2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 979-8-3503-2447-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/EMBC40787.2023.10340564

Assessment of Driver's Stress using Multimodal Biosignals and Regularized Deep Kernel Learning

Vishal Singh Roha[1,4], Nagarajan Ganapathy[2], *Member, IEEE*, Nicolai Spicher[3], *Member, IEEE*, Sriparna Saha[4], *Senior Member, IEEE*, and Thomas M. Deserno[1], *Senior Member, IEEE*

***Abstract*— In this work, we classify the stress state of car drivers using multimodal physiological signals and regularized deep kernel learning. Using a driving simulator in a controlled environment, we acquire electrocardiography (ECG), electrodermal activity (EDA), photoplethysmography (PPG), and respiration rate (RESP) from $\mathrm{N} = 10$ healthy drivers in experiments of 25min duration with different stress states (5min resting, 10min driving, 10min driving + answering cognitive questions). We manually remove unusable segments and approximately 4h of data remain. Multimodal time and frequency features are extracted and employed to regularized deep kernel machine learning based on a fusion framework. Task-specific representations of different physiological signals are combined using intermediate fusion. Subsequently, the fused multimodal features are fed a support vector machine (SVM) and a random forest (RF) for stress classification. The experimental results show that the proposed approach can discriminate between stress states. The combination of PPG and ECG using RF as classifier yields the highest F1-score of 0.97 in the test set. PPG only and RF yield a maximum F1-score of 0.90. Furthermore, subject-specific cross-validation improves performance. ECG and PPG signals are reliable in classifying the stress state of a car driver. In summary, the proposed framework could be extended to real-time stress state assessment in driving conditions.**

I. INTRODUCTION

Stress is defined as the sensation of being overwhelmed or unable to cope with mental or emotional pressure. Driver stress occurs when environmental stresses such as poor visibility, road conditions, delays, and personality factors combine. This can lead to erroneous decisions and serious injuries. It is reported that the road automobile accidents claim the lives of approximately 1.3 million people every year [1]. Acute stress disorder affects one out of every five accident victims. After the occurrence of stress, one out of every four people experiences psychiatric issues, including post-traumatic stress disorder [2][3]. Taking the National Travel Survey of England as an example, the average person spends more than 36 minutes each day driving [3]. This offers the possibility to transform vehicles into personalized diagnostic spaces [4] for healthcare monitoring based on biosignal sensors.

In the majority of related works, single modalities such as ECG, EDA, PPG, or RESP only are considered for measuring stress [5], [6]. However, different physiological responses can be observed across individuals which are subject to similar stress conditions. Thereby, stress assessment using single modalities might fail and it has been reported that the combination of modalities can improve the quantification of the stress states [5]. In a recent study, multimodal signals and the fusion of their features achieved a more robust model to classify human stress [7]. However, multimodal fusion is associated with several technical challenges such as the synchronized acquisition of the signals. In a real-life scenario, there is a delay between the data obtained from sensors which causes the fusion of asynchronous samples, resulting in the degradation of model performance [8].

In this work, we propose a fusion approach to combine multimodal and heterogeneous data to differentiate various stress state in drivers. In particular, we have two research questions:

RQ1: Can the fusion of multimodal biosignals improve the performance of stress state assessment in drivers?

RQ2: How accurate are the fused multimodal features using state-of-the-art machine learning approaches?

II. MATERIALS

1. Study Population: In this study, multimodal data of N = 10 healthy driver volunteers (gender: 5 females, age:25.5 ± 2.3 years, weight: 72.4 ± 10.6 Kg) was recorded, namely ECG, EDA, PPG, and RESP using commercially-available sensors. All participants gave written consents.

2. Study Design: We acquired the multi-modal signals namely ECG, PPG, EDA, RESP using non-invasive sensors (Biosignalplux, Plux, Lisbon, Portugal) in a simulated driving experiment. The signals are digitized at 512Hz per channel on 16 bit resolution. The study used single lead ECG sensor placed on chest whereas PPG and EDA sensor are placed on the non-dominate hand. The RESP signals are obtained using belt-type inductive respiration sensor placed around the chest. The CARLA driving simulator mimics real-world driving in the Logitech driving setup [9]. All volunteers had the following conditions: i) resting (5 min), ii) normal driving (10 min), and iii) answering cognitive questions while driving (10 min). Each subject performed the experiment for a total time period of 25 minutes.

*This work is supported by German Academic Exchange Service (DAAD) under the KOSPIE program.

