

Supervised Learning Techniques for Stress Detection in Car Drivers

Pamela Zontone¹, Antonio Affanni¹, Riccardo Bernardini¹, Leonida Del Linz², Alessandro Piras¹, Roberto Rinaldo^{*1}

¹DPIA, Polytechnic Department of Engineering and Architecture, University of Udine, Udine, 33100, Italy

²VI-grade Srl, Via G. Galilei 42, Tavagnacco, Udine, 33010, Italy

ARTICLE INFO

Article history:

Received: 31 August, 2020

Accepted: 21 October, 2020

Online: 08 November, 2020

Keywords:

Stress Detection in Car Drivers

Electrodermal Activity

Heart Rate Variability

Motion Artifact

Supervised Learning Algorithm

Machine Learning

Deep Learning

ABSTRACT

In this paper we propose the application of supervised learning techniques to recognize stress situations in drivers by analyzing their Skin Potential Response (SPR) and the Electrocardiogram (ECG). A sensing device is used to acquire the SPR from both hands of the drivers, and the ECG from their chest. We also consider a motion artifact removal algorithm that allows the generation of a single cleaned SPR signal, starting from the two SPR signals, which could be characterized by artifacts due to vibrations or movements of the hands on the wheel. From both the cleaned SPR and the ECG signals we compute some statistical features that are used as input to six Machine Learning Algorithms for stress or non-stress episodes classification. The SPR and ECG signals are also used as input to Deep Learning Algorithms, thus allowing us to compare the performance of the different classifiers. The experiments have been carried out in a firm specialized in developing professional car driving simulators. In particular, a dynamic driving simulator has been used, with subjects driving along a straight road affected by some unanticipated stress-evoking events, located at different positions. We obtain an accuracy of 88.13% in stress recognition using a Long Short-Term Memory (LSTM) network.

1 Introduction

Lately¹, the assessment of stress in car drivers is getting more attention, because a highly stressed driver could be less focused, more likely to drive in a risky way [1], and/or to get involved into accidents [2]. The emotional state, corresponding to a condition of tension and pressure in an individual, causes reactions of the sympathetic nervous system. These are defined as stress in literature [3].

Different approaches have been proposed to detect stress in individuals, using physical or physiological parameters [4]. Machine Learning (ML) algorithms are often employed [5]– [8]. In [9], a Support Vector Machine (SVM) classifier is used, with Skin Conductance Response (SCR), Electroencephalogram (EEG), and Heart Rate (HR) as inputs. In [10], various signals are utilized, such as Blood Volume Pulse (BVP), Skin Conductance (SC), Skin Temperature (ST), Electromyography (EMG), and Respiration (RESP). Deep Learning (DL) algorithms have also been successfully applied for stress recognition. Physiological signals are used, such as Electrocardiogram (ECG) [11] and Electrodermal Activity (EDA) [12], or facial expressions [13], or a collection of physiological and motion data, including ECG, EDA, EMG, RESP, Temperature (TEMP),

and Acceleration (ACC) [14]. The system described in [15] uses the time differences between consecutive R-waves of QRS complexes, also called RR intervals, for mental stress identification in firefighters. In several works, a comparison between ML and DL methods for stress recognition has been carried out, also showing that Convolutional Neural Networks (CNNs) can outperform ML conventional methods [16, 17].

There are various stress analysis approaches when we consider car driving scenarios as well. Many use physical characteristics, e.g., facial or eye images [18, 19]. In others, physiological characteristics are considered [20–22]. Among these, EDA can be used jointly with different signals [23, 24]. Authors in [25] study the relation between the vehicle's acceleration and the driver's heart rate in elders. In [26], EEG patterns, belonging to different driving conditions, are classified using an SVM, a Neural Network (NN), and a Random Forest (RF). Some DL algorithms have also been successfully applied for stress recognition in the driving context as well, and are particularly relevant to this work. Authors in [27] present an end-to-end architecture, composed of convolutional layers, that uses ECG signals for stress detection. Their experiments are designed for identification of mental fatigue and for discrimina-

*Corresponding Author: Roberto Rinaldo, roberto.rinaldo@uniud.it

¹This paper is an extension of the work originally presented in [43].

tion between driving and resting conditions. A multimodal system based on CNNs and Long Short-Term Memory (LSTM) networks, using ECG, vehicle data, and contextual data has been proposed in [28]. In [29], CNNs are considered to detect a driver's braking intention, in different driving conditions, using EEG data.

In this work we present a system that allows the recognition of the emotional and stress conditions of subjects while driving. The main features of our system, which differentiate our work from others in the literature, are the relative simplicity of the sensors used, the possibility of being used nearly in real time on 15 s signal blocks, and the ability to discriminate stressful events of short duration during driving. The experiments are organized in a company that designs professional driving simulators. We use a simulator that reproduces a road track in a 3D environment. Along the track we place different tasks to go through, which are expected to cause a reaction in the subjects. Two different signals are considered, the Skin Potential Response (SPR) and the ECG. We have proven that these signals are suitable for the analysis of the emotional condition of a driver and that can be acquired with an acceptably intrusive setup, with no intervention on the vehicle [30].

We analyze the SPR signal instead of the widely utilized SCR, because it can be logged in a more straightforward way, with no electric current in the skin. In addition, the electrode impedance and skin impedance changes have a lower impact on this signal [31]. In detail, our objective is to compare the performance of several ML algorithms in recognizing the stress responses of the drivers that undergo various stress situations. To do this, some features are extracted from both the SPR and ECG signals. In addition, we consider deep learning techniques for driver stress recognition. These require the raw recorded signals as input, contrary to what happens in traditional machine learning techniques, where a feature extraction process is needed. Specifically, we use a CNN and an LSTM network. Compared to other similar studies cited above [24, 27], we apply machine and deep learning techniques for detecting stress during short tasks in a controlled and simulated environment.

