Convolutional Neural Networks Using Scalograms for Stress Recognition in Drivers

Pamela Zontone, Antonio Affanni, Alessandro Piras, Roberto Rinaldo
Polytechnic Department of Engineering and Architecture
University of Udine, Udine, Italy
{pamela.zontone, antonio.affanni, roberto.rinaldo}@uniud.it, piras.alessandro@spes.uniud.it

***Abstract*—In this paper we present a system which allows the detection of stress in drivers by analyzing a two-dimensional representation of their electrodermal activity Skin Potential Response (SPR) signal, and their electrocardiogram signal. Signals were logged during a simulated drive, in an experiment carried out in a company using a professional car driving simulator. Subjects had to overcome some stress-inducing events located at specific positions during the drive. The acquired SPR and heart rate signals are analyzed with scalogram plots, in order to obtain a time-frequency representation of the signals. The 2D scalogram representation is segmented into images, associated to short time segments, which are classified using a Convolutional Neural Network architecture. We show that the use of scalograms can allow the system to perform well in distinguishing among stress and non-stress situations, achieving a 91.78% accuracy. The same system was tested on real driving data available from a public dataset, achieving a 99.24% accuracy.**

***Index Terms*—Stress Detection, Electrodermal Activity, Heart Rate, Scalogram, Convolutional Neural Network**

I. Introduction and Related Work

There are many emotional conditions that can be assessed in car drivers, like fatigue, drowsiness, and stress. In particular, the driver's mental state affects both individual well-being and public road safety [1]. Aggressiveness in driving behaviour has also been increasing during the COVID-19 pandemic [2]. To provide tools to help recognize the onset of potentially dangerous stress situations, several works in the literature have analyzed physiological and behavioural measures such as Electrodermal Activity (EDA), Electroencephalogram (EEG), Electrooculogram (EOG), Electromyogram (EMG), and Electrocardiogram (ECG) signals, facial expressions, and body postures, or combined physiological and vehicle data [3]–[6].

The most prominent approaches are based on Machine Learning (ML) and Deep Learning (DL) techniques. In [7], for example, various ML techniques, such as the Decision Tree, k-Nearest Neighbors, and Naïve Bayes classifiers are compared to evaluate their ability to reveal stress, based on the analysis of the ECG, EMG, respiration rate, and EDA Galvanic Skin Response (GSR) signals. DL architectures are instead proposed in [8] to detect drivers' emotional state as well as their behavioral states (e.g., talking to the passenger or eating). ML algorithms have also been used on EDA and ECG data recorded from subjects while driving on a road with stress-inducing obstacles and in different traffic conditions [9], [10]. The performance results obtained in a scenario where car handling setups change among different tests, still applying ML algorithms, are instead presented in [11], [12].

The 1D signals, which are typically acquired for stress detection, can be also converted into two-dimensional plots. An example of the use of 2D representations is described in [13], where 2D Continuous Recurrence Plots are originated from both hand and foot GSRs, and Heart Rate (HR) signals, and then used as input to multimodal Convolutional Neural Networks (CNNs) for detecting stress in subjects while driving. An overall accuracy of 95.67%, considering 30 s segments, has been achieved in discerning among low-level stress and high-level stress in car drivers.

The main contribution of this paper is the investigation of a novel system where scalograms are used as two-dimensional representations of the input physiological signals, i.e., the EDA Skin Potential Response (SPR) and ECG signals. Scalograms provide localized information about the characteristics of the input signal in time and at different scales, creating a 2D representation of the absolute value of the Continuous Wavelet Transform (CWT) [14], [15], and they have been recently proposed for the analysis of the ECG signal in driving scenarios. In this paper, we extend the work of [16], [17] and propose the use of scalograms as a representation tool of SPR signals. In particular, scalogram plots of both SPR and HR signals are split into images corresponding to signal segments of small duration, and classified via a multidimensional CNN architecture. Two case studies are considered: one using the physiological signals used in [18], logged from individuals driving in a professional driving simulator located in the VI-grade company, and the other using a public dataset of signals logged from individuals while driving in the real world (PhysioNet Stress Recognition in Automobile Drivers (SRAD) dataset [19]). Experimental results are promising, with high accuracy in the two datasets, as it will be discussed in Section III.