[1]Vishal Singh Roha and Thomas M. Deserno are with the Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Brunswick, Germany (corresponding authors e-mail ID: vishal.roha95@gmail.com and thomas.deserno@plri.de)

[2]Nagarajan Ganapathy is with the Department of Biomedical Engineering, Indian Institute of Technology, Hyderabad, India

[3]Nicolai Spicher is with the Institute for Medical Informatics, University Medical Center Göttingen, Göttingen, Germany

[4]Sriparna Saha is with the Department of Computer Science Engineering, Indian Institute of Technology, Patna, India

III. METHODS

The overall architecture of the proposed approach is depicted in Fig. 1. Below, we provide a brief discussion of all the modules included in the approach, along with their respective parameter settings.

A. Dataset Preprocessing and Feature Extraction: The acquired signals are preprocessed using the Neurokit2 toolbox [10] as mentioned below.

1) *ECG:* Fifth-order Butterworth filter with only the low-cut Frequency (LF) of 0.5 Hz is used.
2) *RESP:* Second-order Butterworth filter is used with a high-cut Frequency (HF) of 3 Hz and a LF of 0.05 Hz.
3) *EDA:* It uses Butterworth B/A (Numerator (B) and Denominator (A) Polynomials) filter for cleaning the signal. HF is taken 0.02 along with the Order 4. Then smoothing of signal is performed by using the convolution method along with the boxzen kernel [10].
4) *PPG:* Second-order Butterworth filter with HF of 40Hz is used.

Multimodal features are extracted from the signals (see Tbl. I). We consider the whole timeline of baseline (5min), normal (10min), and cognitive (10min) as a separate window each. We also consider each trial as individual signal vector, each of the subjects had three trials.

B. Multiple kernel learning (MKL): MKL is a prominent technique for fusing features at the feature level. It applies the kernel method to map non-linearly separable features from various modalities to unique high-dimensional feature spaces and consider the combination of several kernels with criterion, thus improving classification performance.

C. Neural-Network based Regularized Deep Kernel Machine Optimization (NRDKMO): The proposed NRDKMO learns a representation for a single ensemble embedding individually. It consists of an embedding layer and the potential feature learning layer (PFLL) to obtain the valuable features for the fusion of data modality. The embedding layer uses a kernel matrix (KM) generated by a kernel function to assess the similarity between the samples. PFLL uses the outcome of embedding layer as input and learns valuable features using a multilayer fully connected network (mFCN). A single multilayer FCN significantly reduces the computation's time and space complexity [8].

1) Embedding Layer: In this layer, we assume a KM generated by the kernel $M \in R^{n\times n}$, where $M_{i,j} = m(y_i, y_j)$ and n is the number of input samples. The similarity between the sample y_i and all other samples y_j is represented by the m_i row of the KM, which is regarded as an embedding for the sample y_i. The values of the KM are sparse in the ideal scenario because samples from the same class have larger values, while samples from other classes have small values close to zero. Furthermore, as the sample size grows, the number of embedding dimensions grows also. Therefore, the original embedding is inappropriate for inference tasks due to its sparsity and large dimensionality. As a result, it is challenging to attain reasonable accuracy by developing a model straight from the original embedding.

For a given KM $M \in R^{n\times n}$, we intend to find an approximation matrix $M_q^{'}$ with a rank considerably lower than the number of samples for a given KM. We first define matrix $P \in R^{n\times s}$, which consists of s columns randomly chosen from the matrix M, using the Nyström method [11], an effective way for generating low-rank matrix approximations. These randomly chosen s columns will be utilized to determine an approximate kernel map $DE \in R^{n\times q}$, with $M \approx (DE)(DE^T)$ and s << n and q $\leq$ s. Following the reconfiguration of the KM M, M and P can be written as

$$M = \begin{bmatrix} Q & U^T \\ U & V \end{bmatrix} \text{ and } P = \begin{bmatrix} Q \\ U \end{bmatrix} \tag{1}$$