The paper is structured as follows. In the next section, a description of the fundamental blocks of the proposed system is provided. We introduce the sensing device utilized for both SPR and ECG recordings, the Motion Artifact (MA) removal algorithm developed to reduce motion artifacts, and the learning algorithms implemented for classification. Section 3 reviews the experimental setup. Section 4 presents the results of our comparative study, using different classification algorithms, on drivers' stress recognition. Finally, in Section 5 we draw some conclusions.

2 Description of the system

The block diagram of the developed system is shown in Figure 1. The inputs are the two SPR signals, recorded from the hands of a test subject ($s_1(t)$ and $s_2(t)$ in the figure), and the ECG signal recorded from the chest of the subject. The signals are acquired synchronously through a sensing device that uses a textile garment with electrodes, in addition to other electrodes placed under the gloves on both hands.

The two SPR signals can exhibit some artifacts that can be generated when the subject is driving, e.g., due to the hands movements

or vibrations. To overcome this problem, we designed a MA removal algorithm that, taking into account the local energy of the two SPR signals, picks the smoother one, with the aim to remove spurious peaks and asymmetric signal spikes, possibly related to MA. As a matter of fact, the SPR component due to the sympathetic nervous system activity in the two hands is highly correlated. In the end, a learning classification algorithm is used to recognize the presence of stress in certain time intervals. In detail, every single time interval is classified into two possible classes, which are the stress class (or "1") and the non-stress class (or "0"). In case of ML algorithms, this is carried out by selecting some statistical features from the cleaned SPR signal ($s(t)$ in Figure 1) and from the ECG signal. These features have been picked out, among others, since they well represent the stress signal peculiarities. In case of DL algorithms, instead, the raw SPR and ECG recordings are directly processed. In the next subsections, we will describe in detail each one of these fundamentals blocks.

2.1 The sensing device

The sensor used for data acquisition is the VI-BioTelemetry system (<https://www.vi-grade.com/en/solutions/vi-biotelemetry/>). This commercial sensor, shown in Figure 2, has been designed in our group and engineered by the VI-grade company. Its detailed description can be found in several recent scientific papers (e.g., [31, 32]).

Summarizing its features and specifications, the sensor is battery operated and transmits data to a laptop via WiFi connection. It acquires two SPR channels and three ECG channels. The SPR signals are acquired from the hands of the tested subject (through Ag/AgCl wet electrodes positioned on the palm and the back of each hand) and the ECG data are acquired using a commercial vest having textile electrodes in contact with the chest of the subject (using an adhesive conductive gel). The analog front end conditions the SPR and ECG signals with proper bandwidth and gain. Analog signals are converted using a 12-bit A/D and an on board DSP which acquires the signals at 1 kSa/s rate. The A/D converter is characterized by an Integral Nonlinearity (INL) lower than 1 LSB (Least Significant Bit) and an Effective number of bits (ENOB) equal to 11.1. The DSP sends the acquired signals to the WiFi module which is responsible to send all the time-aligned data to a server in the local area network.

Regarding the ECG channels, they are basically designed as band-pass differential amplifiers with maximum input range ± 5 mV (i.e., gain 370) and bandwidth in the [0.03, 160] Hz range. The input impedance of each channel is 100 M Ω , in order to reduce the load error to less than 1%, since the skin impedance is lower than 1 M Ω . The resolution of the ECG signals acquired on the chest skin results to be in the order of 2.5 μ V and the accuracy of the ECG readout has been characterized in the past [32] to be as low as 0.1%. As for the ECG signal, we first detect the R-peak locations using the Pan-Tompkins algorithm [33].

The instantaneous RR signal is obtained by converting the RR interval time series with cubic spline interpolation at 100 Sa/s. Similarly to what is done in [34], where RR values and DL are used for Congestive Heart Failure (CHF) detection, interpolation is needed to synchronize signal segments, in our case with the SPR signal blocks. Note that this can be important to preserve correct time and

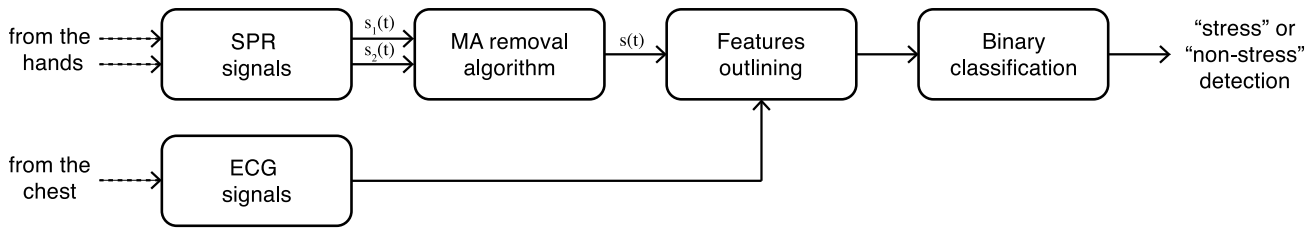


Figure 1: Block diagram of the presented system.

frequency information (see [35] and references therein), which the DL classifiers can possibly extrapolate from the data.

Regarding the SPR channels, the sensors acquire the electrodermal signals on both hands as differential voltages between the palm and the back of each hand. The SPR signals are supposed to be in the ± 10 mV range. They are amplified with a gain equal to 160. The acquisition bandwidth is in the $[0.08, 40]$ Hz range; the input impedance of each channel is 100 M Ω . The resolution of the SPR data acquired on the hands results in the order of 5 μ V and the accuracy of the SPR readout has been characterized in the past [31] to be in the order of 0.2%. The SPR signals are subsampled at 100 Sa/s before further processing.