II. Methods

The block scheme of the proposed system is reported in Figure 1. As mentioned before, we consider the data coming from two different datasets that will be described in detail in the next section. The signals belonging to the different datasets are preprocessed, and the scalograms from the resulting signals are computed. These 2D plots are then split into smaller plots, corresponding to short time interval segments. The segmented

scalogram images are then sent as input to a deep learning architecture for classification into "stress" and "non-stress", or "high-level" or "low-level" stress intervals, depending on the considered experiment, as we will describe below in detail.

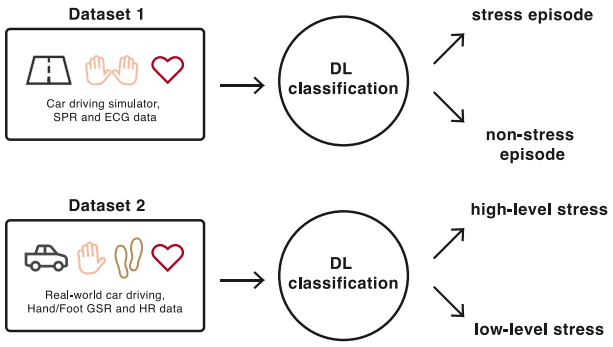

Fig. 1. Fundamental blocks of our system.

A. The datasets

Dataset 1. The first dataset used in our work has been described thoroughly in [18]. In summary, a total of 18 subjects have been tested. They were asked to drive in a driving simulator for about 40 minutes along a highway, trying to overcome 12 obstacles arranged in predefined positions along the course. The 12 obstacles were as follows: 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, Stop. We recorded the data from sensors located in different positions on the body of each subject, i.e., on the two hands for SPR recordings, and on the chest for ECG recordings. It was defined that a stress event started when an obstacle became visible, i.e., 800 meters before the obstacle, and finished 40 seconds after the obstacle. Therefore, for each subject, we were aware of the place and the duration of the stress-evoking events, so we could assign to the SPR and ECG signal segments a label equal to "1" (with stress) to all of the segments belonging to, or intersecting with, a stress event, and a label equal to "0" (without stress) to all of the other segments.

Dataset 2. The second dataset is available in PhysioNet and is called Stress Recognition in Automobile Drivers (SRAD) [3], [19]. It is composed of multiple physiological signals recorded from 17 subjects driving in a real-world scenario, on different road routes, such as a highway and a city route. Various physiological signals are provided, along with an additional signal, denoted as "marker", which allows the identification of both the onset and offset of each session. As in [13], in this work we only focus on the Foot Galvanic Skin Response (FGSR), Hand Galvanic Skin Response (HGSR), and Heart Rate (HR) signals. As a consequence, we only consider the signals belonging to nine individuals, i.e., subjects 06, 07, 08, 09, 10, 11, 12, 15, and 16. We take into account the segments of the signals belonging to the rest sessions and representing a low-level stress (giving them a label equal to "0"), and the segments of the signals belonging to the city sessions and representing a high-level stress (giving them a label equal to "1"). A binary classification can be performed using both datasets, giving us the chance to identify the various stress situations.

B. Scalogram generation

A scalogram represents the absolute value of the CWT, which is able to give us a time-frequency characterization of the signal [14], [15]. The corresponding 2D plots allow the localization of the signal characteristics both in time and frequency. This is crucial for our application because we are interested in time-localized event signals with peculiar frequency domain characteristics. To create the scalogram, we use standard Matlab routines and the generalized Morse analytic wavelet [10], [20], setting the γ and β parameters, which allow the definition of the wavelet shape, as $\gamma = 3$ and $\beta \cdot \gamma = 60$.

