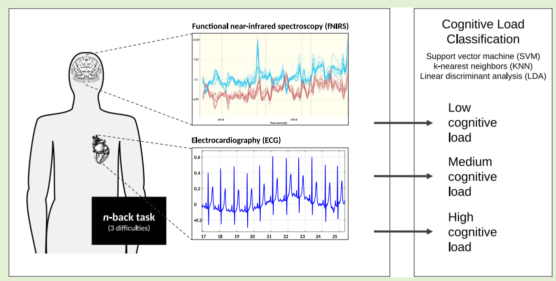
Classification of Cognitive Load Based on Neurophysiological Features From Functional Near-Infrared Spectroscopy and Electrocardiography Signals on *n*-Back Task

Ivan Kesedžić, Marko Šarlija, Jelena Božek, *Member, IEEE*, Siniša Popović, *Member, IEEE*, and Krešimir Ćosić

***Abstract*—Cognitive load can be estimated using individuals' task performance, their subjective measures, and neurophysiological measures. Neurophysiological measures, which among others include brain activation signals obtained with various brain imaging techniques, such as the functional near-infrared spectroscopy (fNIRS), and signals from the peripheral physiology, such as the electrocardiography (ECG) signal, allow an objective and continuous estimation of cognitive load. In this article, the fNIRS and ECG signals were simultaneously collected from 32 participants and used to classify three levels of cognitive load on *n*-back task. A set of 30 fNIRS and ECG features proposed in this article enables the classification of different levels of cognitive load on *n*-back task using the support vector machine (SVM), k-nearest neighbors (KNN), and linear discriminant analysis (LDA) classification models. When combining the fNIRS and ECG features, three difficulties of the *n*-back task were classified with the mean accuracies ranging from 61% to 67%, while two difficulties were classified with the mean accuracy ranging from 70% to 84%. The most important features in the classification are discussed. The results presented in this article extend the existing empirical evidence that combining brain imaging and peripheral physiology features increases the accuracy of multi-level cognitive load classification, thus further underscoring the importance of multimodal approach to cognitive load classification.**

***Index Terms*— Cognitive load classification, electrocardiography, functional near-infrared spectroscopy, sensor fusion.**



I. Introduction

COGNITIVE load refers to the construct representing the load imposed on the human cognitive system while performing a specific task [1]. Various factors can influence the cognitive load, such as task complexity, skill level, age, and spatial ability [2]. Cognitive load can be estimated using various measures, such as task performance, subjective measures, and neurophysiological measures. Task performance is the simplest measure in cognitive load estimation and is often described with task accuracy and response time. The subjective measures can be performed during (e.g. Instantaneous Self Assessment [3]) and after the task (e.g. NASA TLX [4]) but can be biased and can even impose additional cognitive load while performing the task. The neurophysiological measures allow the objective and continuous estimation of cognitive load [5], [6], provide direct measurements of the brain activation and peripheral physiology, even in a real working environment. The continuous cognitive load measurement is of special interest in safety critical professions, such as the air traffic control, since the excessive or insufficient cognitive load is associated with decreased task efficiency [7].

Manuscript received October 16, 2020; accepted November 4, 2020. Date of publication November 16, 2020; date of current version June 30, 2021. The work of Ivan Kesedžić was supported by the Croatian Science Foundation through the Project "Young Researcher's Career Development—Training of Doctoral Students." The work of Jelena Božek was supported by the European Regional Development Fund under Grant KK.01.1.1.01.0009 (DATACROSS). The associate editor coordinating the review of this article and approving it for publication was Dr. Emiliano Schena. *(Corresponding author: Ivan Kesedžić.)*

The authors are with the Faculty of Electrical Engineering and Computing, University of Zagreb, 10000 Zagreb, Croatia (e-mail: ivan.kesedzic@fer.hr; marko.sarlija@fer.hr; jelena.bozek@fer.hr; sinisa.popovic@fer.hr; kresimir.cosic@fer.hr).

