in distinguishing between rest and stress states (with a performance accuracy of 97.95%). More advanced and recently developed techniques proposed in the literature, using EEG data, involve back propagation neural networks [26] and hierarchical neural networks [19].

An architecture that allows the analysis of a subject's stress while driving in a simulated autonomous and manual driving scenario, using a professional dynamic simulator, was proposed in [27]. The results obtained by evaluating the number of labels generated by an SVM model, which performs well in identifying stress situations, show that in general, the subjects seem to be more stressed when driving in a manual scenario. The effect of traffic conditions in subjects while they are driving in a simulated urban scenario is instead described in [28]. The goal was to verify the applicability of various ML approaches, based on the analysis of EDA and ECG signals, in order to identify stress in two different scenarios characterized by the existence, or not, of traffic.

In this paper, we assess the performance of a detection system which identifies the stress level of subjects engaged in driving, using a dynamic simulator, and we compare the results of various ML algorithms extending our previous work [29]. In that work, a single ML algorithm was considered, and a preliminary analysis of the physiological signals logged from the subjects was introduced. Here, we propose the application of a machine learning approach as a support to the comparison and validation of different car setups. The tests were carried out in a specialized firm by professional car test drivers, who drive prototypes and cars before their production and evaluate their performance, safety and comfort. The small number of subjects engaged in the tests is dictated by the common procedures adopted in the firm for such tests, involving a limited number of car test drivers. The proposed scheme, although conceptually similar to others in the literature where the signals are analyzed with ML techniques, is characterized by innovative elements, in particular, a practical procedure for the attenuation of motion artifacts, based on merging the information from two EDA sensors, and the use of the EDA SPR signal, which has some advantages over the more common SCR signal. The main objective of the paper, however, is to verify whether the different setups of a car prototype can influence the driver's emotional response, and how the subjects respond to these different car handling settings. This can provide further insight for evaluating the good or bad qualities of a particular car configuration, including the degree of acceptance in the general public. An objective support in this context is of particular interest for car manufacturers.

The paper is structured as follows. The next section presents an overview of the general blocks of the proposed system. We introduce the sensor we use for SPR and ECG logging, and the eye tracking device. We then briefly describe the MA removal algorithm that we designed to reduce the motion artifacts. In this section, the experimental setup is also presented. Section 3 summarizes the results of our study, taking into consideration all of the signals acquired from the subjects. A detailed SPR signal analysis, using the *t*-test, the Wilcoxon test, and the different ML models, is presented in this section. Section 4 discusses the results and the contributions of the paper, also outlining some of its possible limitations. In Section 5, some conclusions are drawn.

2. Materials and Methods

The experimental procedure used to evaluate the emotional response in drivers is illustrated in Figure 1. We logged different signals from the test subjects, i.e., the EDA SPR signals from the subjects' hands, the ECG from the chest, and the eye movements. We used a dynamic professional simulator which realistically reproduces car movements, considering three different car settings which change the car handling (neutral or "Base", oversteering or OS, understeering or US). The subject's stress reaction to these different car settings, as measured by the acquired signals, is finally evaluated. This gives us the opportunity to assess whether a particular car setting appears to be more comfortable or more stressful than others. In the following sections, we describe each block of Figure 1 in detail.

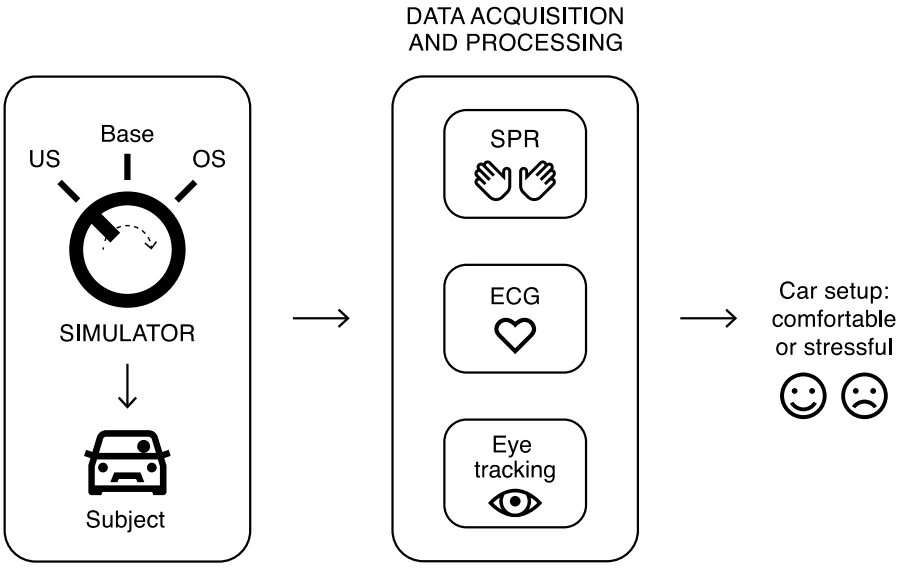

Figure 1. Block diagram of the proposed system.

2.1. The Sensors

Each subject in the test wore the VI-BioTelemetry sensor (produced by the VI-grade company [30]). This sensor was developed by the authors and is similar to the ones presented in [27,31]. It acquires three ECG channels by wearing a vest with wet electrodes posed on the chest, and two SPR channels, one from each hand. The SPR reading is performed posing two Ag/AgCl electrodes on the palm and on the back of each hand. The bio-signals (i.e., the three ECG channels and two SPR channels) are properly conditioned by the VI-BioTelemetry sensor through an analog front end. In particular, the input range for the ECG channels is ± 5 mV with a bandwidth in the range [0.03, 160] Hz, while the SPR channels input range is ± 20 mV with a bandwidth in the range [0.08, 40] Hz. The input impedance of each channel (either ECG and SPR) is 100 MΩ. The sensor, after analog signal conditioning, converts the data into digital information using a 12 bit A/D converter on a DSP board with sampling rate 1 kSa/s, and sends the data to a server using a Wi-Fi module. The accuracy of the sensor was characterized in [32], resulting in being 40 µV (corresponding to 0.1% of full scale) for the SPR channels and 3 µV (corresponding to 0.03% of full scale) for the ECG channels. The resolution is 10 µV for the SPR channels and 2.4 µV for the ECG channels. In addition to the bio-signals, we also recorded the eye movements and the pupil diameter of the driver using the eye tracking system called Smart Eye Aurora [33].

2.2. The EDA SPR Signal

When a person undergoes stress-inducing events, the time and amplitude of the stimuli arising from the autonomic nervous system (ANS) in response to these events can be estimated using EDA [34]. In detail, EDA measures the electrical characteristics of the skin. It reflects the sympathetic nervous system activity, and it is used for physiological measurements due to its correlation with the activity of the ANS [35]. EDA is measured from the eccrine glands, which are mainly concentrated in the palmar and plantar dermatomes, which are the best sites for measuring it [23,36]. There are mainly two types of EDA recordings, namely, endosomatic and exosomatic measurements. In endosomatic measurements, the electric potential differences originating in the skin itself are recorded. This technique refers to the measurement of the skin potential (SP) signal. In exosomatic measurements, a small amount of current is applied to the skin, and the resulting variations in skin conductance (SC), or related electrical properties, are recorded. When the ANS stimulates sweat production, the EDA signals change as a result of sweat secretion and alterations in ionic permeability of sweat gland membranes. EDA signals are typically the combination of two components, namely, the tonic (level—L) component and the phasic

(response—R) component. In particular, the corresponding SCL or SPL signals refer to the slow-varying baseline level of the SC and SP signals, respectively. Phasic EDA is specified by a fast-varying component, known as the skin conductance response (SCR) and skin potential response (SPR). Phasic signals are the most useful to recognize reactions to stress-evoking events, resulting in peculiar signal patterns and increased signal activity. In this work, we propose to use the SPR to detect the stress level of car driver testers in different car setups. We use SPR instead of SCR because SPR can be measured without the need of electrical currents being applied to the body, and because, as reported in the literature, it is less sensitive to the impedance of electrodes and to slow variations of skin impedance [37]. In addition, SPR typically shows a faster reaction to stress stimuli compared to SCR [23,38]. A detailed comparison of different EDA signals, including SCR and SPR, is reported in [38].

2.3. Motion Artifact Removal

The two SPR channels are the inputs of the motion artifact (MA) removal algorithm that we also described thoroughly in [39,40]. This algorithm is based on two assumptions: (1) the artifacts due to motion (e.g., due to the steering wheel action) increase the local energy of the SPR signal, and (2) the motion artifacts rarely appear simultaneously with the same behavior in both hands, thus altering both SPR signals in the same way. As a matter of fact, the second assumption is based on the evidence that the effort due to the rotation of the steering wheel mainly engages one of the two hands. Considering these two assumptions, the algorithm computes the local energy of each SPR signal, and the output is obtained as a weighted combination of the two input signals, giving more weight to the one with the lowest local energy. More specifically, we compute the RMS values σ_1 and σ_2 of the two input signals SPR_1 and SPR_2 on a 1 s moving window (composed of $N = 100$ samples due to the sample rate equal to 100 Sa/s). This way, at the i^{th} sample (with $i > N$), $\sigma_{1,2}$ are determined as follows:

$$\sigma_{1,2}(i) = \sqrt{\frac{\sum_{j=i-N+1}^{i} \mathrm{SPR}_{1,2}^2(j)}{N}}. \tag{1}$$

We then consider a particular smooth threshold function, defined as

$$h(x) = \frac{1}{1 + e^{-2(x-1)}} \tag{2}$$

so that $h(x) \approx 0$ when $x \ll 1$ and $h(x) \approx 1$ when $x \gg 1$. This function is utilized in an adjustment parameter γ, which allows us to get rid of the motion artifacts. Considering the i^{th} sample, the parameter $\gamma(i)$ is set as

$$\gamma(i) = \begin{cases} h\left(\frac{\sigma_1(i)}{\sigma_2(i)}\right) & \text{if} \quad \sigma_2(i) \neq 0 \\ 1 & \text{if} \quad \sigma_2(i) = 0 \end{cases} \tag{3}$$

Looking at (2) and (3), we can deduce that $\gamma \to 0$ when $\sigma_1 < \sigma_2$ and $\gamma \to 1$ when $\sigma_1 > \sigma_2$. In the end, we attain the output $z(i)$ of the MA removal algorithm by computing a weighted combination of the two input signals through the adjustment parameter γ, i.e., at the i^{th} sample, $z(i)$ is calculated as

$$z(i) = \gamma(i) \cdot \mathrm{SPR}_2(i) + [1 - \gamma(i)] \cdot \mathrm{SPR}_1(i) \tag{4}$$

Because of that, the output of the MA removal algorithm is a new clean signal which follows the smoother SPR signal. As an example, Figure 2 shows how the MA algorithm removes the artifacts due to the steering wheel action for one of the four test subjects during the drive. As it can be seen, the SPR_1 signal presents several artifacts due to the motion of the hands (represented by sudden and strong spikes), while the SPR_2 signal results to be less affected by these spikes. The MA algorithm (with the output appearing as a solid

black line) weights and combines the two signals, giving more weight to the signal with the lowest local energy.