Figures 2 and 3 show an example of SPR and HR scalograms generated for subjects belonging to dataset 1. For illustration purposes, the onset and offset of each obstacle is indicated in the figure using a yellow square line, where the onset and offset trigger the line values to 1 and 0, respectively. It is clearly evident how the scalograms can reveal the subject's stress responses when overcoming the 12 obstacles.

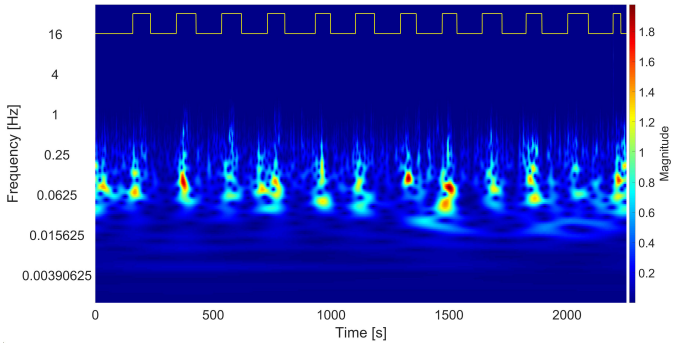

Fig. 2. Example of an SPR scalogram for a subject belonging to dataset 1.

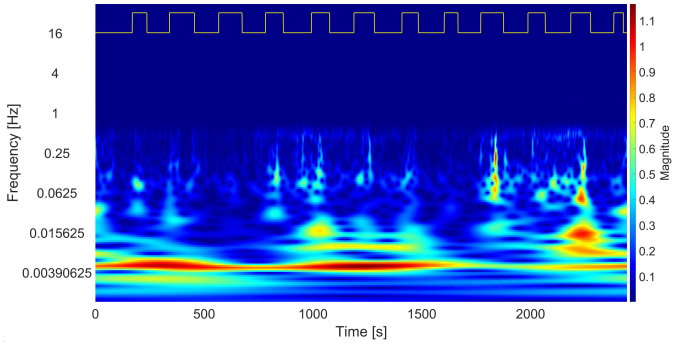

Fig. 3. Example of an HR scalogram for a subject belonging to dataset 1.

Regarding dataset 1, two SPR signals are acquired from the driver's hands to obtain a single SPR signal by using a motion

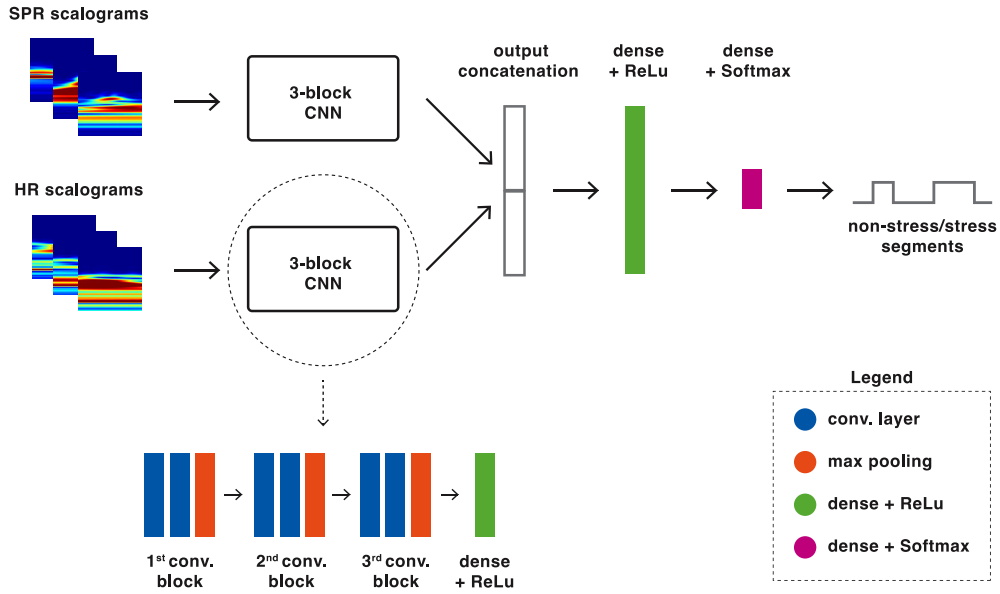

Fig. 4. DL architecture used in our system (with dataset 1).