Digital Object Identifier 10.1109/JSEN.2020.3038032

Brain imaging typically requires large, stationary machines that are not suitable for mobile- or self-monitoring [8]. Beside electroencephalography (EEG) and functional magnetic resonance imaging (fMRI), another neuroimaging technique lately being used in cognitive load measurements is the functional near-infrared spectroscopy (fNIRS) [9], [10]. The basic concept of fNIRS is shining light onto the scalp, detecting it as it exits the head and using the absorption spectra of the light absorbing molecules (chromophores) interpret the detected light levels as changes in chromophore concentrations [11]. The light disperses through most of the biological tissue and is absorbed by the haemoglobin. Hence, the level of oxygenated haemoglobin (HbO) and deoxygenated haemoglobin (HbR) can be estimated using the modified Beer-Lambert law [12]. Depending on the used fNIRS device, some studies scan multiple cortical regions of the brain during the experimental protocol [9], while most of them scan only specific parts of the brain, with the prefrontal cortex (PFC) and the primary motor cortex being the most common brain areas in fNIRS research [13]. The fNIRS device is more affordable than the fMRI scanner, and the results from conducted studies indicate that fNIRS device is a reliable neuromonitoring technique, when compared to the fMRI scanner [14], [15]. Beside the more affordable price, the main advantage of fNIRS compared with other brain imaging techniques is its simplicity of use. Main disadvantage of fNIRS is its low spatial resolution since most of these devices are limited only to the PFC. The fNIRS is often used as a part of the brain computer interface (BCI), a communication system that enables the use of brain activity in controlling a computer or other devices [13].

The fNIRS can detect differences in the blood oxygenation during various cognitive tasks that mostly activate the PFC, such as mental arithmetic tasks, word generation tasks, colour-word matching tasks, mental rotation tasks, working memory tasks, and inhibition tasks [16]. The PFC, and especially its dorsolateral part (dlPFC), is related to the working memory [17] which enables mental manipulation of information over a short period of time [16]. Common manipulations of cognitive load in the working memory tasks involve increases in the size or temporal lag of remembered information. For example, in the *n*-back task, which was firstly introduced in [18], the objects/stimuli are presented to the participants serially with instructions to detect repeating objects/stimuli, either successively, presenting the low cognitive load, (i.e. 1-back task) or with 2 or 3 intervening trials, presenting the higher cognitive load (i.e. 2-back task, and 3-back task) [9]. The stimuli used during the *n*-back task are mostly letters or numbers, but some studies use images or words [19]. Another variation of the *n*-back task is the 0-back task, in which a participant should respond whenever an object/stimulus specified before the task's beginning is presented.

The peripheral physiology signals used in the cognitive load assessment include electrocardiography (ECG), electromyography (EMG), electrodermal activity (EDA), respiration, etc. [6], [20], [21]. Beside these neurophysiological signals, other objective measurements, such as speech/acoustic and linguistic reactions, facial/gesture and oculomotor reactions, etc., can be used in the cognitive load estimation [6], [8]. The fNIRS-assessed brain activation signals, as well as the various signals obtained by measuring the peripheral physiology responses, have been confirmed to be effective indicators of cognitive load due to their reliability and the feasibility of continuous recording [22]–[25].

Research articles in this field can be categorised into several groups: 1) research articles describing the effect of cognitive load on brain activation and/or peripheral physiology responses (but not performing classification of cognitive load); 2) research articles on cognitive load classification using a single modality, e.g. only fNIRS signals or only specific peripheral physiological signals; 3) research articles on cognitive load classification using multiple modalities. More detailed overview of the research in this field is given in section II. When designing a cognitive load classification model based on neurophysiological signals, two approaches are possible. The easier problem is the classification of resting periods compared with the task periods, while a classification between different task difficulties is a more complex problem. Depending on the task, classification model, and computed features, the reported accuracies in cognitive load classification based on the neurophysiological signals vary.

This study aims to classify three cognitive load levels (i.e. three task difficulties) based on the features computed from the simultaneous fNIRS and ECG recordings on the *n*-back task by employing and comparing three machine learning approaches: support vector machine (SVM), k-nearest neighbors (KNN), and linear discriminant analysis (LDA). The hypothesis is that the combination of features derived from the fNIRS and ECG signals will allow for a better classification than the classification using solely one modality irrespective of the applied machine learning algorithm. This article is organised as follows: section II describes the previous work in cognitive load classification using these modalities; section III describes the participants, experimental protocol and setup, data processing and feature extraction for both fNIRS and ECG signals, as well as used classification models; section IV describes the results; and section V discusses the results.

II. Related Work

A number of fNIRS studies have examined the effect of varying *n*-back task on brain activation signals [9], [26]–[29]. The studies have generally found that higher cognitive load tends to produce greater activation within the dlPFC [30]–[33]. A study by Ogawa *et al.* [34] showed that participants with better performance on a visuospatial working memory task had higher levels of HbO activation, measured with the fNIRS. A trend toward an increase of the heart rate in more difficult *n*-back task (i.e. 2-back task) compared to the resting period is shown in a study by Mühl *et al.* [21]. The *n*-back task was used in a study by Fallahi *et al.* [35] in which the authors investigated the effects of cognitive load on physiological and subjective responses. The physiological responses included the ECG and EEG measures, among others. They showed that features obtained from ECG (ratio of low and high frequency powers) and EEG signals (alpha activity) significantly changed by increasing the task difficulty, suggesting that the mentioned features have enough sensitivity to quantify the cognitive load. In a study by Mandrick *et al.* [36] participants' cardiovascular, PFC oxygenation, and pupil diameter measures were recorded while performing the *n*-back task. The results of this study revealed that higher task difficulty induced an increased heart rate and activity in the lateral PFC, and a decreased heart rate variability (HRV).