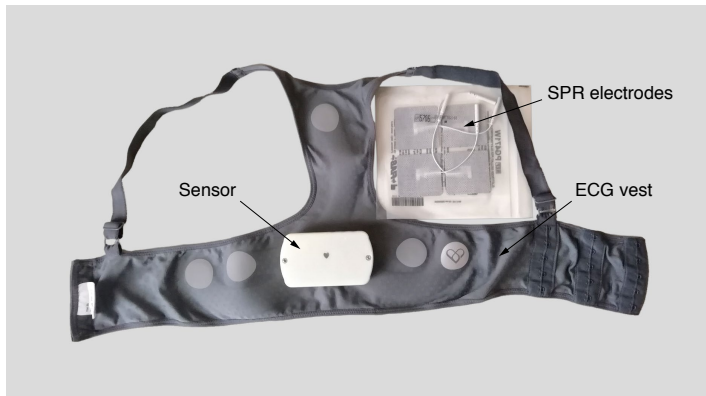


Figure 2: The VI-BioTelemetry sensor used in the experiments. The system is composed by an acquisition box, a vest with textile electrodes for ECG acquisition on the chest and Ag/AgCl electrodes for SPR acquisition on the hands.

2.2 Motion artifact removal block

As introduced above, some artifacts due to the movements and actions of the subject's hands while driving can become visible on the SPR signals, and could alter their shape. To take this issue into account, we propose an MA algorithm (see also [36, 37]) that is built on two assumptions: the first being that motion artifact enhances the local energy of the signal. The second being that the motion artifacts rarely appear simultaneously in the SPR signal of both hands. We observed this behaviour in our testing most of the time, and the proposed solution appears to be a good compromise to lessen the MA. Considering the RMS values σ_1 and σ_2 of the SPR input signals s_1 and s_2 on a $L = 100$ samples wide moving window, corresponding to 1 s at a sample rate of 100 Sa/s, we establish a smoothing threshold function $f(x)$ in order to obtain $f(x) \approx 0$ when

$0 \leq x \ll 1$ and $f(x) \approx 1$ when $1 \ll x \leq 2$, namely,

$$f(x) = \frac{1}{1 + e^{-2(x-1)}} \quad (1)$$

This function is employed in a correction term β , that enables us to reduce motion artifacts. At the n^{th} sample, this correction term is calculated as:

$$\beta(n) = \begin{cases} f\left(\frac{\sigma_1(n)}{\sigma_2(n)}\right) & \text{if } \sigma_2(n) \neq 0 \\ 1 & \text{if } \sigma_2(n) = 0 \end{cases} \quad (2)$$

We can deduce that $\beta \rightarrow 0$ when $\sigma_1 < \sigma_2$ and $\beta \rightarrow 1$ when $\sigma_1 > \sigma_2$. The correction term β , as in (2), is then used to determine the clean single output $s(n)$ of the procedure as

$$s(n) = \beta(n) \cdot s_2(n) + [1 - \beta(n)] \cdot s_1(n) \quad (3)$$

In the end, the output in (3) imitates the signal characterized by the lower energy in the 1 s examined interval.

2.3 Classification algorithms

In the proposed system, we utilize a classification technique to detect the presence (or absence) of a driver's stress condition in a given time interval during a car driving simulation. For this classification task, we utilize and compare six different supervised machine learning algorithms: an SVM classifier, a Naïve Bayes classifier, a k -Nearest Neighbors (k -NN) classifier, a Decision Tree (DT), an RF classifier, and an Artificial Neural Network (ANN). All these algorithms have been widely used in the literature for classification purposes. We also use two deep learning algorithms, in particular a 1D CNN and an LSTM network.

Our goal is to classify all of the time intervals in two possible classes: the stress class (or "1") and the non-stress class (or "0"). For this, we use the processed and cleaned SPR signal, and the ECG signal, with the latter one processed to remove the ectopic beats similarly to what is proposed in [35]. In addition, a further normalization preprocessing step is applied to both SPR and RR signals, because they can differ considerably among the individuals. This step consists of the standardization of each signal, based on the mean and variance of the preceding 5 minutes of the current signal value. We decided to process a 15 s long interval at a time, and we take a new interval every 5 s. In this way each interval covers the preceding one by 10 s. We do this to keep a reasonably low processing delay. The overall number of features extracted from these time intervals is nine: five from the SPR signal and four from the ECG signal. The five statistical features considered for the SPR signal

are: the interval variance, the energy, the mean absolute value, the mean absolute derivative and the max absolute derivative. The four statistical features from the ECG signal are: the mean value of RR intervals, the standard deviation of RR intervals (SDNN), the root mean square of subsequent RR interval differences (RMSSD) and the mean value of the Heart Rate [38]. We gather the features of the two signals together to construct the feature vectors. Each feature is then normalized in the range [0, 1]. All of the ML classifiers have been set up using Matlab (2017.a), except for the ANN that has been set up using Python. Hyperparameters are optimized automatically by the Matlab routine functions used for classification. In addition, for the SVM classifier, a Radial Basis Function (RBF) kernel has been used. The Bayesian optimization has also been used during the training procedure. For the ANN, we implement a four-layer perceptron trained with the back-propagation algorithm. We utilize Keras and the optimization package called “hyperopt” [39], allowing the use of the Bayesian optimization in this case as well. The optimization process also allows us to select the number of nodes in the two hidden layers (ranging from 16 to 128). The drop-out percentage has also been computed by the optimization package. Regarding the cross validation methodology, the 10-fold cross validation is used for all of the ML algorithms. For the ANN classifier we set the batch-size equal to 128.

The DL algorithms are also set up using Python. In detail, the architecture of the CNN used for classification has been inspired by the one proposed in [28], and its parameters are briefly reported in Table 1. The CNN contains two identical sequences of four operations: a convolutional layer, an Exponential Linear Unit (ELU), a batch normalisation layer and a max pooling layer. These are followed by a fully-connected layer (of 128 nodes), and the output layer where the softmax activation function is used. As anticipated, the inputs of the CNN are no longer the feature vectors, but the unprocessed samples of both the SPR and RR signals recorded from the subject at the same moment, in time intervals 15 s long. For each time interval, the values of the SPR and the RR signals must be arranged side by side. Since each 15 s long time interval is made of $N = 1500$ samples, with a sample rate of 100 Sa/s for both the SPR and RR signals, we obtain an input matrix with a size of 1500×2 , to be sent as input to the network. We proceed picking a new interval every 5 seconds and applying the same procedure. The labels corresponding to each interval are the same already derived for the ML algorithms. Other details about how the CNN model is defined are the following: we use the Adam optimizer, the learning rate has been set equal to 0.001, the categorical cross-entropy has been chosen as loss function, the number of epochs has been set equal to 20 and the batch-size to 64 (<https://keras.io>).