where $Q \in R^{s\times s}$ represents the matrix comprising the intersection of P with the corresponding s rows of M. After rearranging the M, the remaining components are U and V. Q is a PSD because the KM M is also a PSD. Then the Nyström method constructs a rank-q approximation for a given $q \leq s$. It can be represented as $M^{'}{}_q = PQ^{'}{}_q + P^T$. Here $Q^{'}{}_q$ is the truncated singular value decomposition's (TSVD) best-q approximation of Q. The mapping function is then obtained and can be represented as $DE = P(\alpha_{Q^{'}{}_q}\lambda^{-1/2}_{Q^{'}{}_q})$. The top q eigenvalues and eigenvectors of Q are $\alpha_{Q^{'}{}_q}$ and $\lambda^{-1/2}_{Q^{'}{}_q}$. Instead of using approximated kernels $M^{'}{}_q$, we directly employ the approximate mappings *DE* to reduce the number of dimensions. Actually, the RKHS will construct completely distinct representations based on the mapping functions generated from various samples. As a result, we model the characters of distinct regions in the input space using alternative representations and compute various mapping functions DE_1, DE_2,..., and DE_p from the distinct sample subsets obtained by using the multiple random sampling process. Finally, the embedding layer employs an ensemble technique to enhance the j-dense embedding obtained by j mapping functions. Thus, an ensemble dense embedding generally takes the following form:

$$DE_{ens} = \sum_{i=1}^{j} \beta_i DE_i \tag{2}$$

where β_i is the ensemble weight which is defined as the reciprocal of the number of embeddings.

2) Potential Feature Learning Layer (PFLL): The embedding obtained from the embedding layer in the NRDKMO model is forwarded to an FCN in the potential feature learning layer to learn the valuable features for the corresponding task. So, the valuable features (PFL) learned by NRDKMO can be represented as $PFL = FCN(DE_{ens})$. Here $FCN(.)$ is the multilayer FCN present in the PFLL.

D. Intermediate Fusion: Upon the learning of potential features from the NRDKMO model, the features are fused into a final representation set using intermediate fusion. The final representation set consists of $2^d - 1$ representation, where d is the total number of psychological signals. Then each representation is sent to the SVM and RF classifiers to predict the stress state of a car driver.

E. Hyperparameters and metrics: SVM classifier kernels are: (RBF, Linear, Poly2, and Poly3), Number of dense

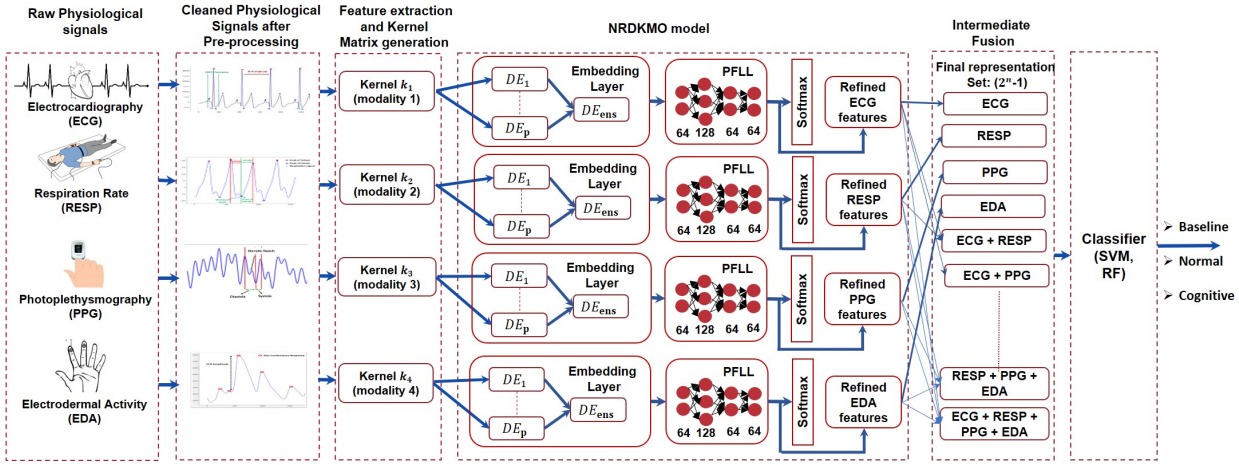

Fig. 1. The proposed framework's pipeline. It is important to note that the abbreviations used are as follows: NRDKMO stands for Neural-Network based Regularized Deep Kernel Machine Optimization, PFLL represents Potential Feature Learning Layer, and DE denotes Dense Embedding.