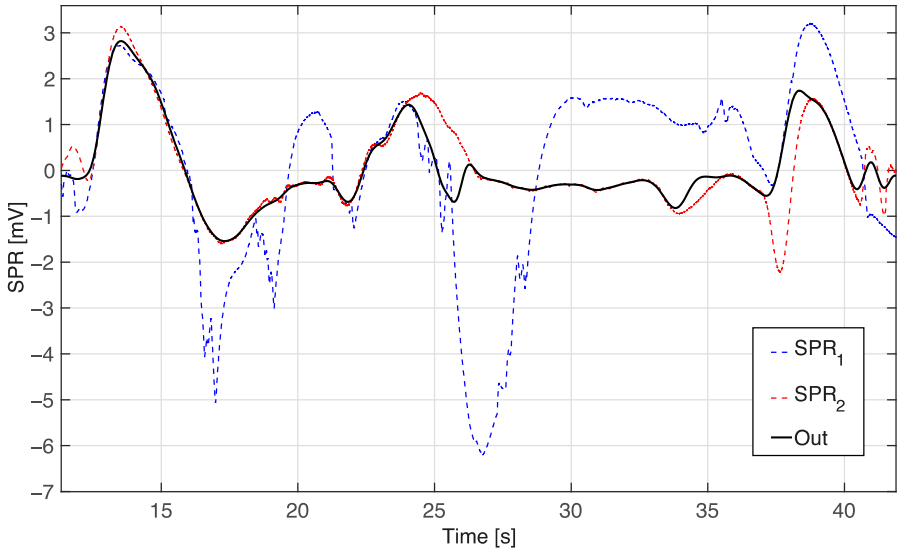

Figure 2. MA removal algorithm application result. Dashed lines: SPR signals from the two hands of a test subject. Continuous line: output of the proposed algorithm.

2.4. Experimental Setup and Stress Level Classification

We estimated the emotional state of the drivers, simulating the various car setups previously mentioned. Data were recorded with experiments performed at the VI-grade company (vi-grade.com). This company is a leading producer of software and hardware products for advanced applications in the field of car driving simulators. We used a 9-DOF (degrees of freedom) dynamic driving simulator, called DiM150 (see Figure 3), in which we changed the vehicle dynamics in three different ways. In detail, the simulated vehicles were designed, implementing three different setups, using different mass distributions, suspension hardness, aerodynamic coefficients and so on, in order to obtain the first vehicle handling model with emphasized oversteering or "OS", the second vehicle with emphasized understeering or "US", and the third vehicle with a neutral or "Base" setup to mimic a daily-use vehicle. In particular, to create an OS or US car behavior, we acted on the dynamic load transfer of the vehicle through the coil spring settings. Supposing a centered center of mass, setting an equal coil spring stiffness for all the four wheels yields a load transfer during turns of 50% on the front wheels and 50% on the rear ones, with a neutral behavior of the car. If we vary the coil spring stiffness by some amount, e.g., by 5%, as we did in our experiments, for example, choosing harder coils on the rear axle and softer coils on the front axle, we force the load transfer to be 55% on the rear wheels and 45% on the front ones. This imbalanced setup makes the rear wheels lose grip, thus providing an oversteering car behavior. Similarly, providing 45% on the rear wheels and 55% on the front ones, we obtain an understeering behavior. We tested four male professional test drivers, with an age in the range 30–45 years and with several years of driving experience, who commonly drive concept cars or cars under testing to assess their behavior before production. They agreed on the logging of their physiological signals through the sensors described above. The subjects under test filled in an informed consent module before getting on the simulator. The principles of the Declaration of Helsinki were also respected during the tests [41].

Figure 3. The DiM150 VI-grade professional dynamic simulator.

The acquired signals were processed and analyzed using three different ML classification algorithms: a support vector machine (SVM), a random forest (RF) and a k-nearest neighbors (k-NN) classifier. These algorithms are commonly used in the literature for classification purposes. All of these models were trained on a larger dataset, characterized by a significant number of elements, collected from a different experiment that we still performed at the VI-grade company with the same dynamic simulator. In that experiment, we tested 18 subjects—14 male and 4 female—coming from the University of Udine and the University of Padua. The subjects were told to manually drive on a straight road, 67 km long, until they reached the stop road sign. They were also told to maintain a steady velocity between 120 and 130 km/h. The complete course took approximately 40 min. During the drive, they had to overcome 12 obstacles positioned in fixed points along the track, which were added in order to stress the subjects during their driving. The obstacles (see Figure 4) were the following: double lane change (right to left or left to right), tire labyrinth, sponsor block (from left or from right), slalom (from left or from right), lateral wind (from left or from right), jersey LR, tire trap, and stop.

Five statistical features were computed from the single SPR signal that was generated by the MA removal algorithm. The derived features were the block variance, the energy, the mean absolute value, the mean absolute derivative, and the maximum absolute derivative (see also Table 1). To determine these features, we considered each 15 s interval at a time, taking into consideration a new 15 s interval 5 s after the previous one (i.e., 10 s overlap between consecutive intervals). A normalization procedure was also applied to the subjects' SPR signals in order to make the signals of the various subjects comparable. During the tests on the driving simulator, we were aware of the position of the 12 obstacles along the course, so we could assume that the individuals were stressed while they crossed them. Therefore, all of the intervals occurring inside or intersecting the obstacles' span were considered stress intervals (giving them a label equal to "1"), whereas all of the intervals occurring outside the obstacle span were considered non-stress intervals (giving them a label equal to "0"). A total of 3195 samples for each class, stress and non-stress, were therefore used for training. In Section 3.1, we report the accuracy results of the classifiers.

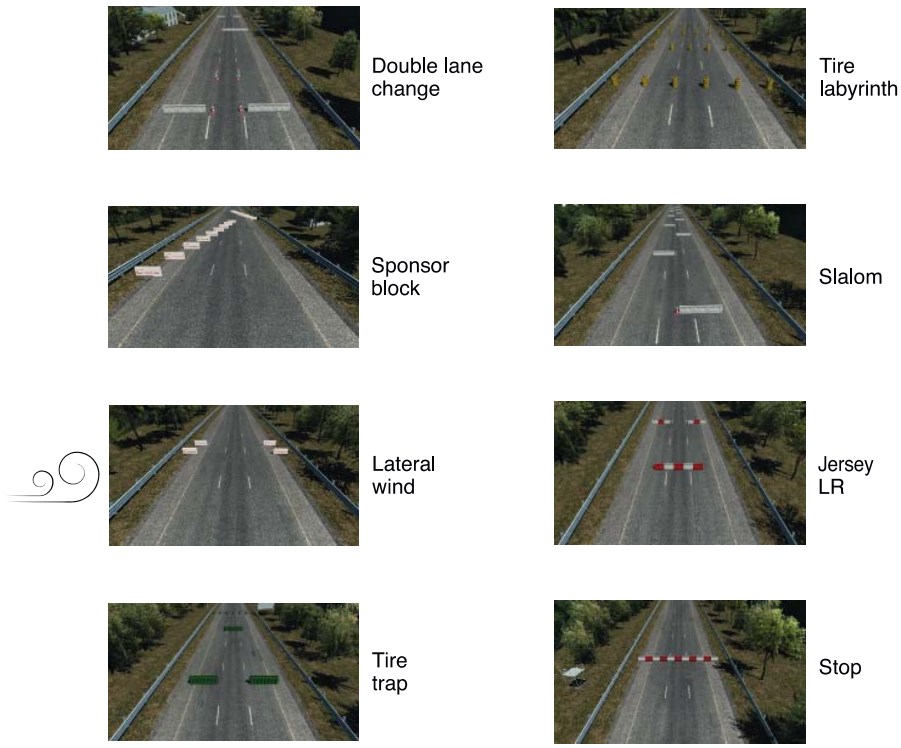

Figure 4. Obstacles used in the experiment to collect data for the ML training procedure [40].

Table 1. Features used in our ML classification algorithms.