Previous studies have adopted a combination of both brain and non-brain imaging measures such as HRV, PFC activation, respiration rate, and eye movement in cognitive load measurement and classification [35], [37], [38], since the combination of modalities typically increases the accuracy of the measurements. Based on the ECG and EDA data collected during the *n*-back task, the psychomotor vigilance task, and the visual search task, a classification of 66% was reported using the KNN classifier in discriminating the three tasks, but no reports were made on classification between task difficulties [39]. Using only the fNIRS recordings on *n*-back task, Herff *et al.* [27] classified the cognitive load with the accuracies up to 80.5% in discriminating the task from the resting period, and accuracies up to 50.3% in discriminating the three performed task difficulties against each other. A study by Keshmiri *et al.* [40] showed classification accuracies between 1-back task and 2-back task of 82.5% and 86.4% for female and male participants, respectively, using the fNIRS, while a study by Liu *et al.* [41] showed classification between three classes on *n*-back task with mean accuracy of 41.7% using the fNIRS features alone, and mean accuracy of 48.7% when using the combination of fNIRS and EEG modalities. In the same study, accuracies mostly increased when comparing only two task difficulties, ranging from 46.7% to 65.6% using only fNIRS modality, and ranging from 38.2% to 83.1% using the combination of fNIRS and EEG modality.

The combination of modalities in cognitive load classification is less frequent [38], [42]. The combination of fNIRS and cardiovascular physiology modalities in a study by Gurel *et al.* [42] resulted with the accuracy of 85% in classifying three cognitive load states: resting periods, mental arithmetic task, and *n*-back task. In a study by Liu *et al.* [38] the participants performed the 0-back, 2-back, and 3-back task, with the fNIRS, EEG, and physiological data being simultaneously measured. The authors classified three difficulties using the LDA approach based on only EEG data, only fNIRS data, only physiological data, and all of the combinations of the three modalities. The combination of fNIRS and physiological data improved the classification when comparing it to using only one modality from 42.2% for physiology-based features and 53.9% for fNIRS-based features, to 55% when combining these two modalities.

III. Methods

A. Participants

The participants in this study were 32 right-handed students from the University of Zagreb (28 male and 4 female; mean age = 24.35; SD = 1.01 years; age range = 22-26; mean Edinburgh Handedness Inventory [43] score = 76.72; SD = 23.87). All participants signed informed consent according to the guidelines of the University of Zagreb. The study was approved by the Ethics Committee of the University of Zagreb Faculty of Electrical Engineering and Computing.[1]

B. Experimental Protocol

The protocol consisted of *n*-back task and a questionnaire about participants' handedness using the Edinburgh Handedness Inventory [43]. During the *n*-back task, the participants were instructed to respond with a left arrow on the keyboard whenever they were presented the same stimulus as the one presented *n* trials previously, and with a right arrow whenever the stimulus was not the same as the one presented *n* trials previously, where *n* is a prespecified integer (n = 1, 2, or 3). At the bottom of the screen there were always shown words "yes" on the left side of the screen representing the positive response, and "no" on the right side representing that the stimulus is not the same as the one presented *n* trials previously. Once answered, the white box around the selected answer would appear. The reaction time (time between the stimulus presentation and the participant's response) and the accuracy (percentage of the correct answers) were measured for each participant. Letters were used as stimuli during this task (all uppercase consonant letters from the English alphabet without letters "Q", "X", "Y", "W" because they are not included in the Croatian alphabet). Each letter was presented on the screen for 0.5 s, with a period of 2.5 s for the participants' response. The participants had to respond after each stimulus. The experimental paradigm consisted of 13 blocks, which included six blocks of tasks, and seven blocks of resting periods (Fig. 1). During the resting periods, the participants were asked to relax and continue looking at the screen in front of them. A white "+" sign was shown on the screen during the resting period. The subjects underwent a total of 150 trials of the *n*-back task that were evenly distributed into three difficulty levels and six task blocks. The total duration of the protocol was 590 s.

C. Experimental Setup

The fNIRS Imager 1100 with 16 optodes (Fig. 2) and COBIStudio software [44] were used for data collection. The system consists of four LEDs emitting light at two different wavelengths (730 and 850 nm), and ten light detectors. The light-source separation was 2.5 cm and the imager was placed over the participants' forehead. The sampling rate of the imager was 2 Hz.