Another DL algorithm, i.e., an LSTM network, has been used in the experiments. This is a particular type of Recurrent Neural Network (RNN) that overcomes the issue of the short-term memory of the RNNs (also known as the vanishing gradient problem). The LSTM internal gates keep the relevant information in a long sequence of data, throwing away the non-relevant data, and this can help us recognizing stress with a long dynamic temporal behavior. The architecture of the implemented LSTM network is reported in Table 2. The same settings introduced before for the CNN (the learning rate, the number of epochs, etc.) have been used for the LSTM as well, with batch-size 128. Regarding this network, we

also consider a variant of it which uses the hyperopt optimization package (it will be denoted as LSTM2 in Section 4). In this case the number of nodes (ranging from 16 to 128) and the dropout value (varying between 0 and 1), for each layer, have been computed by the optimization package.

Table 1: Architecture of the implemented CNN

CNN	
Convolutional layer	Filters = 20, Kernel size = 10, Stride = 2
Exponential Linear Units (ELU)	Alpha = 0.1
Batch norm. + dropout (0.15)	
Max pooling	Pool size = 2, Stride = 2
Convolutional layer	Filters = 20, Kernel size = 10, Stride = 2
Exponential Linear Units (ELU)	Alpha = 0.1
Batch norm. + dropout (0.15)	
Max pooling	Pool size = 2, Stride = 2
Flatten	
Dense + dropout (0.5)	Nodes = 128, Activation = ReLU
Dense	Nodes = 2, Activation = Softmax

We employ the “leave-one-person-out” technique for all of the learning algorithms, such that the training of each classifier is performed using the data of all individuals, except one, on which the classifiers will be tested. This process is carried out for each individual, and in the end the overall system performance is determined by computing the average of the results originated from all of the individuals.

Table 2: Architecture of the implemented LSTM network

LSTM	
LSTM layer	Nodes = 128, Activation = ReLU
Dropout	0.4
LSTM layer	Nodes = 128, Activation = ReLU
Dropout	0.4
Dense	Nodes = 2, Activation = Softmax

3 Experimental setup

The experiments, as introduced above, are carried out in a firm that develops professional driving simulators, both in hardware and software. For our tests, we use a dynamic driving simulator that allows the reproduction of the same movements of a real car. In fact, it provides motion feedback to the drivers through a nine degrees-of-freedom moving platform. We enrolled 16 healthy individuals, 22-47 years old, some from the University of Udine and some from the University of Padua. The subjects were asked to drive on a 28 km long straight highway track, with four stress-inducing obstacles demanding some concentration to get over. These four obstacles are shown in Figure 3. From left to right and from top to bottom, they are: Sponsor block (from right to left), Tire labyrinth, Double lane change (right to left) and Sponsor block (from left to right).

All subjects prepare themselves before entering the simulator. They wear the sensing devices both on chest and hands. They gave consent having their physiological signals recorded, and the test were performed according to the principles of the Declaration of Helsinki.

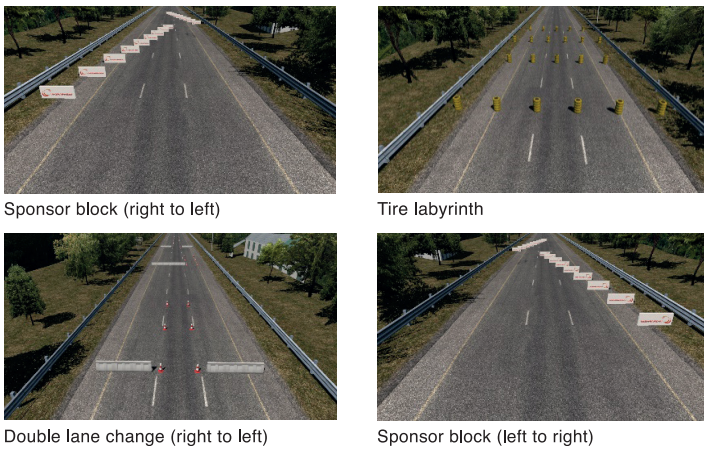


Figure 3: The four stress-inducing events located along the road.

Figure 4 displays the cleaned SPR signal, after the MA removal block, and the instantaneous HR signal of a subject (derived from the RR signal) during the simulation. The start of an event is signaled with a vertical line with a circle marker, and the end of an event with a vertical line with a square marker. We also assume an event starts when the obstacle enters the field of view of the subject (and this happens approaching a distance of 800 meters from the obstacle), and ends 30 seconds after surpassing the obstacle. Passing each obstacle takes approximately 65 seconds. The peculiar stress peaks shape in the SPR signal during the stress event can be seen evidently in the figure, just as the distinctive rise in Heart Rate.

4 Experimental Results

The classifiers, which allow the recognition of stress in each 15 s long interval, are built using the cleaned SPR signal and the ECG signal of each subject. For the traditional ML algorithms the features of the signals are extracted from each interval. For the DL algorithms there are no features to be extracted, and the raw signals themselves are considered for each interval. In all cases, the intervals overlap each other by 10 s as already specified.

We assume that all the intervals happening in a stretch of the road where obstacles are present, to have the value “1”, with stress, and that all the intervals happening in a stretch of the road without obstacles to have the value “0”, with no stress. As already mentioned, we employ the “leave-one-person-out” method in the classification process. So, for the training, we examine 15 subjects at a time out of the 16 in total. In this way, we obtain a total number of intervals ranging from 1176 to 1209, for both stress and non-stress classes. The total number of intervals for training is not fixed because each individual overcomes the obstacles with slightly different timing and speed. During the test we obtain approximately 187 intervals to analyze for each individual (this is an average value taking all of the subjects into consideration), where the majority of them, about two thirds, belong to the non-stress class.