Input Signal	Computed Features
SPR	Variance Energy Mean absolute value Mean absolute derivative Max absolute derivative

The models, after being trained on the larger database, were used to classify the SPR signals acquired from the subjects while they were driving with a particular car handling setting (that is, the SPR signals provide a new test set for these models). The four individuals, after entering the simulator's cockpit, had to drive on a straight highway for 15 km. Ten successive double lane changes were prearranged along the course such that the distance between them was about 1500 m. Cones on the road allow the definition of each lane change section, and the complete double lane change maneuver spans over 170 m. This comprises a 30 m entry lane, followed by a 25 m long side lane and a 60 m long exit lane. The entry lane is 3 m wide, whereas the side and exit lanes have a width of 3.2 m. There are offsets, lateral and longitudinal, between the entry and side lane, respectively of 0.8 and 30 m. Side and exit lanes have this identical lateral offset and 85 m of a longitudinal offset. Two road signs are also located at −100 and −200 m before the beginning of the double lane change in order to inform the driver of the incoming task. The driver had to overcome the lane changes, not hitting the cones. Each subject had to repeat the whole drive (15 km long) three times, each time with one of the three different car setups. Figure 5 shows in detail the consecutive phases of our test protocol.

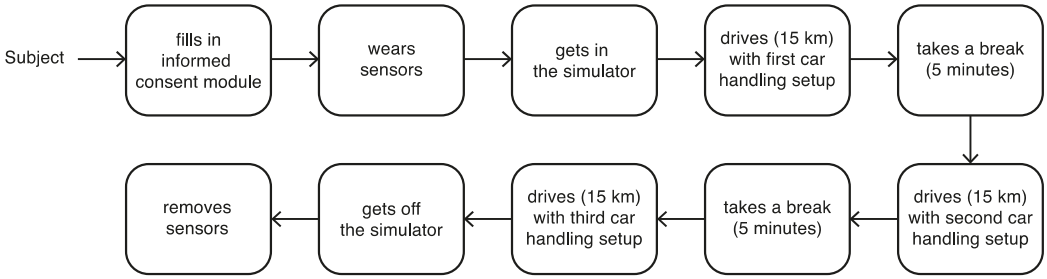

Figure 5. Phases of our test protocol.

3. Results

3.1. SPR Signal Analysis

As a first step to evaluate how a car setup can be more or less stressful than the others, we analyze the single SPR signal at the output of the MA removal block. As previously described, the simulated track for each car setup is composed of ten identical double lane change tasks. For this reason, we divide each track into ten segments so that a task results in being located in the center of each segment. The positions of the tasks are at $d_i = 1500\,i$ m, $i = 1, 2, \ldots, 10$, and we observe the SPR signals in the intervals $[d_i - 750,\ d_i + 750]$ m. We start evaluating the RMS value of the SPR signal in these intervals to take a first look at the characteristics of this signal and the possible differences among the subjects. This provides a first indication about the signal behavior for each driver and each car setup. Figure 6 shows the mean and standard deviation (SD), in mV, of the 10 RMS values, for each subject and each car setup. It can be seen that the base setup is characterized by a smaller RMS value, except for Driver 4, where the US setup results in a slightly smaller RMS value.

Even if the number of calculated values is small (i.e., ten values, corresponding to the track segments), we perform an analysis with a paired t-test and a Wilcoxon test. For each driver, we compare the corresponding three SPR data arrays (Base, OS and US) and evaluate if there is a statistical difference between the three populations, which would mean that different car setups generate different emotional responses on the subjects (see also [42] for a detailed discussion about the applicability of statistical tests to small sample sizes).

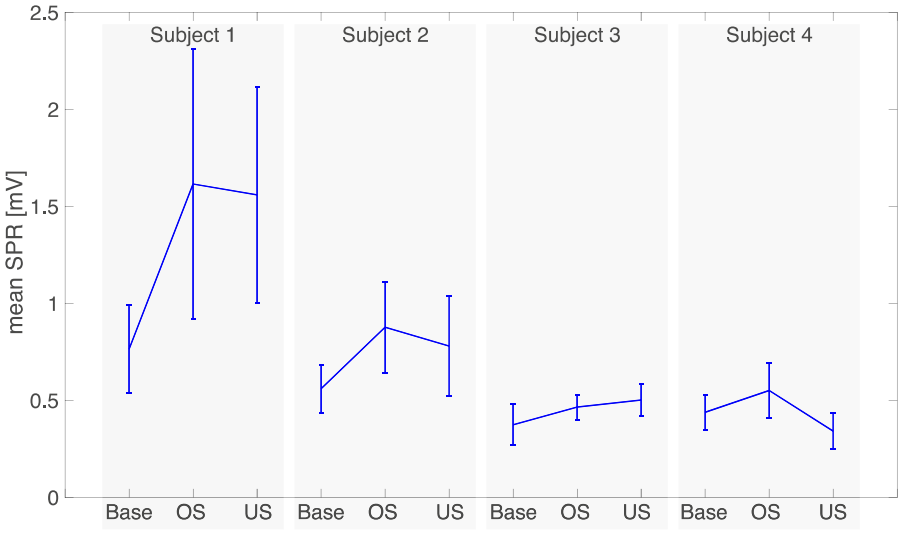

Figure 6. Mean SPR and standard deviation for each driver and each car setup.

As a first indication of the significance of the mean values shown in Figure 6, we performed the paired t-test on the Base–OS, Base–US, and OS–US pairs. The t-test assumes that the data are instances of independent normal random variables, a rough approximation that, however, is in accordance with the ML classifier results that we will present later.

Table 2 shows the p-value resulting from the test. Bold numbers in the table highlight the cases with $p \leq 0.05$, indicating a significant difference between the RMS value relative to the first setup, compared to the second one under test. When $p \leq 0.05$, we also report within parenthesis the power of the test, calculated by considering the sample average and combined standard deviation of the samples, for a significance level $\alpha = 0.05$ (with sample size $n = 10$). In particular, we use Cohen's d formula to measure the effect size of the difference between the means of the two distributions. Note that the power indicates the probability of accepting the hypothesis that the samples come from Gaussian distributions with different means (which we do when $p \leq 0.05$), given that this hypothesis is actually true [43].

Table 2. Paired t-test probability p-value considering Base and OS, Base and US, and US and OS car setups for each subject. Power of the test within parenthesis when $p \leq 0.05$.

Subject	1	2	3	4
Base vs. OS	**0.01 (0.94)**	**0.01 (0.95)**	**0.05 (0.60)**	0.08
Base vs. US	**0.03 (0.98)**	0.06	**0.03 (0.82)**	**0.02 (0.61)**
OS vs. US	0.62	0.23	0.25	**0.004 (0.96)**

Looking at the data in Table 2, we can see that the stress response in the Base setup appears to be less stressful than the OS one (with good statistical significance) by drivers 1, 2 and 3. For driver 4, instead, the Base setup appears to be less stressful than the OS one, but with lower statistical significance ($p = 0.08$). On the other hand, the Base setup appears to be less stressful than the US setup with good significance for drivers 1 and 3. Similar to the previous case, for driver 2, the Base setup appears to be equally less stressful than the US setup, albeit with slightly lower significance ($p = 0.06$). For driver 4, instead, the Base setup appears to be more stressful than the US one. The OS setup appears to be comparable or slightly more stressful than the US one for drivers 1, 2 and 4. However, the statistical significance of the paired t-test is low for three out of four drivers. For subjects 1 and 2, the OS car setup appears to be more stressful than the US one, but from a statistical point of view, they provide equivalent responses ($p = 0.62$ and $p = 0.23$, respectively). For driver 3, the OS setup results to be less stressful than the US one, but from a statistical point of view, they provide equivalent responses ($p = 0.25$). For subject 4, the OS setup results to be more stressful than the US one with good significance.

We also performed a non-parametric test (Wilcoxon signed rank test) between the two data sets (see Table 3). In particular, the Base setup results to be significantly less stressful than the OS one by all the drivers ($p = 0.003$, $p = 0.001$, $p = 0.05$, and $p = 0.04$, respectively). The Base setup results to be significantly less stressful than the US one by drivers 1, 2, and 3 ($p = 0.001$, $p = 0.037$, and $p = 0.01$, respectively) but significantly more stressful for driver 4 ($p = 0.01$). As in the paired t-test results, the OS setup results to be more stressful than the US one with poor statistical significance, with the exception of driver 4 ($p = 0.001$). In Table 3, we also show the power of the test for a significance level $\alpha = 0.05$ (with sample size $n = 10$). To compute it, we assume that the samples come from Gaussian distributions and, as before, we used Cohen's d as a measure of the effect size of the difference between the means of the two distributions.

Table 3. Paired Wilcoxon test probability p considering Base and OS, Base and US, and US and OS car setups, for each subject. Power of the test within parenthesis when $p \leq 0.05$.

Subject	1	2	3	4
Base vs. OS	**0.003 (0.99)**	**0.001 (0.99)**	**0.05 (0.82)**	**0.04 (0.73)**
Base vs. US	**0.001 (0.99)**	**0.04 (0.84)**	**0.01 (0.96)**	**0.01 (0.82)**
OS vs. US	0.96	0.42	0.34	**0.001 (0.99)**

To sum up, analyzing the sympathetic activity observed through the acquired SPR signals, we can conclude that the Base setup is significantly less stressful, as measured by the SPR signal activity, than the others for the majority of drivers (with the exception of driver 4), while the OS setup appears to be the most stressful one. As mentioned, the limited number of car test drivers was dictated by the common procedures adopted in the company for the tests. For each test subject, we were also able to extract a given number of intervals to be processed through the ML algorithms (as we will see soon). The results of both these analysis, considering the SPR signal, are consistent with each other.

In addition to the previous evaluations, we observe a high variability of the SPR values for driver 1 (see Figure 6). Looking at the SPR data along the entire track, we can notice that there is an increasing engagement of this driver to the task. In Figure 7, we can see that the SPR activity of driver 1 increases along the track (the vertical lines in Figure 7 denote the task positions). More specifically, starting from the fifth lane change, the SPR activity increases task after task. This behavior does not appear in the other subjects. We include in Figure 8 a plot of the SPR for another subject, where the SPR signal response is flatter and does not present this increase in amplitude.