BIOPAC MP150 system with the accompanying modality-specific module (Fig. 2) was used for collecting the ECG data, at a sampling frequency of 1000 Hz. ECG electrodes (EL503 from BIOPAC) were placed on both wrists and above the right ankle (Einthoven's triangle) and Lead I was measured. The electrodes were considered successfully placed if QRS complexes in the ECG signals were easily visible to the experimenter during manual visual inspection of the signals. Visual inspection of signals after electrode placement, lasting approximately one minute, additionally

[1]Ethics Committee approval dated October 30, 2018, classification: 023-01/18-01/56, register number: 251-67/301-18/217.

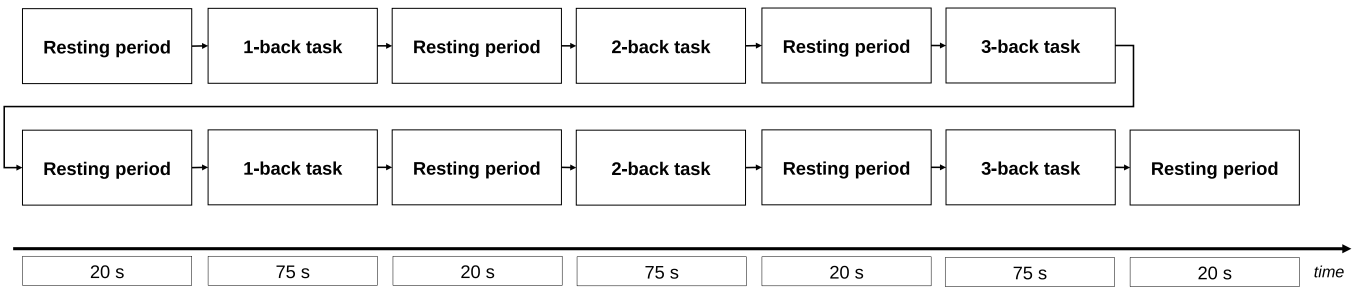


Fig. 1. Sequence of the blocks during the experimental paradigm.

Fig. 2. The equipment used in the experimental protocol: fNIRS Imager 1100 and BIOPAC MP150 system with the accompanying ECG-specific module.

helped the participants' signals to settle to a basal rest level before the start of the experiment.

During the experiment, the participants were seated facing a screen showing the tasks and a keyboard for their response. The 10 seconds baseline preceded the recording phase. The purpose of the baseline recording was the acquisition of fNIRS baseline responses used in the later haemoglobin concentration calculations. The participants were instructed to keep their natural behaviour (including swallowing, blinking, etc.) but also to minimise their movements. The participants were instructed that their reaction time and accuracy will be measured during the experiment. The participants were shown the examples of the tasks prior to the scanning. The examples were shown after the placement of the fNIRS headband on the participants' forehead and the electrodes for ECG data acquisition.

D. Data Processing and Feature Extraction

1) fNIRS: Raw light intensities recorded from the PFC were first visually inspected to reject signals from the optodes with inadequate skin contact. Raw light intensities were converted into concentration changes in HbO and HbR using the modified Beer-Lambert law [12]. HbO and HbR signals were filtered with low-pass filter using the Butterworth 20th order filter with the cut-off frequency of 0.1 Hz in order to reduce artefacts from physiological signals (e.g. heartbeat and respiration). Motion artefacts were removed using the sliding-window motion artefact rejection algorithm [45]. These cleaned signals were used in the subsequent analysis.

The HbO and HbR signals were averaged across 16 optodes for each participant in order to calculate a set of 10 fNIRS-based features. These features include: mean activation during the task (HbO_{mean} and HbR_{mean}), ratio of the mean activation during the task and the preceding resting period (HbO_{ratio} and HbR_{ratio}), variance during the task (HbO_{var} and HbR_{var}), peak value of a signal during the task (HbO_{peak} and HbR_{peak}), and slope of the signal during the task (HbO_{slope} and HbR_{slope}).

Furthermore, the HbO and HbR signals were averaged across the optodes 3 and 4, and across optodes 13 and 14, making the measures for left and right dlPFC respectively. From these signals, a set of 10 additional fNIRS-based features was calculated: mean activation of dlPFC during the task ($HbO_{dlPFCmean}$ and $HbR_{dlPFCmean}$), slope of the signal during the task ($HbO_{dlPFCslope}$ and $HbR_{dlPFCslope}$), mean activation of the left dlPFC ($HbO_{LdlPFCmean}$ ans $HbR_{LdlPFCmean}$), mean activation of the right dlPFC ($HbO_{RdlPFCmean}$ and $HbR_{RdlPFCmean}$), and ratio of left and right dlPFC activation ($HbO_{dlPFCratio}$ and $HbR_{dlPFCratio}$), making a total of 20 features calculated from fNIRS signals.