According to the assumed ground-truth, the True Positives (TP), the False Negatives (FN), the True Negatives (TN) and the False Positives (FP) can be computed, as well as the performance indicators

like Accuracy, Sensitivity and Specificity:

$$\text{Accuracy (\%)} = \frac{TP+TN}{TP+TN+FP+FN} \cdot 100 \quad (4)$$

$$\text{Sensitivity (\%)} = \frac{TP}{TP+FN} \cdot 100 \quad (5)$$

$$\text{Specificity (\%)} = \frac{TN}{FP+TN} \cdot 100 \quad (6)$$

Due to the fact that our test data are unbalanced (in fact we have less stress intervals than non-stress intervals), two additional indicators are also important, the Balanced Accuracy (BA) and the Geometric Mean (GM), defined as

$$\text{BA (\%)} = \frac{1}{2} \left(\frac{TP}{TP+FN} + \frac{TN}{FP+TN} \right) \cdot 100 \quad (7)$$

$$\text{GM (\%)} = \sqrt{\frac{TP}{TP+FN} \cdot \frac{TN}{FP+TN}} \cdot 100 \quad (8)$$

Table 3 presents the overall performance of our system (MEAN \pm STD), calculated by averaging the results of the 16 subjects, for all of the considered ML algorithms. The accuracy and balanced accuracy values among the classifiers are close, even if the ANN classifier is slightly better than the other ones in terms of accuracy, and the RF in terms of balanced accuracy. Moreover, the sensitivity of the RF is higher than the others, as it is the specificity of the Naïve Bayes classifier. The GM values are similar for all of the ML algorithms, with the RF classifier still performing marginally better than the others.

As introduced before, as a first attempt to use a deep learning algorithm to recognize stress and non-stress episodes in our scenario, we set up a 1D CNN and an LSTM. We compute the overall performance of the CNN and the LSTM using the test set, and we obtain the values reported in Table 3. So, comparing the results of the DL algorithms with the ones obtained applying the aforementioned ML algorithms, we can notice that the CNN architecture performs similarly to the other ML architectures. In particular, the accuracy of the ANN and the specificity of the Naïve Bayes classifier exceed the values corresponding to the CNN classifier, whereas the CNN sensitivity is better than that obtained with the other ML classifiers. However, when looking at the LSTM, we can see that it works well in detecting the stress episodes, outperforming all the other algorithms in terms of accuracy, sensitivity, balance accuracy and geometric mean. The specificity of the LSTM, instead, is lower than the Naïve Bayes one. The variant LSTM2 performs even better, and all of the performance indicators are higher than the other DL and ML algorithms.

The overall performance of the system looks favourable. It should be noted that some of our hypotheses are critical and can affect the performance indicators. The first hypothesis we make is that all the time intervals happening during the overcoming of an obstacle by a driver should originate stress, even if it is possible that the driver is not constantly stressed during the manoeuvre. The second one is that a driver is not stressed when not overcoming an obstacle. This might not be true at all times, since a driver can be in a condition of stress during those time intervals for causes unknown to us.

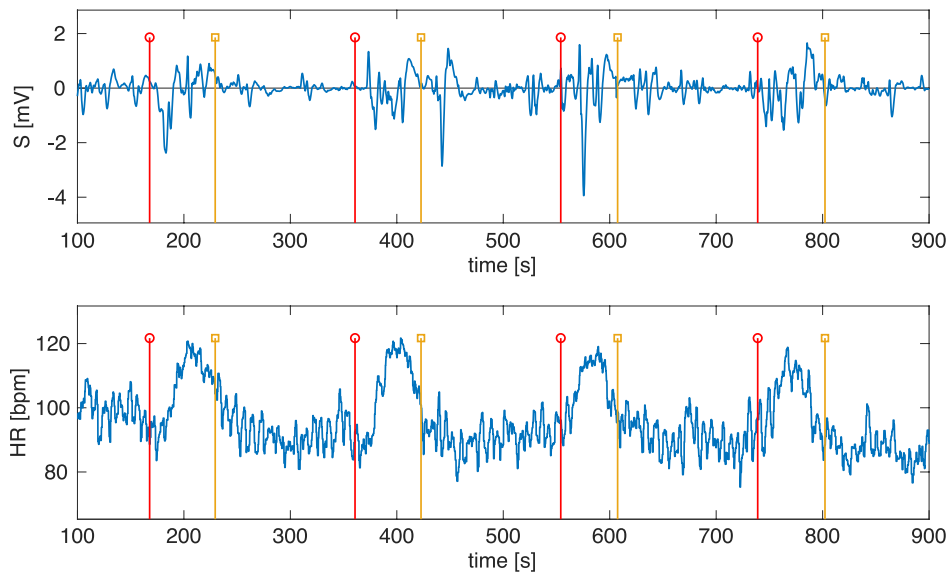


Figure 4: SPR and HR signals (at the top and the bottom, respectively) for a subject during the driving simulation. Each event onset is evidenced with a red vertical line with a circle marker, and each event offset with a yellow vertical line with a square marker.