TABLE I

FEATURES USED IN THE EXPERIMENT [8]

Signal	Number of features	Feature set
ECG	81	Mean, Standard Deviation(SD), Skewness, Kurtosis,Power Spectral Density(PSD) $(log(P_x(f)), f \in [0.1, 0.2], [0.2, 0.3], [0.3, 0.4], [0.4, 0.6], [0.6, 1.0], [1.0, 1.5], [1.5, 2.0], [2.0, 2.5], [2.5, 5.0])$, Heart Rate Variability (Time, Frequency, and Non-linear Domain)
RESP	15	Mean, SD, Skewness, Kurtosis, PSD $(log(P_x(f)), f \in [0.0, 0.1], [0.1, 0.2], [0.2, 0.3], [0.3, 0.4], [0.4, 0.5], [0.5, 0.6], [0.6, 0.7], [0.7, 0.8], [0.8, 0.9], [0.9, 1.0])$, Main Frequency (Frequency at which PSD reaches its maximum value $(f \in [0.16, 0.6])$)
EDA	5	Mean, SD, Number of peaks, Maximum rise time, Maximum Amplitude
PPG	15	Mean, SD, Skewness, Kurtosis, PSD $(log(P_x(f)), f \in [0.0, 1.0]$ with range varying from step size difference of 0.1 Main Frequency (Frequency at which PSD reaches its maximum value $(f \in [0.16, 0.6],))$

TABLE II

STRESS CLASSIFICATION ON SAMPLE AND SUBJECT BASED LOO CV USING SVM AND RF CLASSIFIERS. THE SYMBOL '+' INDICATES CONCATENATION OF FEATURES. BEST RESULTS OBTAINED FROM A SINGLE AND FUSION OF 2, 3, AND 4 SIGNALS ARE HIGHLIGHTED IN BOLD.

	SVM										RF									
	Sample based					Subject based					Sample based					Subject based				
Signals	**Acc**	**Pr**	**F1-score**	**Re**	**ROC**	**Acc**	**Pr**	**F1-score**	**Re**	**ROC**	**Acc**	**Pr**	**F1-score**	**Re**	**ROC**	**Acc**	**Pr**	**F1-score**	**Re**	**ROC**
ECG	0.80	0.80	0.80	0.80	0.85	0.83	0.83	0.83	0.83	0.88	0.80	0.80	0.80	0.80	0.85	0.80	0.80	0.80	0.80	0.85
EDA	0.43	0.43	0.43	0.43	0.58	0.60	0.60	0.60	0.60	0.70	0.60	0.60	0.60	0.60	0.70	0.67	0.67	0.67	0.67	0.75
RESP	0.60	0.60	0.60	0.60	0.70	0.70	0.70	0.70	0.70	0.78	0.60	0.60	0.60	0.60	0.70	0.73	0.73	0.73	0.73	0.80
PPG	**0.80**	**0.80**	**0.80**	**0.80**	**0.85**	**0.87**	**0.87**	**0.87**	**0.87**	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**	**0.93**	**0.90**	**0.90**	**0.90**	**0.90**	**0.93**
ECG + EDA	0.80	0.80	0.80	0.80	0.85	0.87	0.87	0.87	0.87	0.90	0.83	0.83	0.83	0.83	0.88	0.83	0.83	0.83	0.83	0.88
ECG + RESP	0.83	0.83	0.83	0.83	0.88	0.87	0.87	0.87	0.87	0.90	0.87	0.87	0.87	0.87	0.90	0.87	0.87	0.87	0.87	0.90
ECG + PPG	**0.93**	**0.93**	**0.93**	**0.93**	**0.95**	**0.97**	**0.97**	**0.97**	**0.97**	**0.98**	**0.97**	**0.97**	**0.97**	**0.97**	**0.98**	**0.97**	**0.97**	**0.97**	**0.97**	**0.98**
EDA + RESP	0.80	0.80	0.80	0.80	0.85	0.83	0.83	0.83	0.83	0.88	0.70	0.70	0.70	0.70	0.78	0.73	0.73	0.73	0.73	0.80
EDA + PPG	0.77	0.77	0.77	0.77	0.83	0.87	0.87	0.87	0.87	0.90	0.80	0.80	0.80	0.80	0.85	0.90	0.90	0.90	0.90	0.93
RESP + PPG	0.87	0.87	0.87	0.87	0.90	0.90	0.90	0.90	0.90	0.93	0.87	0.87	0.87	0.87	0.90	0.83	0.83	0.83	0.83	0.88
ECG + EDA + RESP	0.87	0.87	0.87	0.87	0.90	0.87	0.87	0.87	0.87	0.90	0.90	0.90	0.90	0.90	0.93	0.90	0.90	0.90	0.90	0.93
ECG + EDA + PPG	0.87	0.87	0.87	0.87	0.90	**0.97**	**0.97**	**0.97**	**0.97**	**0.98**	**0.97**	**0.97**	**0.97**	**0.97**	**0.98**	**0.97**	**0.97**	**0.97**	**0.97**	**0.98**
ECG + RESP + PPG	**0.90**	**0.90**	**0.90**	**0.90**	**0.93**	**0.97**	**0.97**	**0.97**	**0.97**	**0.98**	0.93	0.93	0.93	0.93	0.95	**0.97**	**0.97**	**0.97**	**0.97**	**0.98**
EDA + RESP + PPG	0.87	0.87	0.87	0.87	0.90	0.97	0.97	0.97	0.97	0.98	0.80	0.80	0.80	0.80	0.85	0.83	0.83	0.83	0.83	0.88
ECG + EDA + RESP + PPG	**0.93**	**0.93**	**0.93**	**0.93**	**0.95**	**0.97**	**0.97**	**0.97**	**0.97**	**0.98**	**0.97**	**0.97**	**0.97**	**0.97**	**0.98**	**0.97**	**0.97**	**0.97**	**0.97**	**0.98**