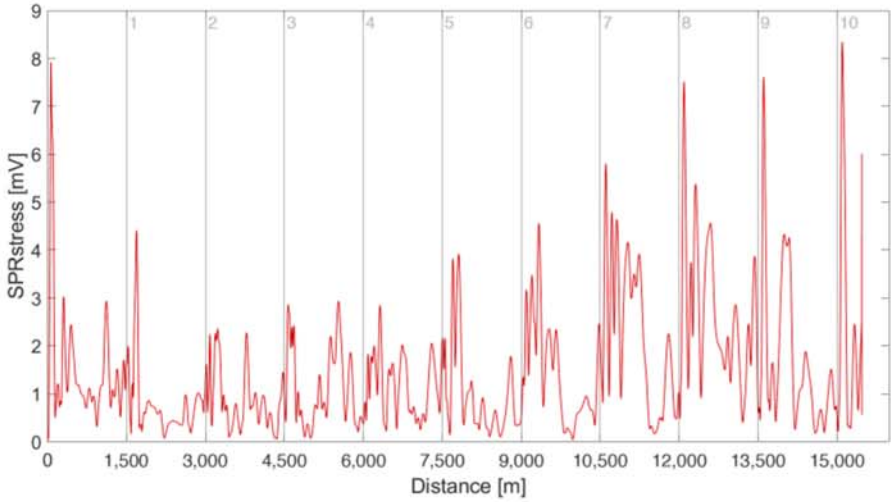

Figure 7. SPR signal for subject 1 (local RMS value).

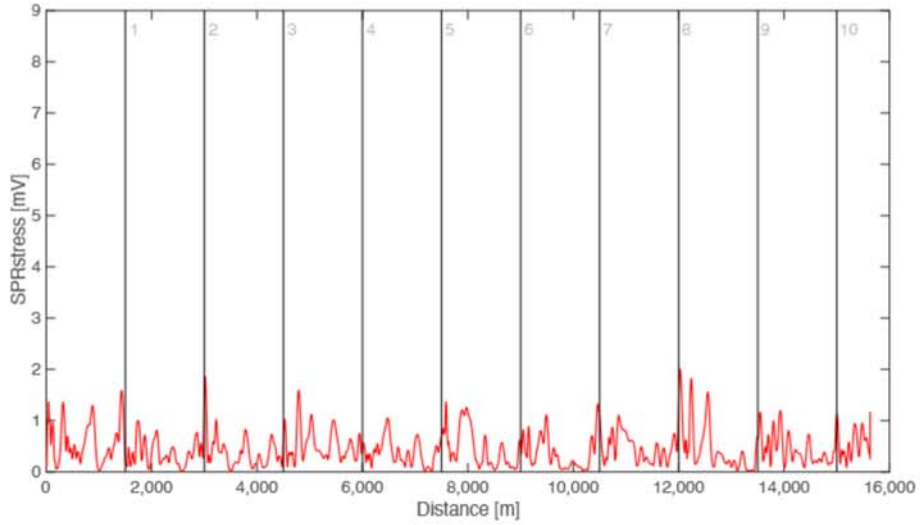

Figure 8. SPR signal for subject 3 (local RMS value).

As introduced above, for the classification task, we compared the performance results of an SVM, an RF, and a k-NN classifier on a larger dataset with 18 subjects. Their performance resulted in being similar in terms of accuracy (approximately 73%), with

just the k-NN classifier showing a slightly lower accuracy value (68%). To compute these accuracy values, we used an approach which is a nested procedure, where the outer loop excludes one subject at a time (the leave-one-subject-out step), while in the inner loop we performed a 10-fold cross validation step. We then computed the mean of the test results for all individuals, and we applied a relabel procedure to overcome the issue of single isolated "1" labels along the course [44]. All of these ML models were developed in Matlab (2017a), also using the Bayesian optimization during the training phase for hyperparameter optimization (see also [45]). The classifiers' hyperparameters are obtained by selecting the auto hyperparameter optimization of the Matlab Statistics and Machine Learning Toolbox [46]. In particular, a radial basis function (RBF) kernel was used for the SVM classifier, and the hyperparameters are the C and γ values. The hyperparameters of the RF and k-NN classifiers are selected by the auto hyperparameter optimization procedure, as well.

The classification models are used to classify the SPR signals recorded from the four subjects while driving on a car characterized by a particular setup. For each subject, we standardize the corresponding SPR signal (after the MA removal block) by subtracting its mean and dividing it by the maximum standard deviation among the three values that we obtained, considering the three different configurations (US, OS, and Base). In this way, we are able to obtain comparable signals for evaluating the impact of the different car handling setups on each subject, making the comparison of different subjects meaningful. As was done for training, the SPR signals are divided into 15 s intervals, extracting the same five features described above. Running the various ML classifiers only as a test gives us the chance to count the output labels equal to "1" or "0", that is, stress or non-stress intervals, for each driver and each different car setup. In this way, we obtain evidence of the possible stress states arising in the subjects while completing the various runs.

Figures 9–11 show the percentage of the positive or "stress" labels for all the classifiers, i.e., the ratio between the number of positive labels and the total labels obtained considering the whole course with that particular car setup. The results are in accordance with those obtained by analyzing only the RMS of the SPR signals. In particular, the number of positive labels considering the Base car setup results to be generally smaller than the US and OS ones. We can thus state that this setup appears to be less stressful than the others. The OS setup, instead, results in being the most stressful setup (only excluding subject 3). For clarity, we also report in Table 4 the exact values of the number of positive labels (in %) represented in Figures 9–11, for the different classifiers.

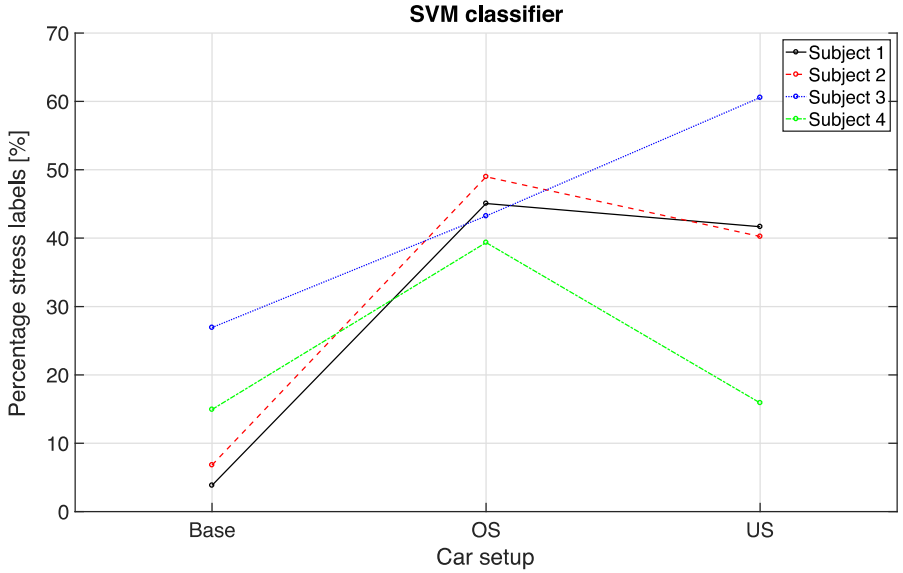

Figure 9. Percentage of labels labeled as "stress" by the SVM classifier, for the four subjects in the different car setups.

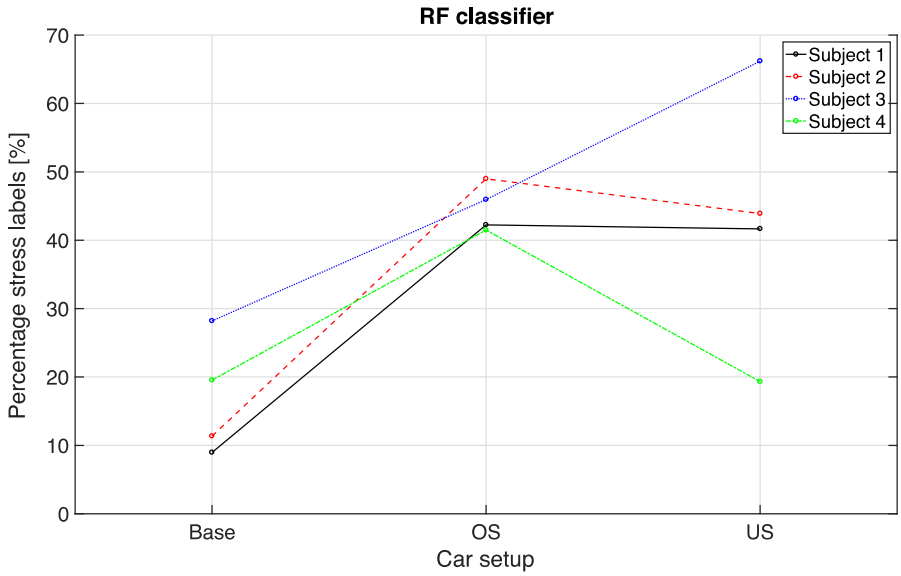

Figure 10. Percentage of labels labeled as "stress" by the RF classifier, for the four subjects in the different car setups.