Table 3: Performance of all the supervised learning algorithms

	Accuracy (%)	Sensitivity (%)	Specificity (%)	BA (%)	GM (%)
SVM	72.86 ± 8.77	86.69 ± 7.79	66.46 ± 13.02	76.57 ± 6.91	75.40 ± 7.85
Naïve Bayes	74.04 ± 6.53	82.83 ± 8.04	70.27 ± 10.85	76.55 ± 4.65	75.81 ± 5.01
k-NN	72.07 ± 6.01	86.60 ± 8.30	65.99 ± 8.77	76.30 ± 4.96	75.26 ± 5.23
DT	70.76 ± 8.88	85.16 ± 6.61	64.80 ± 13.13	74.98 ± 5.98	73.69 ± 7.20
RF	74.25 ± 6.17	87.05 ± 7.60	68.83 ± 8.83	77.94 ± 5.03	77.11 ± 5.35
ANN	74.45 ± 8.65	86.19 ± 8.13	69.00 ± 13.38	77.59 ± 6.50	76.56 ± 7.30
CNN	73.97 ± 6.23	87.99 ± 6.40	67.76 ± 8.30	77.87 ± 4.82	76.99 ± 5.27
LSTM	82.56 ± 3.04	88.61 ± 7.55	69.79 ± 17.61	79.20 ± 5.96	77.46 ± 9.23
LSTM2	88.02 ± 0.97	93.60 ± 1.92	76.26 ± 4.22	84.93 ± 1.55	84.43 ± 1.78

Another possible critical aspect of our hypotheses is that part of the signal responses may be due to the increased physical effort during the events. However, in some previous works, e.g., [40], we observed that the intensities of the signals during designated stress events were significant, and could not be confused with the levels recorded during non-stress intervals, even if a comparable physical activity was required for driving. In [41], we also evaluated a car driver’s stress condition in a simulated autonomous driving scenario. In that experiment, we submitted questionnaires to the drivers and evaluated other parameters such as the User Experience (UX) measure and the NASA Task Load Index. Questionnaires supported our findings derived by analyzing the SPR and ECG signals (see also [42]).

To account for these considerations, and enhance the stress recognition ability, we apply a re-label step to the output of the classifiers (see Figure 5). To do so, we assume stress is happening if a minimum of four consecutive intervals are marked as positive (thus covering a time span of 30 s). In the end, the blocks of four or more consecutive positive intervals are kept, whereas blocks of up to three consecutive positive intervals are instead re-labelled as negative.

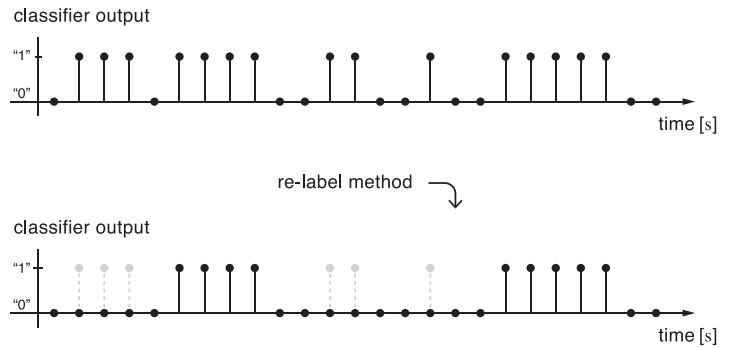


Figure 5: The re-label method applied to the output of each classifier.

In Table 4 the performance indicators applying the re-label method are presented. The values are slightly higher than the ones in Table 3. On the other hand, sensitivity is a little smaller for all of the classifiers, since we noticed that a number of stress intervals, which were correctly detected before, are now discarded by the re-label step. The performance indicators of the LSTM2 are clearly higher, when compared to the other algorithms, in this scenario as well.

Table 4: Performance of all the supervised learning techniques when the re-label method is applied to the output of each classifier

	Accuracy (%)	Sensitivity (%)	Specificity (%)	Balanced Accuracy (%)	Geometric Mean (%)
SVM	77.61 ± 8.14	85.11 ± 8.82	74.05 ± 11.67	79.58 ± 6.89	79.00 ± 7.30
Naïve Bayes	77.39 ± 6.69	82.83 ± 8.04	70.27 ± 10.85	76.55 ± 4.65	75.81 ± 5.01
k-NN	75.86 ± 5.84	83.85 ± 11.04	72.61 ± 8.95	78.23 ± 5.47	77.56 ± 5.63
DT	76.17 ± 7.87	82.42 ± 8.09	73.69 ± 11.41	78.06 ± 5.85	77.50 ± 6.16
RF	77.15 ± 6.39	86.16 ± 7.42	73.32 ± 8.89	79.74 ± 5.35	79.22 ± 5.59
ANN	78.17 ± 9.13	84.73 ± 9.72	75.16 ± 14.23	79.94 ± 6.88	79.15 ± 7.50
CNN	79.47 ± 6.45	85.31 ± 8.33	77.09 ± 8.48	81.20 ± 5.79	80.87 ± 5.90
LSTM	83.21 ± 3.27	88.23 ± 7.78	72.62 ± 17.39	80.42 ± 5.94	78.91 ± 9.00
LSTM2	88.13 ± 0.87	93.30 ± 1.92	77.20 ± 4.05	85.25 ± 1.42	84.83 ± 1.60

5 Conclusions

We presented a scheme to recognize individuals' stress while driving in a car driving simulator. The proposed system uses relatively simple sensors, with an acceptably intrusive setup. Due to the short signal block analysis, it can be used nearly in real time and can discriminate stressful events of short duration. In our system, we record two SPR signals for each individual, one for each hand, and the ECG signal. We then apply an MA removal algorithm to remove the artifacts that could appear in the SPR signals during the drive, so that the output is a single cleaned SPR signal. Some features are extracted from both the cleaned SPR signal and the ECG, and sent as input to an ML algorithm. The raw signals are instead directly used as input to DL algorithms. In particular, we evaluate the performance of an SVM classifier, a Naïve Bayes classifier, a k -NN classifier, a DT, an RF classifier and an ANN. The performance of all the ML classifiers are similar one to each other, however, the ANN classifier performs better than the others in terms of accuracy. The RF classifier outperforms the others in terms of sensitivity, balanced accuracy and geometric mean, while the Naïve Bayes classifier in terms of specificity. We also report the results obtained with a 1D CNN and an LSTM. The CNN outperforms the other ML classifiers in terms of sensitivity. The LSTM, when some parameters are selected by the optimization procedure, outperforms all the other learning algorithms. A re-labelling procedure is also considered, and the performance indicators of all of the algorithms are compared again in this case. We obtain the highest accuracy, equal to 88.13%, using the optimized LSTM.