artifact removal algorithm, as explained in [18]. The HR and cleaned SPR signals are normalized to have zero mean and unit variance before computing the scalogram. The signals are sampled at 100 sa/s. Regarding dataset 2, the FGSR, HGSR, and HR signals are available at a sampling rate of 15.5 Hz. After normalization, we apply a 3rd-order median filter to smooth the signals and reduce possible artifacts, as in [13].

The scalogram plots are then segmented into smaller images (or sub-images), corresponding to 15 s time intervals [18] and 30 s time intervals [13] for dataset 1 and 2, respectively, with a fixed overlap in both cases. The resulting images, however, are subsampled along both x and y axes, to have a final dimension of 224x224x3 (considering the three RGB components along the z-axis).

C. Deep learning architecture

The scalogram sub-images, corresponding to the different time intervals, are used as input to a deep learning architecture which uses a combination of multiple convolutional neural networks, with a dedicated CNN for each signal scalogram representation (see Figure 4). In particular, for dataset 1, two scalogram sub-images corresponding to each segment of the SPR and HR signals, respectively, are the input of the neural network, whose output is a binary level corresponding to the "stress" and "non-stress" classes. For dataset 2, the input to the network consists of three scalogram sub-images, corresponding to the FGSR, HGSR, and HR signal segments, again with a binary output indicating "high-level" or "low-level" stress.

The proposed architecture falls within the VGG model specifications [21]. We will denote the whole network as "multi-VGG-CNN". We tried different architectures with 1, 2, and 3 CNN convolutional blocks, finally choosing a 3-block architecture, which in our tests outperformed the other solutions. We set each convolutional layer, in each block, to use the ReLu activation function, with each block to be followed by a max pooling operation. Dropout layers have also been included to reduce possible overfitting on training data. Finally the outputs of the CNNs are concatenated to provide the final classification. The architecture and the parameters of the multi-VGG-CNN model are shown in Table 1.

TABLE I
IMPLEMENTED MULTI-VGG-CNN ARCHITECTURE WITH RELATED PARAMETERS: N=2 WITH DATASET 1, AND N=3 WITH DATASET 2

multi-VGG-CNN architecture	
Input 1	
Conv. layer 2D	Filters = 32, Kernel size = (3, 3), Activation = ReLu
Conv. layer 2D	Filters = 32, Kernel size = (3, 3), Activation = ReLu
Max pooling 2D	Pool size = (2, 2)
Dropout	0.2
Conv. layer 2D	Filters = 64, Kernel size = (3, 3), Activation = ReLu
Conv. layer 2D	Filters = 64, Kernel size = (3, 3), Activation = ReLu
Max pooling 2D	Pool size = (2, 2)
Dropout	0.2
Conv. layer 2D	Filters = 128, Kernel size = (3, 3), Activation = ReLu
Conv. layer 2D	Filters = 128, Kernel size = (3, 3), Activation = ReLu
Max pooling 2D	Pool size = (2, 2)
Dropout	0.2
Flatten	
Dense	Nodes = 128, Activation = ReLu
Dropout	0.2
Output 1	
...	
Input N	
⋮	
Output N	
Concatenate	**Output 1, ⋯, Output N**
Dense	Nodes = 128*N, Activation = ReLu
Dropout	0.2
Dense	Nodes = 2, Activation = Softmax
Final Output	