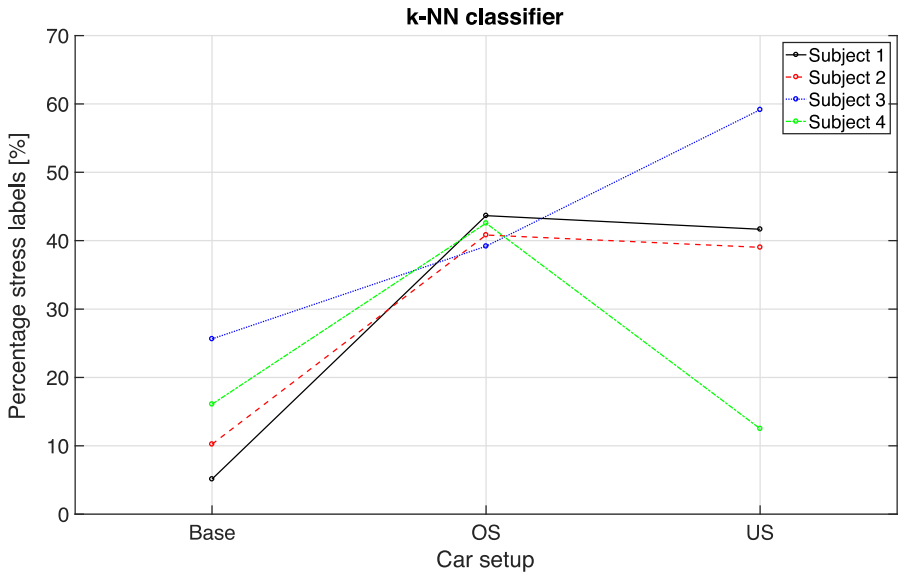

Figure 11. Percentage of labels labeled as "stress" by the k-NN classifier, for the four subjects in the different car setups.

Table 4. Percentage of intervals labeled as "stress" by the different classifiers, for each subject and each car setup.

SVM: % Positive Labels				**RF: % Positive Labels**				**k-NN: % Positive Labels**			
Subj	Base	OS	US	Subj	Base	OS	US	Subj	Base	OS	US
1	3.85	45.07	41.67	1	8.97	42.25	41.67	1	5.13	43.66	41.67
2	6.82	48.98	40.24	2	11.36	48.98	43.90	2	10.23	40.82	39.02
3	26.92	43.24	60.56	3	28.21	45.95	66.20	3	25.64	39.19	59.15
4	14.94	39.36	15.91	4	19.54	41.49	19.32	4	16.09	42.55	12.50

Figure 12 shows the output of the SVM classifier for subject 2, where we can notice that the number of positive labels in the OS car setup is greater than the Base and US ones. For clarity, we only plot the labels equal to "1" using a grey stem, added at the end of the corresponding 15 s interval. We do not include the labels equal to "0", which

correspond to the non-stress case. The SPR signals in the three different car setups, after being cleaned from the MA and then normalized, are also plotted in a blue continuous line. Figure 13 instead displays the output of the RF classifier for subject 3. In this case the number of the positive labels always increases, going from the Base setup, through the OS setup, to the US one. Finally, Figure 14 shows the output of the k-NN classifier for subject 4 (also in this figure, each positive label is placed at the end of each corresponding 15 s interval). This is the subject where the OS setup appears remarkably more stressful than the other two setups, no matter the model used for classification. As we can see, all of the classifiers well recognize the emotional responses arising in the individuals while driving on a professional simulator, characterized by different car handling settings.

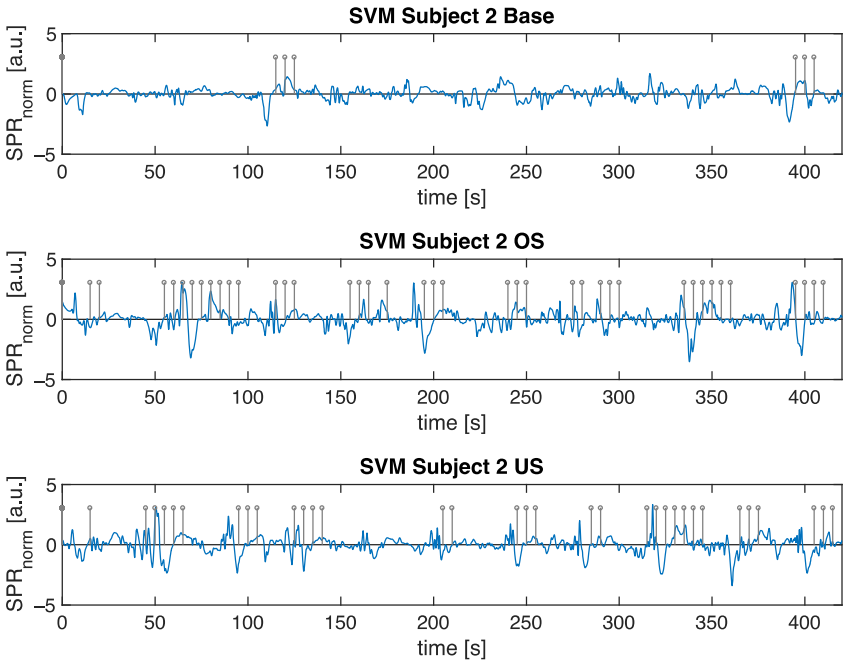

Figure 12. Output of the SVM classifier for subject 2 in the three different car setups. Only the labels equal to "1" (represented by a grey stem), added at the end of each corresponding 15 s interval, are displayed.

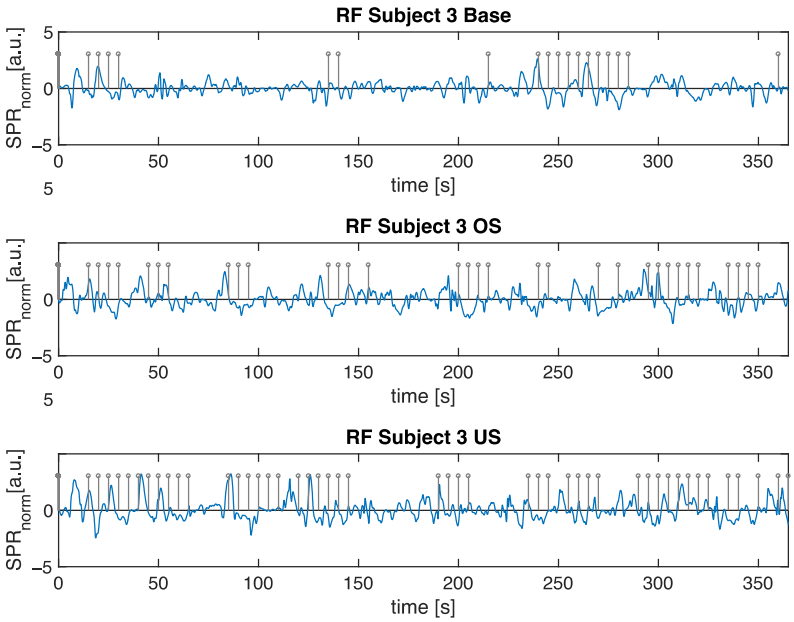

Figure 13. Output of the RF classifier for subject 3 in the three different car setups. Only the labels equal to "1" (represented by a grey stem), added at the end of each corresponding 15 s interval, are displayed.

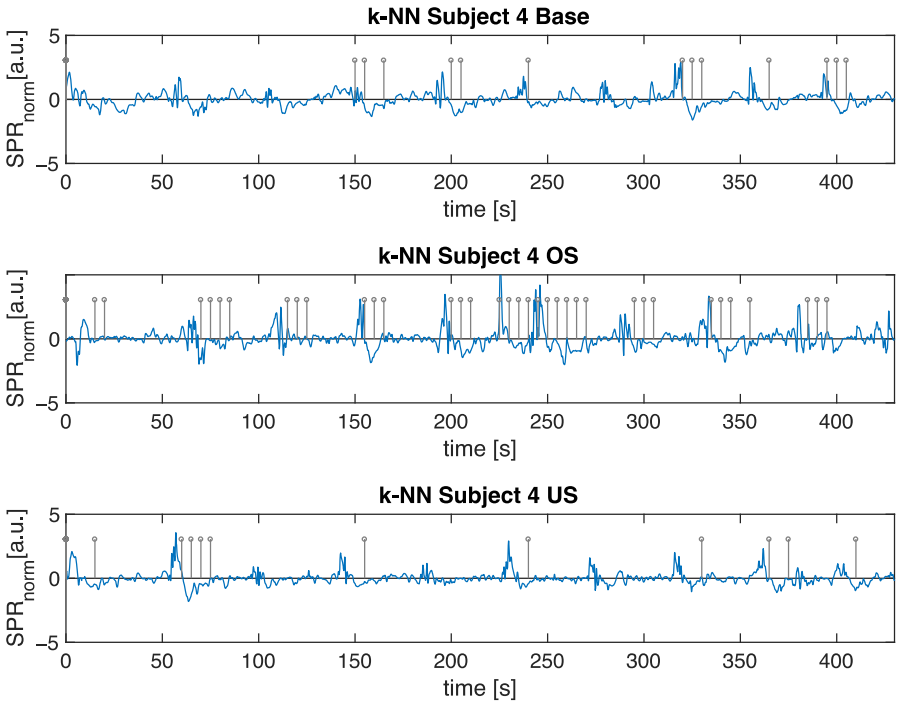

Figure 14. Output of the k-NN classifier for subject 4 in the three different car setups. Only the labels equal to "1" (represented by a grey stem), added at the end of each corresponding 15 s interval, are displayed.

3.2. ECG Signal Analysis

As we described in Section 2, we logged three ECG signals from the chest of the subjects [32]. However, we chose the signal (among the three) with the highest signal-to-noise ratio, depending on the subject under test. We then proceeded by evaluating how the instantaneous heart rate (HR) changes in the three tested car settings. We could not identify a significant difference in the mean HR in the various cases. In fact, the mean HR of the subjects was comparable in all the different setups, with average values of 83.20, 98.65, 80.45, and 96.67 bpm for the 1st, 2nd, 3rd, and 4th subjects, respectively. However, we also evaluated both the *t*-test and Wilcoxon test for the HR values computed in the 10 segments in which the whole track was divided. The significance level values are reported in Tables 5 and 6 for the *t*-test and Wilcoxon test, respectively. In most cases, they do not reveal a significant difference between the HR values when considering the different car handling settings, except when comparing the OS and US setups for subjects 1 and 3 (in the *t*-test) and only for subject 3 (in the Wilcoxon test).