Some critical aspects of the proposed experiment are due to the assumption that stress is present only during the arranged events, and that part of the signal responses may be due to the increased physical effort while driving through the events. We have observed that these issues still allow us to interpret the results with a good level of confidence. We conclude that the overall performance of our system is encouraging, showing that by using the SPR and ECG signals we can identify the stress in drivers, deriving from some demanding driving activities, with acceptable accuracy.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment 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.

References

- [1] M. Navon, O. Taubman-Ben-Ari, "Driven by emotions: The association between emotion regulation, forgivingness, and driving styles," *Transportation Research Part F: Traffic Psychology and Behaviour*, **65**, 1–9, 2019, doi:10.1016/j.trf.2019.07.005.
- [2] C. D. Katsis, N. Katertsidis, G. Ganiatsas, D. I. Fotiadis, "Toward emotion recognition in car-racing drivers: A biosignal processing approach," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, **38**(3), 502–512, 2008, doi:10.1109/TSMCA.2008.918624.
- [3] W. Boucsein, *Electrodermal Activity*, Springer, 2012.
- [4] S. Greene, H. Thapliyal, A. Caban-Holt, "A survey of affective computing for stress detection: Evaluating technologies in stress detection for better health," *IEEE Consumer Electronics Magazine*, **5**(4), 44–56, 2016, doi:10.1109/MCE.2016.2590178.
- [5] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, J.-H. Chen, "EEG-based emotion recognition in music listening," *IEEE Transactions on Biomedical Engineering*, **57**(7), 1798–1806, 2010, doi:10.1109/TBME.2010.2048568.
- [6] J. Zhang, W. Wen, F. Huang, G. Liu, "Recognition of Real-Scene Stress in Examination with Heart Rate Features," in *2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, Hangzhou, 2017, doi:10.1109/IHMSC.2017.13.
- [7] H. Becker, J. Fleureau, P. Guillotel, F. Wendling, I. Merlet, L. Albera, "Emotion recognition based on high-resolution EEG recordings and reconstructed brain sources," *IEEE Transactions on Affective Computing*, **11**(2), 244–257, 2020, doi:10.1109/TAFFC.2017.2768030.
- [8] Z. Qin, M. Li, L. Huang, Y. Zhao, "Stress level evaluation using BP neural network based on time-frequency analysis of HRV," in *2017 IEEE International Conference on Mechatronics and Automation (ICMA)*, Takamatsu, 2017, doi:10.1109/ICMA.2017.8016090.
- [9] M. Moghimi, R. Stone, P. Rotshtein, "Affective recognition in dynamic and interactive virtual environments," *IEEE Transactions on Affective Computing*, **11**(1), 45–62, 2020, doi:10.1109/TAFFC.2017.2764896.
- [10] C. Maaoui, A. Pruski, F. Abdat, "Emotion Recognition for Human-Machine Communication," in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nice, 2008, doi:10.1109/IROS.2008.4650870.
- [11] G. Giannakakis, E. Trivizakis, M. Tsiknakis, K. Marias, "A novel multi-kernel 1D convolutional neural network for stress recognition from ECG," in *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, Cambridge, United Kingdom, 2019, doi:10.1109/ACIIW.2019.8925020.
- [12] K. T. Chui, M. D. Lytras, and R. W. Liu, "A generic design of driver drowsiness and stress recognition using MOGA optimized deep MKL-SVM," *Sensors*, **20**(5):1474. 2020, doi:10.3390/s20051474.