TABLE II
PERFORMANCE RESULTS OBTAINED IN CASE 1 AND CASE 2, CONSIDERING BOTH DATASETS

case 1					
	Accuracy (%)	Sensitivity (%)	Specificity (%)	BA (%)	GM (%)
Dataset 1	91.78	89.32	94.30	91.81	91.78
Dataset 2	99.24	98.92	99.52	99.22	99.22
case 2					
	Accuracy (%)	Sensitivity (%)	Specificity (%)	BA (%)	GM (%)
Dataset 1 (mean $\pm$ std)	84.35 $\pm$ 3.08	85.41 $\pm$ 4.32	82.71 $\pm$ 7.10	84.06 $\pm$ 3.40	83.91 $\pm$ 3.46
Dataset 2 (mean $\pm$ std)	96.64 $\pm$ 2.17	94.59 $\pm$ 4.97	98.07 $\pm$ 2.09	96.33 $\pm$ 2.45	96.27 $\pm$ 2.53

We use two different approaches to compute the final performance indicators, considering the two different datasets. In the first approach (also denoted as "case 1" from now on), which is the most typical, for training we extract some random scalograms from each class (80% of the data), while the remaining samples are used for testing. In the second approach, that will be denoted as "case 2", we consider for the training process all the data coming from all the subjects, except one, which will be the one on which the algorithm will be tested. In this case, however, we include 20% of the scalograms of the excluded subject in the training set, whereas testing is carried out on the unseen 80% samples. In a practical scenario, this will require that some samples, coming from the subject under test, are acquired in advance and included in the training set. Indeed, this procedure becomes necessary due to the limited number of subjects available in the training set. Performance indicators are then computed as the average of the results obtained for each subject. The results will be discussed in the next section.

III. EXPERIMENTAL RESULTS

The datasets used to assess the performance of the deep learning architecture have been already described in Section II-A. The first dataset is built by considering successive 15 s intervals, with a 5 s overlap, each one belonging to a different class, i.e., the stress or non-stress class [18]. A 224x224x3 scalogram is associated to each 15 s interval. Having tested a total of 18 subjects driving in a simulator, we end up with a balanced dataset with 6390 intervals for both classes. The second dataset is built by considering successive 30 s time intervals, with a 50% overlap, each one belonging to the low-level or high-level stress class [13]. With 9 subjects, considering only the time while they are driving along the highway and city routes, we are able to extract 1972 for both classes.

As mentioned before, for case 1, we extract 80% of scalogram time intervals for training and 20% for test. To optimize the network, 20% of the training data are used as a validation set. For case 2, we collect the time intervals coming from all subjects except one, which will be used as test, and this is done for each test subject. As in case 1, 20% of the training data has been used as a validation set. In addition, for both cases, we choose the SGD optimizer, with a learning rate equal to 0.001, the categorical cross-entropy as a loss function, the number of epochs equal to 300, and the batch-size equal to 32 [22]. An early stopping procedure is also applied during each training phase, monitoring the loss on the validation set (with a "patience" value equal to 20). In particular, for each epoch, it evaluates the loss function on the validation set, and stops the training when the validation loss is not getting smaller than the value computed some epochs before it, according to the defined patience value. The early-stopping procedure allows our DL algorithm to avoid overfitting the training set. We choose the model which gives us the best classification accuracy on the validation set, to be then applied on the test dataset, thus evaluating the model accuracy on previously unseen data.