Table 5. Paired *t*-test probability *p* considering Base and OS, Base and US, and US and OS car setups, for all subjects and their HR values in the 10 track segments.

Subject	1	2	3	4
Base vs. OS	0.34	0.80	0.70	0.19
Base vs. US	0.77	0.32	0.24	0.76
OS vs. US	**0.03**	0.25	**0.05**	0.11

Table 6. Paired Wilcoxon test probability p considering Base and OS, Base and US, and US and OS car setups, for all subjects and their HR values in the 10 track segments.

Subject	1	2	3	4
Base vs. OS	0.27	0.73	0.62	0.34
Base vs. US	0.97	0.34	0.19	0.79
OS vs. US	0.31	0.52	**0.05**	0.16

In Figure 15, we also show the mean and SD, in bpm, of the HR values extracted from the ten segments in which the track is divided for each subject and each car setting. We can note that there is not much variability among the HR results.

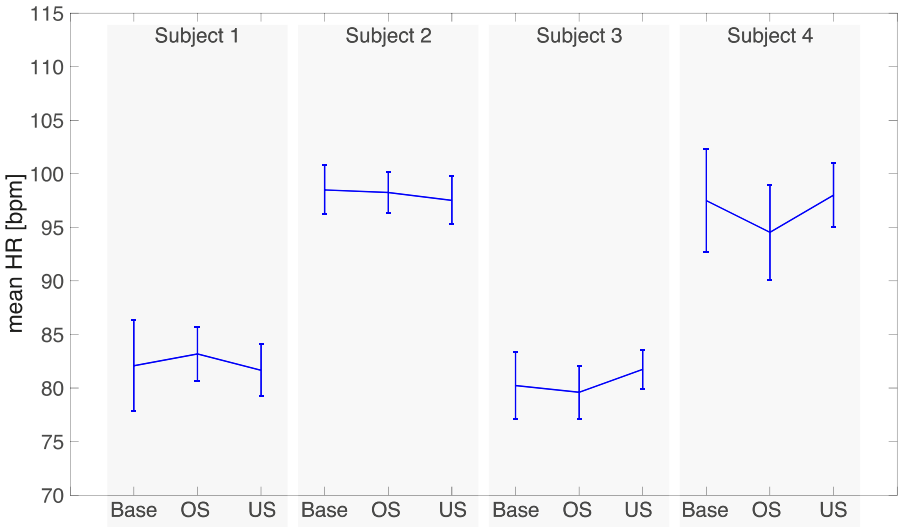

Figure 15. Mean HR (in bpm) and standard deviation, for each driver and each car setup.

As an example, Figure 16 shows the recorded HR for subject 2, along the entire course, with the three setups.

In addition to the previous analysis based on HR values, we evaluated several parameters of HRV, such as the mean value of normal-to-normal RR intervals, the standard deviation of RR intervals (SDNN), the root mean square of subsequent RR interval differences (RMSSD), the number of subsequent RR intervals differing more than 50 ms (NN50), the corresponding relative value in percentage (PNN50), the low frequency (LF) and high frequency (HF) power spectra, and the ratio LF/HF to be used as input features to the ML algorithms (similarly to what we did in [40]). However, in this work, the HRV parameters did not provide significant differences on the final experimental results. An additional example showing the SDNN signal for subject 2 along the entire course, with the various setups, is reported in Figure 17. This signal is obtained by interpolating the original RR signal time series in order to obtain a signal with equidistant sampled points [47], and for each sample, the previous 120 s interval is used to compute the corresponding SDNN value. As confirmed by the classifier performance, the signal behavior does not suggest a visible difference among the three setups. Similar signal behavior was observed for the other subjects.

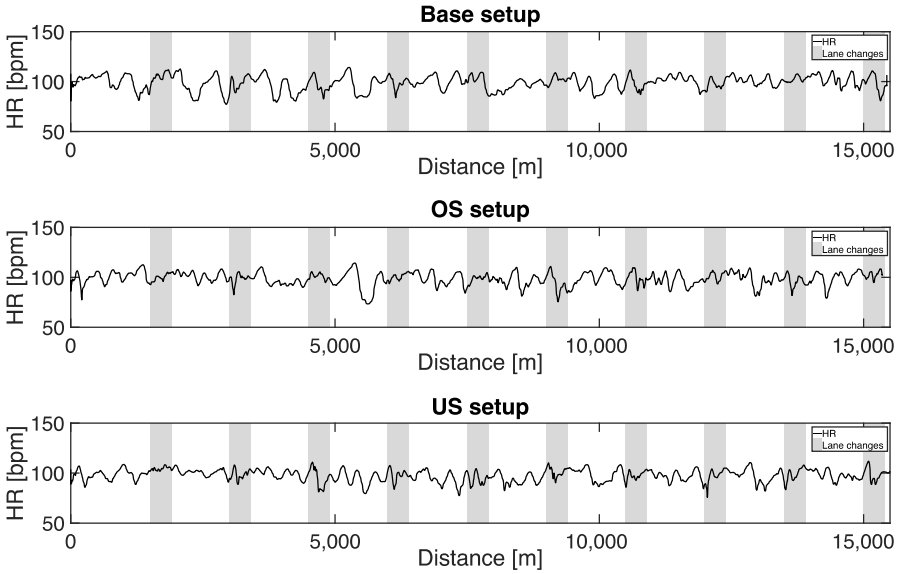

Figure 16. HR signal recorded from subject 2 during the drive, with Base (top), OS (center), and US (bottom) car setups, respectively. Grey rectangles indicate the sections of double lane changes.

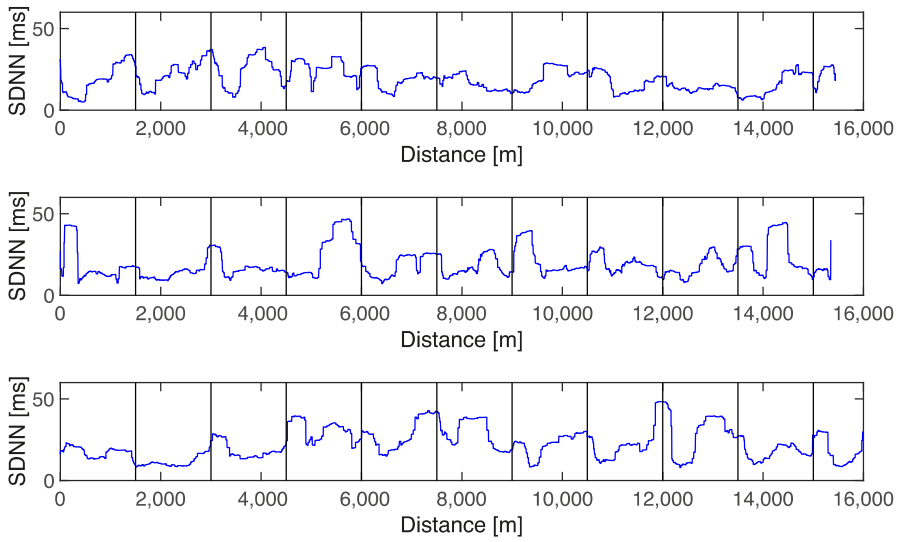

Figure 17. SDNN signal recorded from subject 2 during the drive, with Base (top), OS (center), and US (bottom) car setups, respectively. Vertical lines denote the task positions.

3.3. Eye Tracking Signal Analysis

In the experiments we recorded real-time data of eye movements in the subjects. Studies in the literature, evaluating mental stress in subjects, show that the pupil diameter can be considered an effective indicator of autonomic nervous activity [48]. So, in this work, as a first step in our research employing eye signals, we focused on the pupil diameter. We observed that, no matter what the car setup is used, subject's pupil will increase in diameter as they advance toward the cones. For example, for subject 1, Figure 18 shows this phenomenon in the Base, OS, and US car setups. As before, for all the subjects, there is not any major difference in eye response with the various car setups. We can thus conclude that the drivers react when approaching the tasks, but without any remarkable difference among the three setups.