2) ECG: The raw ECG signals were preprocessed using a state-of-the-art algorithm for offline automatic detection of heartbeats [46] and in-lab custom developed MATLAB-based tool for manual checking and correction of a small percentage of heartbeats incorrectly detected by the algorithm. This step resulted in an inter-beat interval (IBI) time series on a 1-ms precision scale, which is critical for accurate assessment of the fluctuations in the obtained IBI time series, known as HRV. Changes in HRV are a direct result of autonomic regulation of the cardiovascular system in response to various environmental and psychological challenges [47], including the varying levels of cognitive load which were induced in this study. The Lomb-Scargle periodogram was used to estimate the power spectral density of the obtained non-uniformely sampled IBI signal during the task-relevant periods.

The following set of HRV features was calculated from the obtained IBI time series: mean IBI value during task (IBI_{mean}), standard deviation of IBI during task (IBI_{std}), linear slope of the IBI signal during task (IBI_{slope}), root mean square of successive differences (RMSSD) during task, ratio of the mean IBI value during task and mean IBI value during the preceding rest period (IBI_{ratio}), ratio of the mean RMSSD value during task and mean RMSSD value during the preceding rest period ($RMSSD_{ratio}$), low frequency (0.04-0.15 Hz) HRV power during task (LF), high frequency (0.18-0.4 Hz) HRV power

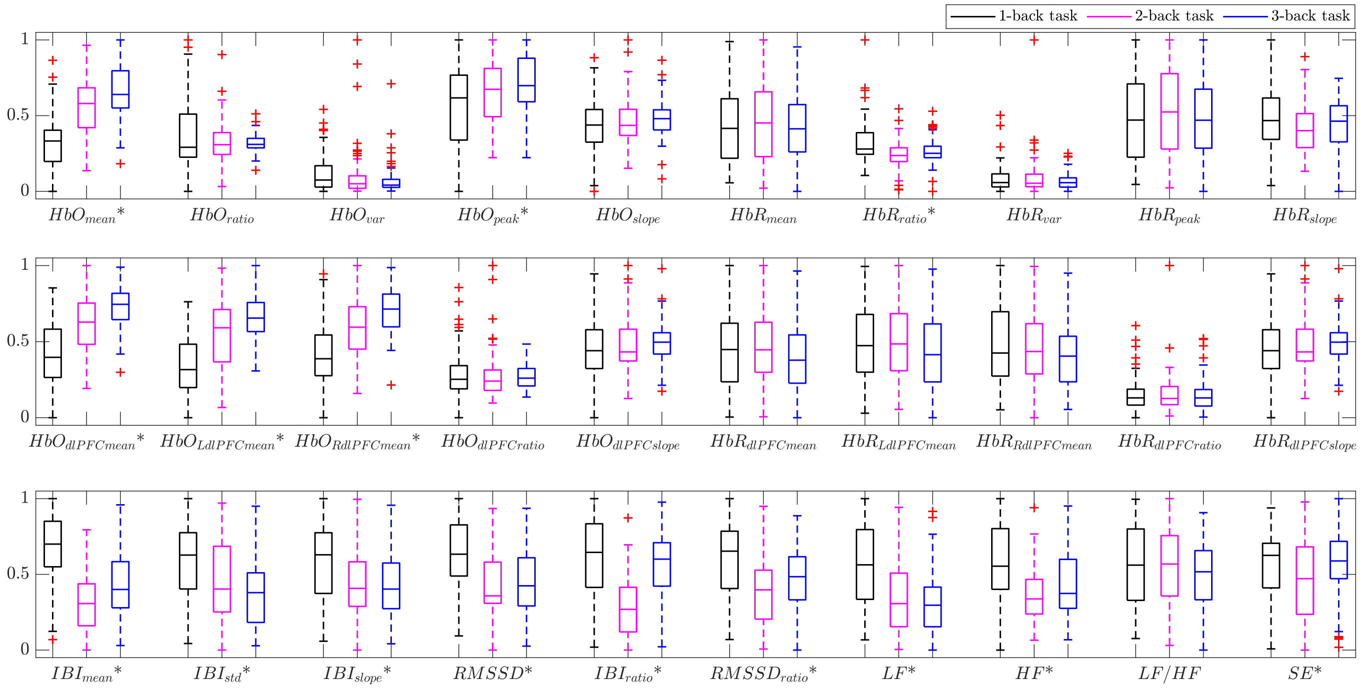


Fig. 3. Features distribution across all the participants for the performed tasks. Every feature has been computed on each of the 2 blocks of the same difficulty and joined into a single boxplot corresponding to the respective 1-back, 2-back or 3-back task. All features were scaled to the range [0,1] for the purpose of this illustration. Features containing statistically significant difference ($p<0.05$) between task segments are marked with "*".

during task (HF), ratio of low to high frequency powers during task (LF/HF), Shannon entropy (embedding dimension $L = 3$, $\xi = 10$ quantization levels [48]) of the obtained IBI time series during task (SE), making a total of ten HRV features calculated from the ECG signals.