embedding(p): 3, Size of dense embedding(n): 10, Regularization technique: Dropout(0.1), Loss: categorical cross entropy, Optimizer: Adam, Learning rate: $1e^{-3}$, Batch size: 8, Epochs: 500. In RF Classifier, max depth of tree: 3, number of estimators: 200. Perforamcne metrics considered are accuracy, precision, recall, F1-score, and ROC curve using Leaving-one-out (LOO) cross-validation on both the sample-based and subject-based features.

IV. RESULTS

Our experiments are based on evaluating different kernels in the KM and SVM classifier as described in sec. III. Moreover, three types of feature combinations were applied to the features obtained from the representation learning on the different physiological signals, namely, concatenation, summation, and multiplication. Tbl. II contains the best results obtained from the RBF kernel and concatenation of features. In the case of individual signals (without fusion), PPG outperforms ECG, EDA, and RESP for both, sample and subject-based, validation. Analyzing the results of the multi-modal fusion answers **RQ1** by clearly demonstrating that the fusion of different improves stress state classification (see Tbl. II). The use of MKL results in the approach to be not sensititve to synchronization of signals. It also demonstrates the robustness of the approach as the fusion of different psychological signals never degrades the classification performance, which answers **RQ2**.

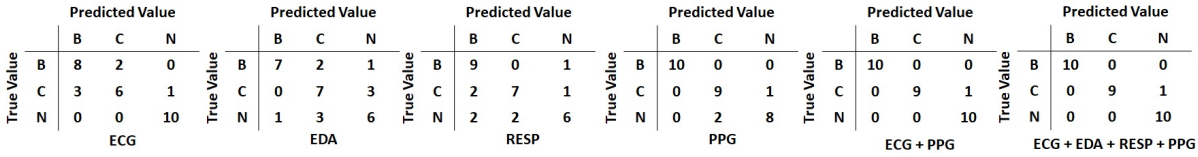

Fig. 2. Confusion matrices of ECG, EDA, RESP, PPG, ECG+PPG, ECG+EDA+RESP+PPG for subject-based LOO CV using the RF classifier for all the 10 subjects. B represents baseline, i.e. the first part of the experiments while the subject is sitting, N represents normal driving, and C represents driving combined with cognitive stress.

Analyzing F1-score in sample-based LOO, the fusion of ECG + PPG features shows an improvement of *16.25%, 16.28%, 55.00%, 16.25%* using SVM and *21.25%, 61.67%, 61.67%, 7.78%* using RF for ECG, EDA, RESP, and PPG signals, respectively. In subject-based LOO, the fusion of features shows an increment in F1-score of *16.87%, 61.67%, 38.57%, 11.49%* using SVM and *21.25%, 53.97%, 25.97%, 7.78%* using RF for ECG, EDA, RESP, and PPG signals respectively.

Subject-based LOO has performed better than sample-based LOO on most of the fusions for both the SVM and RF classifiers. Tbl. II shows that PPG and ECG signals better classify stress than RESP and EDA. PPG and ECG outperform RESP and EDA by *86.04*% and *33.33*% in terms of F1-score for SVM Sample-based LOO. A similar trend can be observed for other cases as well. Additionally, PPG outperforms ECG by *12.50*% for RF for the sample-based and subject-based LOO and *4.82*% for SVM in case of Subject-based LOO.