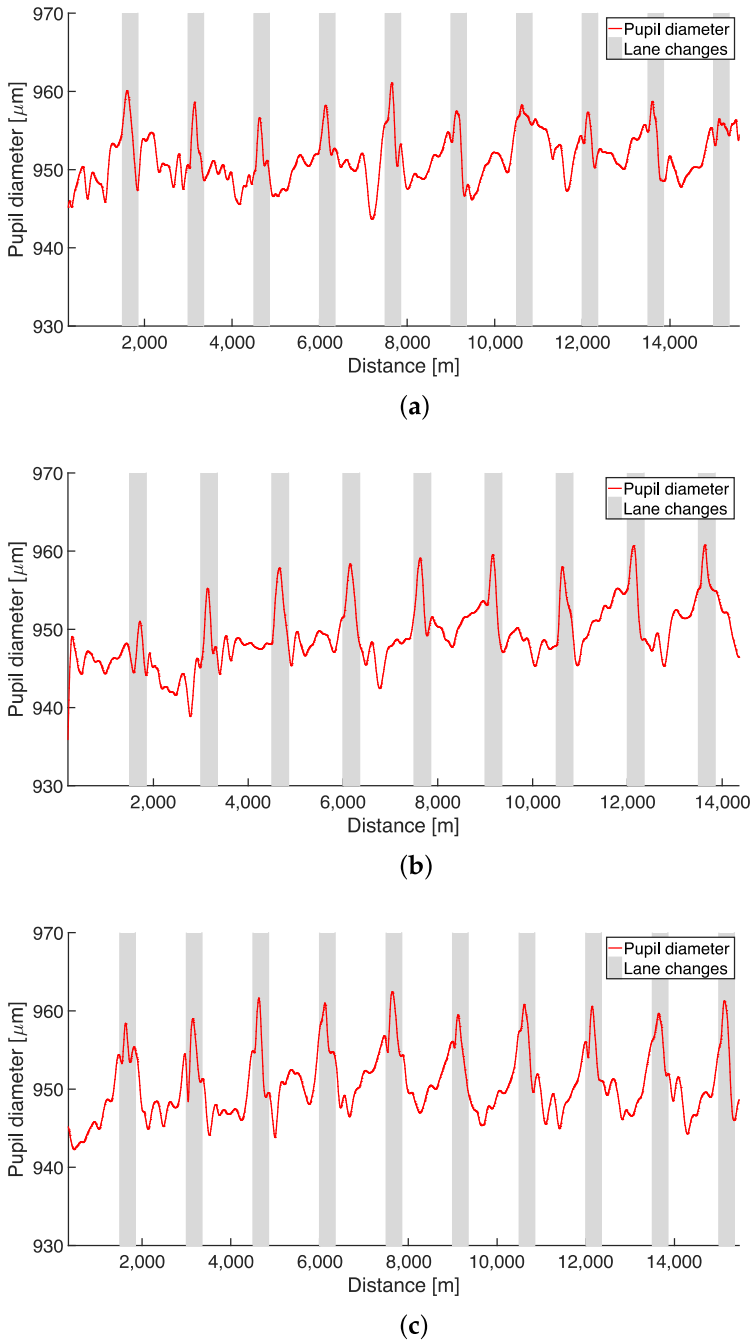

Figure 18. Changes on the pupil diameter recorded from subject 1 during the drive, with Base (**a**), OS (**b**), and US (**c**) car setups, respectively. Grey rectangles indicate the sections of double lane changes.

4. Discussion

The results presented in the previous section bring out some aspects that are worth discussing. First, we observe that the SPR signals, after the MA removal block, can discriminate among different drivers' emotional responses in the Base, OS or US car setups. The Wilcoxon test and the paired *t*-test showed that the SPR activity in the three car setups is different, with statistical significance in Base vs. OS and Base vs. US. Thus, looking only at SPR signals, the Base setup resulted in being significantly less stressful than both US and OS ones. We also computed the power of both statistical tests, and we showed that, even if the number of samples considered in each data set is limited ($n = 10$), the obtained

values are consistent and acceptable, at least considering the rough assumptions used in the computation.

Moreover, using the SPR signals to train different ML algorithms, we observe a neutral EDA response and only a small number of time intervals classified with the "stress" label in the Base configuration. When the car handling is set to US or OS, instead, the subject's stress reaction, as measured by the signals, increases. This supports the findings obtained with the *t*-test and Wilcoxon on SPR activity.

As a second aspect, ECG signals and eye movements did not provide useful information to improve the results obtained with just the SPR. In particular, we observed that there were limited changes in the HR values, and the HRV parameters did not provide significant differences on the final ML classification performance. This was also confirmed by both the *t*-test and Wilcoxon test results based on HR values. One reason for this behavior could be that the movement and physical activity of the professional pilots during driving could have masked the ECG variability related to stress, which instead we observed in other experiments with sudden stress-evoking events [40] in a situation which was different from the one we are facing here. Focusing on eye movements, we noted that the diameter of the pupil always increases when drivers approach the tasks, regardless of the car setup they are testing. This may appear to be in contrast with the results reported in [48], but we have to point out that our experimental setup is very different from the one presented in [48]. In that case, the volunteers had to stare at the screen waiting for stimuli to appear. In our case, instead, the drivers had to change the gaze direction quite often in order to monitor the dashboard, the mirrors, and the road ahead. This continuous change in the gaze direction may have altered the response of the pupil diameter due to the stress induced by the different car setups. Therefore, in our setup, the pupil diameter has not enough sensitivity to discern among the effects of the vehicle dynamics.

The capacity to recognize different emotional responses, arising with different car-handling settings, is a crucial aspect in the automotive market, in particular when designing a new car. As already stated, the proposed system has some innovative elements, such as an MA algorithm to reduce artifacts, and the use of the SPR signal, which can result in some advantages over the more commonly used SCR signal. There are, however, some limitations in our study. The first one is due to the low number of subjects involved in our experiment, and of course the availability of further acquisitions would make the conclusions safer. Unfortunately, because of the type of experiment involving professional drivers and complex dynamic models of car prototypes, we did not have the chance to collect additional data with more subjects. Due to the great variability of individual responses (see Figure 6), we focused, therefore, on the evaluation of each driver's response to the different setups. We believe that the collected data are rather consistent, and that the paper provides a clear indication about the applicability of the proposed approach for the application we are taking into account. In addition, as introduced before, we performed the paired *t*-test and Wilcoxon test on the Base–OS, Base–US, and OS–US pairs, using ten values for each subject and each setup. This is a limited number which, however, can be acceptable for statistical purposes [42].

The second limitation is due to the hypothesis, made during the training phase of the classification procedure, that the driver's stress appears only during the stress-inducing episodes. This way, we are supposing that the subjects are continuously stressed during these events, but this could not be always true. We are also supposing that they are not stressed outside of these episodes, but it could happen that an EDA stimulus is evoked for reasons that we do not know and cannot control.

Finally, we observe that the proposed system is capable of classifying stress or no-stress conditions in real time. In fact, once the training phase of the classifiers is completed, the classification of new signal segments can be carried out without excessive use of resources. Therefore, this system is able to provide real-time classification every 15 s signal segment, allowing the recognition of stressful episodes of a short length while subjects are driving.

5. Conclusions

We present a system that allows us to compare the physiological responses of subjects driving on a professional simulator where we change the vehicle dynamics. As indicated by the SPR signal characteristics and the classifier outputs, the Base car setup resulted in being the least stressful setup for most drivers. These findings, obtained by using some ML algorithms, are also validated by the statistical *t*-test and the Wilcoxon test on the SPR RMS values. Even if the number of the considered subjects is somehow limited, with this study, we observed that the SPR signal and the proposed classification system can help us to evaluate the stress level of the subject when considering different car-handling settings. As a first step in comparing and evaluating the physiological signals in people driving with different car settings, the results show that it is feasible to automatically assess their acceptability. We think that this is a valuable outcome, since the driver's response to different car setups is a crucial aspect when designing a car for safety, comfort, and performance reasons. In our experiment, the ECG and eye movement do not appear to add meaningful insight about the different situations. The development of new test scenarios involving more subjects—as well as additional data coming from various devices, such as wearables, that allow us to design a multimodal acquisition system—could provide the opportunity for further research. Another possible future development of this study is the application of the proposed measurement setup also in a real driving scenario, since the wearable SPR and ECG sensors can be easily worn by drivers in a real car.

Author Contributions: The authors contributed equally to the research investigation and validation, and to the writing of this paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by VI-grade S.r.l., Tavagnacco (UD), Italy, grant number CV19_023.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We are grateful to VI-grade S.r.l., Tavagnacco (UD), Italy, and in particular, to Fabio Formaggia, Diego Minen, Michela Minen, and Carlo Savorgnan, for their assistance in data acquisition with their professional driving simulator.

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

Abbreviations

The following abbreviations are used in this manuscript:

SPR	Skin Potential Response
EDA	Electrodermal Activity
ECG	Electrocardiogram
MA	Motion Artifact
ML	Machine Learning
SVM	Support Vector Machine
RF	Random Forest
k-NN	k-Nearest Neighbors