To summarise, 20 fNIRS-based features and 10 ECG-based features comprise a total of 30 neurophysiological features. All ratio-based features, i.e. $\mathrm{HbO_{ratio}}$, $\mathrm{HbR_{ratio}}$, $\mathrm{HbO_{dlPFCratio}}$, $\mathrm{HbR_{dlPFCratio}}$, $\mathrm{IBI_{ratio}}$, $\mathrm{RMSSD_{ratio}}$, and LF/HF, were log-transformed in order to account for the skewness of the obtained feature distributions. The total number of labeled instances for each participant was six (one for each task block), making a total of 192 labeled instances for machine learning of the 3-level classification model. Fig. 3 shows the features distribution across all the participants for three difficulties of the performed tasks. For each feature, the repeated measures analysis of variance (ANOVA) was performed between task conditions, i.e. task difficulties. A statistical significance level of $p<0.05$ was considered in this calculation and the results are presented in Fig. 3.

E. Classification Models

Classification was performed using three machine learning models: support vector machine (SVM), k-nearest neighbors (KNN), and linear discriminant analysis (LDA). All data analyses, including the feature extraction, were performed in Matlab R2018b. The classification models were trained to discriminate three classes of tasks difficulty (1-back task as difficulty 1, 2-back task as difficulty 2, 3-back task as difficulty 3). The classification problem was predicting the task difficulty of the unknown data in the test set into one of the three categories (1-back task vs. 2-back task vs. 3-back task). Both single modality (fNIRS features or ECG features), as well as a combination of two modalities were incorporated in the models and evaluated. Beside these models, additional models were trained to discriminate two levels of task difficulties (1-back task vs. 2-back task, 2-back task vs. 3-back task, 1-back task vs. 3-back task), also using single modality (fNIRS or ECG), as well as using the combination of modalities.

The training process for all models was carried out using nested leave-one-subject-out (LOSO) crossvalidation, which incorporates both hyperparameter tuning and model evaluation. The data were first divided into 32 folds, where each fold comprised instances from a single participant. Data from one participant, i.e. one fold, was always held as the testing set, and the remaining data, i.e. 31 folds, were used for training. For each hold-out fold representing the testing set, the sequential feature selection algorithm was performed over the corresponding training set, and the chosen features were used in the classification model evaluation on the testing set. This process was repeated 32 times, testing on a different participant in each iteration. Classification accuracy for each testing set was calculated by dividing the number of correctly predicted labels by the total number of testing samples from each subject (i.e. correct/total). Mean classification accuracy and standard deviation (SD) of all models were calculated from the obtained 32 classification accuracies on each of the test folds. The receiver operating characteristic (ROC) curve was calculated for all the 2-level classifiers as the graph of two parameters: true positive rate and false positive rate. The area under ROC curve (AUC) was calculated for all the 2-level classifiers, as well as for the 3-level classifiers as the average of three AUC measures: AUC for 1-back task vs. all other

TABLE I
CLASSIFICATION ACCURACIES (IN %) OF ALL THE TRAINED MODELS

		1-back vs. 2-back vs. 3-back			1-back vs. 2-back			2-back vs. 3-back			1-back vs. 3-back		
		Accuracy			Accuracy			Accuracy			Accuracy		
	Feature set	Mean	SD	AUC	Mean	SD	AUC	Mean	SD	AUC	Mean	SD	AUC
SVM	fNIRS	58.85	20.73	0.7607	72.66	20.44	0.8351	62.50	16.80	0.6678	83.59	18.63	0.8859
	ECG	61.98	20.41	0.7669	80.47	18.77	0.8448	75.78	22.23	0.7954	71.09	22.99	0.7687
	fNIRS+ECG	63.54	21.77	0.8259	81.25	17.96	0.9058	76.56	20.14	0.8021	84.38	19.83	0.9045
KNN	fNIRS	52.08	16.80	0.6367	66.41	20.68	0.7234	54.69	22.39	0.5582	76.56	15.47	0.8210
	ECG	55.73	19.68	0.7511	77.34	19.43	0.8093	69.53	23.53	0.6736	60.94	26.13	0.5714
	fNIRS+ECG	61.46	19.14	0.7586	82.03	16.23	0.8725	70.31	21.48	0.7054	83.59	18.63	0.8270
LDA	fNIRS	53.65	17.83	0.7307	82.03	17.52	0.8782	64.84	20.93	0.6753	82.81	18.45	0.8958
	ECG	62.50	20.30	0.7821	79.69	17.32	0.8567	72.66	24.06	0.8113	75.00	21.06	0.7954
	fNIRS+ECG	67.71	21.56	0.8266	83.59	15.47	0.9292	80.47	20.80	0.8486	84.38	19.52	0.8987