In Table II we show the final performance results of our system, applying the multi-VGG-CNN architecture on the test sets, for both cases and both datasets. As performance indicators, we include the accuracy, the sensitivity, the specificity, the balanced accuracy (BA), and the geometric mean (GM), which can be computed as follows:

$$\text{Accuracy } (\%) = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \cdot 100 \tag{1}$$

$$\text{Sensitivity } (\%) = \frac{\text{TP}}{\text{TP+FN}} \cdot 100 \tag{2}$$

$$\text{Specificity } (\%) = \frac{\text{TN}}{\text{FP+TN}} \cdot 100 \tag{3}$$

$$\text{BA } (\%) = \frac{1}{2}\left(\frac{\text{TP}}{\text{TP+FN}} + \frac{\text{TN}}{\text{FP+TN}}\right) \cdot 100 \tag{4}$$

$$\text{GM } (\%) = \sqrt{\frac{\text{TP}}{\text{TP+FN}} \cdot \frac{\text{TN}}{\text{FP+TN}}} \cdot 100 \tag{5}$$

These figures are based on the number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), which can be calculated considering each case and each dataset. In case 2, we indicate the (MEAN $\pm$ STD) values, computed by averaging the results of all subjects.

As far as case 1 is concerned, we notice that the accuracy values are very high, for both datasets. In particular, with dataset 1 we attain accuracy and GM values equal to 91.78%, and a balanced accuracy of 91.81%. With dataset 2, the accuracy and the other performance indicators are much higher, achieving an accuracy up to 99.24%, and both BA and GM up to 99.22%. The specificity and sensitivity values are similar, for both datasets. These results confirm that the

model is accurate in learning how to discern between the two different classes.

Regarding case 2, all of the performance indicators are worse than the ones obtained for case 1. We expected this behaviour, since in this scenario only a limited number of intervals belonging to a subject are included in the training procedure, and the majority of them are excluded. Nevertheless, the performance results are favorable. We obtain an accuracy of 84.35% and 96.64% with dataset 1 and dataset 2, respectively. It can be seen that the performance indicators are higher for dataset 2. This could be due to the characteristics of the physiological signals belonging to the public dataset, and the related scalograms, which allow a better classification of the signals. As an additional test, similarly to [16], [17] where only the scalograms of the ECG signal are used, we also compute (in case 2) the accuracy of our architecture when considering the HR scalograms only (i.e., by removing the SPR channel in Figure 4). We obtain an accuracy value of $(80.55 \pm 3.77)(\%)$ for dataset 1, and $(89.85 \pm 5.32)(\%)$ for dataset 2, which are smaller that the values reported in Table II.

Our findings, in terms of performance of DL algorithms used in different driving scenarios, appear to be acceptable, and higher when combining different physiological signals compared to the use of a single signal. However, we are aware of some of the limitations of the proposed system, such as the use of some scalograms belonging to a subject under test in the training process. In future works we will try to overcome this limit, e.g., by increasing the size of the pool of subjects in the training set.

IV. Conclusion

We described a system which employs a deep learning architecture for stress detection in subjects driving in a simulated environment, trying to overcome several stress-inducing events, and in a real-world context, with different road routes inducing different stress levels. We use the physiological signals logged from the subjects in these two different scenarios to compute a 2D representation using scalograms. Scalograms sub-images are then extracted, and sent to a multi-VGG-CNN architecture which is able to capture and extract the features that better describe the characteristics of the scalograms belonging to two different classes. The performance of the proposed system is promising, showing that by using the scalograms computed from the physiological signals logged from subjects while driving, we can recognize their emotional state with acceptable accuracy.

References

[1] Y. Amichai-Hamburger, *Technology and psychological well-being*. Cambridge University Press, 2009.

[2] A. Stephens, S. Trawley, J. Ispanovic, and S. Lowrie, "Self-reported changes in aggressive driving within the past five years, and during COVID-19," *PLoS one*, vol. 17, no. 8, 2022.

[3] J. Healey and R. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 156–166, 2005.

[4] N. Kim, M. Choe, J. Park, J. Park, H. Kim, J. Kim, M. Hussain, and S. Jung, "Analysis of relationship between electroencephalograms and subjective measurements for in-vehicle information system: A preliminary study," *International Journal of Environmental Research and Public Health*, vol. 18, no. 22, 2021.