- [13] L. Lombardi, F. Marcolin, Psychological Stress Detection by 2D and 3D Facial Image Processing, *Progresses in Artificial Intelligence and Neural Systems*, Springer, 2020.
- [14] M. T. Uddin, S. Canavan, "Synthesizing Physiological and Motion Data for Stress and Meditation Detection," in 2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), Cambridge, United Kingdom, 2019, doi:10.1109/ACIIW.2019.8925245.
- [15] A. Oskooei, S. Mai Chau, J. Weiss, A. Sridhar, M. Rodríguez Martínez, B. Michel, "DeStress: deep learning for unsupervised identification of mental stress in firefighters from heart-rate variability (HRV) data," arXiv:1911.13213v1, 2019.
- [16] B. Hwang, J. You, T. Vaessen, I. Myin-Germeys, C. Park, B.-T. Zhang, "Deep ECGNet: An optimal deep learning framework for monitoring mental stress using ultra short-term ECG signals," *Telemedicine and e-Health*, **24**(10), 753–772, doi:10.1089/tmj.2017.0250.
- [17] F. Al Machot, A. Elmachot, M. Ali, E. Al Machot, K. Kyamakya, "A deep-learning model for subject-independent human emotion recognition using electrodermal activity sensors," *Sensors*, **19**(7):1659, 2019, doi:10.3390/s19071659.
- [18] T. Kundu, C. Saravanan, "Advancements and recent trends in emotion recognition using facial image analysis and machine learning models," in 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), Mysuru, 2017, doi:10.1109/ICEECCOT.2017.8284512.
- [19] Y. Ishii, T. Ogitsu, H. Takemura, H. Mizoguchi, "Real-time eyelid open/closed state recognition based on HLAC towards driver drowsiness detection," in 2014 IEEE International Conference on Robotics and Biomimetics (ROBIO), Bali, 2014, doi:10.1109/ROBIO.2014.7090707.
- [20] D. S. Lee, T. W. Chung, B. G. Lee, "Stress events detection of driver by wearable glove system," *IEEE Sensors Journal*, **17**(1), 194–204, 2017, doi:10.1109/JSEN.2016.2625323.
- [21] Y. J. Yu, Z. Yang, B.-S. Oh, Y. K. Yeo, Q. Liu, G.-B. Huang, Z. Lin, "Investigation on driver stress utilizing ECG signals with on-board navigation systems in use," in 2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV), Phuket, 2016, doi:10.1109/ICARCV.2016.7838780.
- [22] D. Tomoi, W. Wen, H. Yamakawa, S. Hamasaki, K. Takakusaki, Q. An, Y. Tamura, A. Yamashita, H. Asama, "Estimation of stress during car race with factor analysis," in 2015 International Symposium on Micro-NanoMechatronics and Human Science (MHS), Nagoya, 2015, doi:10.1109/MHS.2015.7438258.
- [23] G. Rigas, Y. Goletsis, D. I. Fotiadis, "Real-time driver's stress event detection," *IEEE Transactions on Intelligent Transportation Systems*, **13**(1), 221–234, 2012, doi:10.1109/TITS.2011.2168215.
- [24] J. A. Healey, R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on Intelligent Transportation Systems*, **6**(2), 156–166, 2005, doi:10.1109/TITS.2005.848368.
- [25] D.-W. Koh, S.-G. Lee, "An evaluation method of safe driving for senior adults using ECG signals," *Sensors*, **19**(12):2828, 2019, doi:10.3390/s19122828.
- [26] Z. Halim, M. Rehan, "On identification of driving-induced stress using electroencephalogram signals: A framework based on wearable safety-critical scheme and machine learning," *Information Fusion*, **53**, 66–79, 2020, doi:10.1016/j.inffus.2019.06.006.
- [27] H.-M. Cho, H. Park, S.-Y. Dong, I. Youn, "Ambulatory and laboratory stress detection based on raw electrocardiogram signals using a convolutional neural network," *Sensors*, **19**(20):4408, 2019, doi:10.3390/s19204408.
- [28] M. N. Rastgoo, B. Nakisa, F. Maire, A. Rakotonirainy, V. Chandran, "Automatic driver stress level classification using multimodal deep learning," *Expert Systems with Applications*, **138**:112793, 2019, doi:10.1016/j.eswa.2019.07.010.
- [29] L. G. Hernández, O. M. Mozos, J. M. Ferrández and J. M. Antelis, "EEG-based detection of braking intention under different car driving conditions," *Frontiers in Neuroinformatics*, **12**:29, 2018, doi:10.3389/fninf.2018.00029.
- [30] P. Zontone, A. Affanni, R. Bernardini, A. Piras, R. Rinaldo, F. Formaggia, D. Minen, M. Minen, C. Savorgnan, "Car driver's sympathetic reaction detection through electrodermal activity and electrocardiogram," *IEEE Transactions on Biomedical Engineering*, 1–12, 2020, doi:10.1109/tbme.2020.2987168.
- [31] A. Affanni, G. Chiorboli, "Design and characterization of a real-time, wearable, endosomatic electrodermal system," *Measurement*, **75**, 111–121, 2015, doi:10.1016/j.measurement.2015.07.047.
- [32] A. Affanni, "Wearable instrument to measure simultaneously cardiac and electrodermal activities," in 2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Benevento, Italy, 2016, doi:10.1109/MeMeA.2016.7533749.
- [33] J. Pan, W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Transactions on Biomedical Engineering*, **BME-32**(3), 230–236, 1985, doi:10.1109/TBME.1985.325532.
- [34] W. Chen, G. Liu, S. Su, Q. Jiang, and H. Nguyen, "A CHF detection method based on deep learning with RR intervals," 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Seogwipo, 2017, doi:10.1109/EMBC.2017.8037578.
- [35] M. P. Tarvainen, J.-P. Niskanen, J. A. Lipponen, P. O. Ranta-Aho, P. A. Karjalainen, "Kubios HRV – Heart rate variability analysis software," *Computer Methods and Programs in Biomedicine*, **113**(1), 210–220, 2014, doi:10.1016/j.cmpb.2013.07.024.
- [36] A. Affanni, A. Piras, R. Rinaldo, P. Zontone, "Dual channel Electrodermal activity sensor for motion artifact removal in car drivers' stress detection," in 2019 IEEE Sensors Applications Symposium (SAS), Sophia Antipolis, France, 2019, doi:10.1109/SAS.2019.8706023.
- [37] P. Zontone, A. Affanni, R. Bernardini, A. Piras, R. Rinaldo, "Low-complexity classification algorithm to identify drivers' stress using electrodermal activity (EDA) measurements," *Biomedical Engineering and Computational Intelligence*, **20**, 25–33, 2020, doi:10.1007/978-3-030-21726-6_3.
- [38] M. Malik, "Heart rate variability. Standards of measurement, physiological interpretation, and clinical use," *European Heart Journal*, **17**(3), 354–381, 1996, doi:10.1161/01.CIR.93.5.1043.
- [39] J. Bergstra, D. Yamins, D. D. Cox, "Hyperopt: A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms," in 2013 12th Python in Science Conference (SCIPY), Austin, Texas, 2013, doi:10.1088/1749-4699/8/1/014008.
- [40] A. Affanni, R. Bernardini, A. Piras, R. Rinaldo, P. Zontone, "Driver's stress detection using skin potential response signals," *Measurement*, **122**, 264–274, 2018, doi:10.1016/j.measurement.2018.03.040.
- [41] P. Zontone, A. Affanni, R. Bernardini, L. Del Linz, A. Piras, R. Rinaldo, "Stress evaluation in simulated autonomous and manual driving through the analysis of skin potential response and electrocardiogram signals," *Sensors*, **20**(9):2494, 2020, doi:10.3390/s20092494.
- [42] A. Affanni, P. Zontone, R. Fenici, D. Brisinda, D. Bacchin, L. Gamberini, P. Pluchino, M. Bruschetta, C. Savorgnan, F. Formaggia et al., "Assisted / autonomous vs. human driving assessment on the DiM driving simulator using objective / subjective characterization," in 2019 10th International Munich Chassis, Munich, Germany, 2019, Pfeffer P. (eds), Springer Vieweg, Wiesbaden, 307–321, 2020, doi:10.1007/978-3-658-26435-2_23.
- [43] P. Zontone, A. Affanni, R. Bernardini, A. Piras, R. Rinaldo, "Stress Detection Through Electrodermal Activity (EDA) and Electrocardiogram (ECG) Analysis in Car Drivers," in 2019 European Signal Processing Conference (EUSIPCO), A Coruña, Spain, 2019, doi:10.23919/EUSIPCO.2019.8902631.