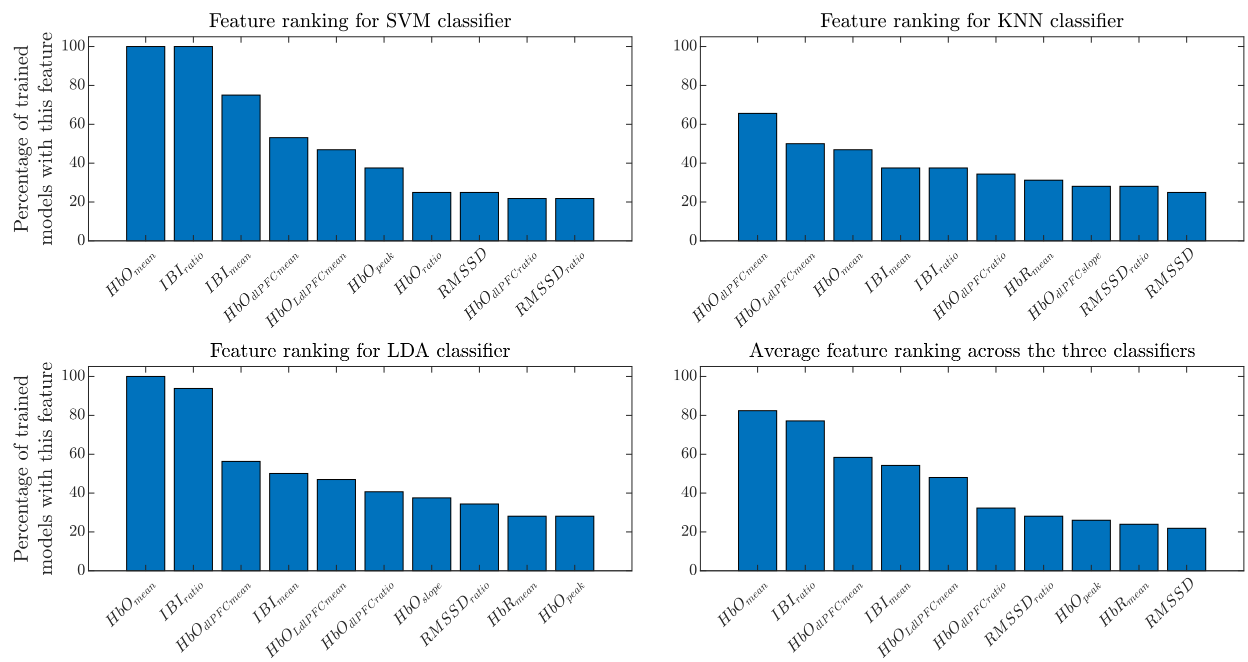


Fig. 4. Feature ranking in the SVM, KNN, and LDA model for 3-level classification (1-back task vs. 2-back task vs. 3-back task) using both fNIRS and ECG features. The final ranking score for each feature was calculated by averaging feature rankings of the each 3-level classification model. The averaged feature ranking for these three classifiers is presented at the bottom-right graph. More important features in the final model have the higher values of bar plots.

tasks, AUC for 2-back task vs. all other tasks, and AUC for 3-back task vs. all other tasks.

The kernel used for the SVM model was the radial basis function (RBF) kernel, and the hyperparameters tuned for the SVM classifier included gamma (the kernel coefficient) and C (the penalty for the error). Regarding the KNN classifier, the tuned hyperparameters were the number of neighbors (a parameter specifying the number of nearest neighbors for classifying each point when predicting) and distance (a parameter specifying the distance metric), and the hyperparameters tuned in the LDA classifier were gamma (parameter for regularizing the correlation matrix of predictors) and delta (threshold on linear coefficients). The importance of each feature, i.e. feature ranking, in all classification models was obtained by averaging rankings of each feature in all the folds of a single trained model.

IV. RESULTS

The task performance measures were calculated for each participant. The average overall performance accuracy, calculated as the ratio of the correct answers and the total number of tasks per each level, of all the participants was 95.18% during 1-back task, 94.29% during 2-back task, and 89.35% during 3-back task, while their average response time was 0.886 s on 1-back task, 1.083 s on 2-back task, and 1.189 s on 3-back task.