The confusion matrices of ECG, EDA, RESP, and PPG and their combination with and without fusion for RF are presented in Fig. 2. On the one hand, it can be observed that EDA and RESP fail to correctly predict the cognitive and normal stress level resulting in a degraded performance (see Fig. 2). On the other hand, ECG and PPG differentiate the baseline and normal stress level successfully but they face problems in correctly predicting the cognitive stress level. This problem can be mitigated by fusing ECG and PPG, resulting in 90% accuracy for cognitive stress classification. It can also be observed that features obtained from ECG and PPG dominate the fusion even when combined with other signals as the confusion matrix of ECG + EDA + RESP + PPG is exactly the same as in case of ECG + PPG. The findings in Tbl. II provide further evidence to support this argument. One possible explanation for this phenomenon is the relatively small sample size of only 30 data samples. As the number of data samples increases, the impact of EDA and RESP on the results is expected to become more prominent. Moreover, it is noteworthy that the confusion matrix of EDA + RESP + PPG differs from that of PPG alone, and a similar observation can be made regarding the confusion matrix of ECG + RESP + PPG compared to ECG. Notably, the fusion of ECG and PPG has yielded better results than any other fusion method. The maximum F1-score achieved by ECG+PPG is 0.93 and 0.97 for SVM and RF for Sample-based LOO and 0.97 for both SVM and RF for Subject-based LOO. ECG+PPG outperforms RESP+EDA by *16.25*% and *38.57*% for SVM and RF for Sample-based LOO and *16.87*% and *32.88*% for SVM and RF for Subject-based LOO.

V. Conclusion & Future Work

In this paper, we considered multimodal psychological signals to classify different stress levels in a driver: ECG, EDA, RESP, and PPG. The multimodal fused features are more effective than single modality in stress level prediction. We observed that the performance improves with *16.25%, 116.28%, 55.00%, 16.25%* using SVM as classifier, and *21.25%, 61.67%, 61.67%, 7.78%* for RF. RF outperforms the SVM when it comes to sample-based validation. The proposed method can be extended to other heterogeneous data and is found to be insensitive by data synchronization issues. In the future, other signals could be added to the setup, e.g. camera-based modalities. Regarding signal processing, convolutional neural networks with an attention mechanism will be explored to pay closer attention to the features extracted from the intermediate fusion.

References

[1] "Road traffic injuries," Who.int. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries. [Accessed: 08-Apr-2022].

[2] D. K. Ścigała and E. Zdankiewicz-Ścigała, "The role in road traffic accident and anxiety as moderators attention biases in modified emotional Stroop test," Front. Psychol., vol. 10, p. 1575, 2019.

[3] D. Stillwell, A. Evans -Andrew, and K.-M. Slocombe, "National Travel Survey: England 2018," Gov.uk, 2019. [Online]. [Accessed: 08-Apr-2022].

[4] T. M. Deserno, "Transforming smart vehicles and smart homes into private diagnostic spaces," Proc. ACM APIT. 2020; 165-71.

[5] W.-Y. Chung, T.-W. Chong, and B.-G. Lee, "Methods to detect and reduce driver stress: A review," Int. J. Automot. Technol., vol. 20, no. 5, pp. 1051–1063, 2019.

[6] N. Ganapathy, Y. R. Veeranki, and R. Swaminathan, "Convolutional neural network based emotion classification using electrodermal activity signals and time-frequency features," Expert Syst. Appl., vol. 159, no. 113571, p. 113571, 2020.

[7] L.-L. Chen, Y. Zhao, P.-F. Ye, J. Zhang, and J.-Z. Zou, "Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers," Expert Syst. Appl., vol. 85, pp. 279–291, 2017.

[8] X. Zhang et al., "Emotion recognition from multimodal physiological signals using a regularized deep fusion of kernel machine," IEEE Trans. Cybern., vol. 51, no. 9, pp. 4386–4399, 2021.

[9] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, CARLA: An open urban driving simulator. InConference on robot learning. PMLR, 2017.

[10] "Functions — NeuroKit 0.1.7 documentation," Readthedocs.io. [Online]. Available: https://neurokit2.readthedocs.io/en/latest/functions.html.

[11] S. Kumar, M. Mohri, and A. Talwalkar, "Sampling methods for the Nyström method," J. Mach. Learn. Res, vol. 13, pp. 981–1006, 2012.