[5] W. Li, R. Tan, Y. Xing, G. Li, S. Li, G. Zeng, P. Wang, B. Zhang, X. Su, D. Pi, G. Guo, and D. Cao, "A multimodal psychological, physiological and behavioural dataset for human emotions in driving tasks," *Scientific Data*, vol. 9, pp. 1–20, 2022.

[6] G. Oh, E. Jeong, R. Kim, J. Yang, S. Hwang, S. Lee, and S. Lim, "Multimodal data collection system for driver emotion recognition based on self-reporting in real-world driving," *Sensors*, vol. 22, no. 12, 2022.

[7] M. Alnashashibi, W. Hadi, and N. El-Khalili, "Predicting stress levels of automobile drivers using classical machine learning classifiers," in *2022 International Conference on Business Analytics for Technology and Security (ICBATS)*, 2022, pp. 1–5.

[8] M. Tauqeer, S. Rubab, M. A. Khan, R. Naqvi, K. Javed, A. Alqahtani, S. Alsubai, and A. Binbusayyis, "Driver's emotion and behavior classification system based on internet of things and deep learning for advanced driver assistance system (ADAS)," *Computer Communications*, vol. 194, pp. 258–267, 2022.

[9] P. Zontone, A. Affanni, R. Bernardini, L. Del Linz, A. Piras, and R. Rinaldo, "Supervised learning techniques for stress detection in car drivers," *Advances in Science, Technology and Engineering Systems Journal*, vol. 5, no. 6, pp. 22–29, 2020.

[10] P. Zontone, A. Affanni, A. Piras, and R. Rinaldo, "Exploring physiological signal responses to traffic-related stress in simulated driving," *Sensors*, vol. 22, no. 3, 2022.

[11] P. Zontone, A. Affanni, R. Bernardini, L. Del Linz, A. Piras, and R. Rinaldo, "Emotional response analysis using electrodermal activity, electrocardiogram and eye tracking signals in drivers with various car setups," in *2020 28th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 1160–1164.

[12] ——, "Analysis of physiological signals for stress recognition with different car handling setups," *Electronics*, vol. 11, no. 6, 2022.

[13] J. Lee, H. Lee, and M. Shin, "Driving stress detection using multimodal convolutional neural networks with nonlinear representation of short-term physiological signals," *Sensors*, vol. 21, no. 7, 2021.

[14] M. Vetterli and J. Kovačevic, *Wavelets and Subband Coding*. USA: Prentice-Hall, Inc., 1995.

[15] E. Sejdic, I. Djurovic, and L. Stankovic, "Quantitative performance analysis of scalogram as instantaneous frequency estimator," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3837–3845, 2008.

[16] M. Amin, K. Ullah, M. Asif, A. Waheed, S. U. Haq, M. Zareei, and R. R. Biswal, "ECG-based driver's stress detection using deep transfer learning and fuzzy logic approaches," *IEEE Access*, vol. 10, pp. 29 788–29 809, 2022.

[17] S. Arefnezhad, A. Eichberger, M. Frühwirth, C. Kaufmann, M. Moser, and I. Koglbauer, "Driver monitoring of automated vehicles by classification of driver drowsiness using a deep convolutional neural network trained by scalograms of ECG signals," *Energies*, vol. 15, no. 2, 2022.

[18] P. Zontone, A. Affanni, R. Bernardini, A. Piras, R. Rinaldo, F. Formaggia, D. Minen, M. Minen, and C. Savorgnan, "Car driver's sympathetic reaction detection through electrodermal activity and electrocardiogram measurements," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 12, pp. 3413–3424, 2020.

[19] A. Goldberger *et al.*, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals." *Circulation [Online]*, vol. 101, no. 23, p. e215–e220, 2020.

[20] J. Lilly and S. Olhede, "Generalized morse wavelets as a superfamily of analytic wavelets," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 6036–6041, 2012.

[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, arXiv:1409.1556.

[22] F. Chollet *et al.*, "Keras," 2015, uRL: https://keras.io.