References

1. Casaccia, S.; Revel, G.M.; Cosoli, G.; Scalise, L. Assessment of Domestic Well-Being: From Perception to Measurement. *IEEE Instrum. Meas. Mag.* **2021**, *24*, 58–67. [CrossRef]
2. Bowen, L.; Budden, S.L.; Smith, A.P. Factors underpinning unsafe driving: A systematic literature review of car drivers. *Transp. Res. Part F Traffic Psychol. Behav.* **2020**, *72*, 184–210. [CrossRef]
3. Miyama, G.; Fukumoto, M.; Kamegaya, R.; Hitosugi, M. Risk factors for collisions and near-miss incidents caused by drowsy bus drivers. *Int. J. Environ. Res. Public Health* **2020**, *17*, 4370. [CrossRef] [PubMed]
4. Wu, W.T.; Tsai, S.S.; Wang, C.C.; Lin, Y.J.; Wu, T.N.; Shih, T.S.; Liou, S.H. Professional driver's job stress and 8-year risk of cardiovascular disease: The Taiwan bus driver cohort study. *Epidemiology* **2019**, *30*, S39–S47. [CrossRef] [PubMed]
5. Jensen, A.; Kaerlev, L.; Tüchsen, F.; Hannerz, H.; Dahl, S.; Nielsen, P.S.; Olsen, J. Locomotor diseases among male long-haul truck drivers and other professional drivers. *Int. Arch. Occup. Environ. Health* **2008**, *81*, 821–827. [CrossRef] [PubMed]
6. Lohani, M.; Payne, B.R.; Strayer, D.L. A review of psychophysiological measures to assess cognitive states in real-world driving. *Front. Hum. Neurosci.* **2019**, *13*, 57. [CrossRef] [PubMed]
7. Kim, S.; Rhee, W.; Choi, D.; Jang, Y.J.; Yoon, Y. Characterizing driver stress using physiological and operational data from real-world electric vehicle driving experiment. *Int. J. Automot. Technol.* **2018**, *19*, 895–906. [CrossRef]
8. Zhang, M.; Ihme, K.; Drewitz, U. Discriminating drivers' emotions through the dimension of power: Evidence from facial infrared thermography and peripheral physiological measurements. *Transp. Res. Part Traffic Psychol. Behav.* **2019**, *63*, 135–143. [CrossRef]
9. Din, I.U.; Almogren, A.; Guizani, M.; Zuair, M. A decade of Internet of Things: Analysis in the light of healthcare applications. *IEEE Access* **2019**, *7*, 89967–89979. [CrossRef]
10. Wang, H.; Ma, C.; Zhou, L. A brief review of machine learning and its application. In Proceedings of the 2009 International Conference on Information Engineering and Computer Science, Wuhan, China, 19–20 December 2009; pp. 1–4. [CrossRef]
11. Shalev-Shwartz, S.; Ben-David, S. *Understanding Machine Learning: From Theory to Algorithms*; Cambridge University Press: Cambridge, UK, 2014.
12. Kumar, M.; Weippert, M.; Vilbrandt, R.; Kreuzfeld, S.; Stoll, R. Fuzzy evaluation of heart rate signals for mental stress assessment. *IEEE Trans. Fuzzy Syst.* **2007**, *15*, 791–808. [CrossRef]
13. Liew, W.S.; Seera, M.; Loo, C.K.; Lim, E.; Kubota, N. Classifying stress from heart rate variability using salivary biomarkers as reference. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *27*, 2035–2046. [CrossRef] [PubMed]
14. Zhang, J.; Wen, W.; Huang, F.; Liu, G. Recognition of real-scene stress in examination with heart rate features. In Proceedings of the 2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Hangzhou, China, 26–27 August 2017; Volume 1, pp. 26–29. [CrossRef]
15. Hwang, B.; You, J.; Vaessen, T.; Myin-Germeys, I.; Park, C.; Zhang, B.T. Deep ECGNet: An optimal deep learning framework for monitoring mental stress using ultra short-term ECG signals. *Telemed. e-Health* **2018**, *24*, 753–772. [CrossRef]
16. Qin, Z.; Li, M.; Huang, L.; Zhao, Y. Stress level evaluation using BP neural network based on time-frequency analysis of HRV. In Proceedings of the 2017 IEEE International Conference on Mechatronics and Automation (ICMA), Takamatsu, Japan, 6–9 August 2017; pp. 1798–1803. [CrossRef]
17. Saeed, S.M.U.; Anwar, S.M.; Khalid, H.; Majid, M.; Bagci, U. EEG based classification of long-term stress using psychological labeling. *Sensors* **2020**, *20*, 1886. [CrossRef] [PubMed]
18. Lin, Y.P.; Wang, C.H.; Jung, T.P.; Wu, T.L.; Jeng, S.K.; Duann, J.R.; Chen, J.H. EEG-based emotion recognition in music listening. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 1798–1806. [CrossRef] [PubMed]
19. Yang, Y.; Wu, Q.J.; Zheng, W.L.; Lu, B.L. EEG-based emotion recognition using hierarchical network with subnetwork nodes. *IEEE Trans. Cogn. Dev. Syst.* **2018**, *10*, 408–419. [CrossRef]
20. Becker, H.; Fleureau, J.; Guillotel, P.; Wendling, F.; Merlet, I.; Albera, L. Emotion recognition based on high-resolution EEG recordings and reconstructed brain sources. *IEEE Trans. Affect. Comput.* **2020**, *11*, 244–257. [CrossRef]
21. Rigas, G.; Goletsis, Y.; Bougia, P.; Fotiadis, D.I. Towards driver's state recognition on real driving conditions. *Int. J. Veh. Technol.* **2011**, *2011*, 1–14. [CrossRef]
22. Rigas, G.; Goletsis, Y.; Fotiadis, D.I. Real-time driver's stress event detection. *IEEE Trans. Intell. Transp. Syst.* **2012**, *13*, 221–234. [CrossRef]
23. Boucsein, W. *Electrodermal Activity*; Springer: New York, NY, USA, 2012.
24. Choi, M.; Koo, G.; Seo, M.; Kim, S.W. Wearable device-based system to monitor a driver's stress, fatigue, and drowsiness. *IEEE Trans. Instrum. Meas.* **2018**, *67*, 634–645. [CrossRef]
25. Halim, Z.; Rehan, M. On identification of driving-induced stress using electroencephalogram signals: A framework based on wearable safety-critical scheme and machine learning. *Inf. Fusion* **2020**, *53*, 66–79. [CrossRef]
26. Movahedi, F.; Coyle, J.L.; Sejdić, E. Deep belief networks for electroencephalography: A review of recent contributions and future outlooks. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 642–652. [CrossRef] [PubMed]
27. Zontone, P.; Affanni, A.; Bernardini, R.; Del Linz, L.; Piras, A.; Rinaldo, R. Stress Evaluation in Simulated Autonomous and Manual Driving through the Analysis of Skin Potential Response and Electrocardiogram Signals. *Sensors* **2020**, *20*, 2494. [CrossRef] [PubMed]

28. Zontone, P.; Affanni, A.; Piras, A.; Rinaldo, R. Exploring Physiological Signal Responses to Traffic-Related Stress in Simulated Driving. *Sensors* **2022**, *22*, 939. [CrossRef] [PubMed]
29. Zontone, P.; Affanni, A.; Bernardini, R.; Del Linz, L.; Piras, A.; Rinaldo, R. Emotional response analysis using electrodermal activity, electrocardiogram and eye tracking signals in drivers with various car setups. In Proceedings of the 2020 28th European Signal Processing Conference (EUSIPCO), Amsterdam, The Netherlands, 18–21 January 2021; pp. 1160–1164. [CrossRef]
30. VI-Grade S.r.l. Available online: https://www.vi-grade.com (accessed on 15 July 2021).
31. Affanni, A. Dual-channel electrodermal activity and an ECG wearable sensor for measuring mental stress from the hands. *Acta IMEKO* **2019**, *8*, 56–63. [CrossRef]
32. Affanni, A. Wearable instrument to measure simultaneously cardiac and electrodermal activities. In Proceedings of the 2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Benevento, Italy, 15–18 May 2016; pp. 1–5. [CrossRef]
33. Smart Eye Aurora. Available online: https://smarteye.se (accessed on 15 July 2021).
34. Posada-Quintero, H.F.; Chon, K.H. Innovations in Electrodermal Activity Data Collection and Signal Processing: A Systematic Review. *Sensors* **2020**, *20*, 479. [CrossRef] [PubMed]
35. Tronstad, C.; Kalvøy, H.; Grimnes, S.; Martinsen, Ø. Waveform difference between skin conductance and skin potential responses in relation to electrical and evaporative properties of skin. *Psychophysiology* **2013**, *50*, 1070–1078. [CrossRef] [PubMed]
36. Tronstad, C.; Grimnes, S.; Martinsen, Ø.; Amundsen, V.; Wojniusz, S. PC-based instrumentation for electrodermal activity measurement. *J. Phys. Conf. Ser.* **2010**, *224*. [CrossRef]
37. Affanni, A.; Chiorboli, G. Design and characterization of a real-time, wearable, endosomatic electrodermal system. *Measurement* **2015**, *75*, 111–121. [CrossRef]
38. Bari, D.; Yacoob, H.; Tronstad, C.; Kalvøy, H.; Martinsen, Ø. Electrodermal responses to discrete stimuli measured by skin conductance, skin potential, and skin susceptance. *Ski. Res. Technol.* **2018**, *24*, 108–116. [CrossRef]
39. Affanni, A.; Piras, A.; Rinaldo, R.; Zontone, P. Dual channel Electrodermal activity sensor for motion artifact removal in car drivers' stress detection. In Proceedings of the 2019 IEEE Sensors Applications Symposium (SAS), Sophia Antipolis, France, 11–13 March 2019; pp. 1–6. [CrossRef]
40. Zontone, P.; Affanni, A.; Bernardini, R.; Piras, A.; Rinaldo, R.; Formaggia, F.; Minen, D.; Minen, M.; Savorgnan, C. Car driver's sympathetic reaction detection through electrodermal activity and electrocardiogram measurements. *IEEE Trans. Biomed. Eng.* **2020**, *67*, 3413–3424. [CrossRef] [PubMed]
41. Association, W.M. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *JAMA* **2013**, *310*, 2191–2194. [CrossRef]
42. De Winter, J.C.F. Using the Student's t-test with extremely small sample sizes. *Pract. Assess. Res. Eval.* **2013**, *18*, *1–12*. [CrossRef]
43. Ellis, P.D. *The Essential Guide to Effect Sizes: An Introduction to Statistical Power, Meta-Analysis and the Interpretation of Research Results*; Cambridge University Press: Cambridge , UK, 2010.
44. Zontone, P.; Affanni, A.; Bernardini, R.; Piras, A.; Rinaldo, R. Stress detection through electrodermal activity (EDA) and electrocardiogram (ECG) analysis in car drivers. In Proceedings of the 2019 27th European Signal Processing Conference (EUSIPCO), A Coruna, Spain, 2–6 September 2019; pp. 1–5. [CrossRef]
45. Zontone, P.; Affanni, A.; Bernardini, R.; Linz, L.D.; Piras, A.; Rinaldo, R. Supervised learning techniques for stress detection in car drivers. *Adv. Sci. Technol. Eng. Syst. J.* **2020**, *5*, 22–29. [CrossRef]
46. Statistics and Machine Learning Toolbox. Available online: https://mathworks.com/products/statistics.html (accessed on 9 January 2022).
47. Tarvainen, M.P.; Niskanen, J.P.; Lipponen, J.A.; Ranta-Aho, P.O.; Karjalainen, P.A. Kubios HRV—Heart rate variability analysis software. *Comput. Methods Programs Biomed.* **2014**, *113*, 210–220. [CrossRef] [PubMed]
48. Yamanaka, K.; Kawakami, M. Convenient evaluation of mental stress with pupil diameter. *Int. J. Occup. Saf. Ergon.* **2009**, *15*, 447–450. [CrossRef] [PubMed]