The evaluation scores of all the trained classifiers are presented in Table I. The mean accuracies in 3-level cognitive load classification vary between 52.08% and 67.71%, while the mean accuracies in 2-level cognitive load classification vary between 54.69% and 84.38%.

The feature ranking, i.e. the importance of each feature in the cognitive load classification, was calculated for all trained 3-level classification models (1-back task vs. 2-back task vs. 3-back task) based on the fNIRS and ECG features combined, and their ranking is presented in Fig. 4. The average number of features selected into the 3-level SVM classification model based on both fNIRS and ECG features was 4.81, average of 4.38 features for KNN model, and average of 4.53 features for the LDA 3-level classification model. In these three

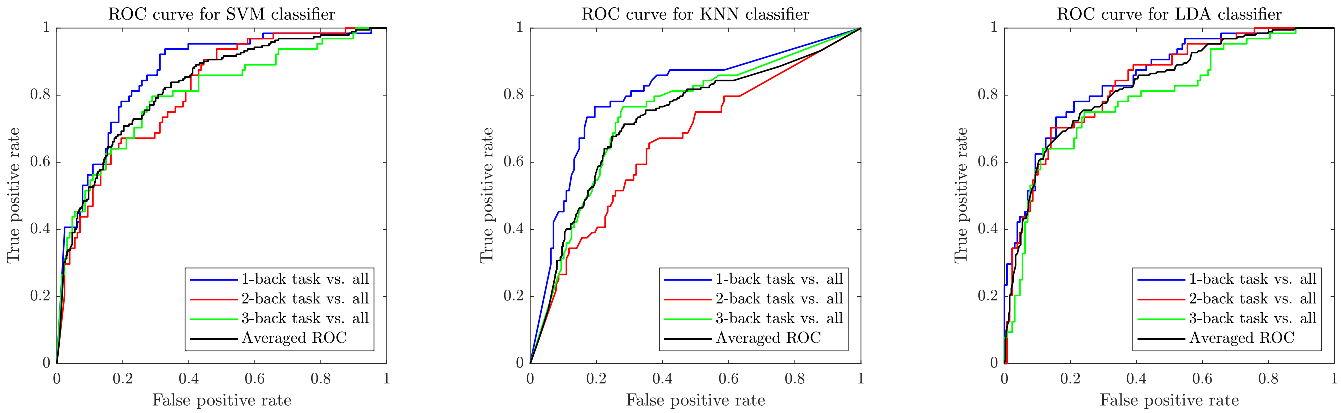


Fig. 5. ROC curves for SVM, KNN, and LDA 3-level classifiers based on fNIRS and ECG features.

models, the 10 most important features in the 3-level cognitive load classification across all participants were $\mathrm{HbO}_{\mathrm{mean}}$, $\mathrm{HbO}_{\mathrm{dlPFCmean}}$, $\mathrm{HbO}_{\mathrm{LdlPFCmean}}$, $\mathrm{HbO}_{\mathrm{dlPFCratio}}$, $\mathrm{HbO}_{\mathrm{peak}}$, and $\mathrm{HbR}_{\mathrm{mean}}$ from the fNIRS feature set, and $\mathrm{IBI}_{\mathrm{ratio}}$, $\mathrm{IBI}_{\mathrm{mean}}$, $\mathrm{RMSSD}_{\mathrm{ratio}}$, and RMSSD from the ECG feature set. The ROC curves for SVM, KNN, and LDA 3-level classification models based on fNIRS and ECG data are presented in Fig. 5.

V. Discussion

In this work, the added value of the combination of two simultaneously recorded sensing modalities, fNIRS and ECG, was investigated in the context of cognitive load classification on *n*-back task. The three machine learning methods, SVM, KNN, and LDA, were applied in order to assess the robustness of the conclusion related to our hypothesis, i.e. that conclusion is not dependent on a specific machine learning method. The *n*-back task consisted of 3 difficulties: 1-back task, 2-back task, and 3-back task. The results of the classification indicate that a combination of these modalities outperforms each modality alone when discriminating both three classes of cognitive load, as well as two classes of cognitive load, in all of the three applied machine learning methods.

We first observed the task performance metrics for each participant. No participant showed unusually poor performance accuracies on any of the tasks, thus no participant was excluded from the study. The performance accuracy averaged across all participants on each task level ranged from 95.18% for 1-back task to 89.35% on 3-back task. The response time averaged across all participants on each task level ranged from 0.886 s on 1-back task to 1.083 s on 3-back task. These performance metrics are in line with the previous literature findings, which indicate lower performance accuracies, i.e. more errors, and longer response time on more difficult tasks [19]. The performance accuracy and the response time were not used as the features in the later cognitive load classification.

We then showed that the 3-level cognitive load can be classified using only fNIRS and only ECG data collected during the *n*-back task with better than chance level classification, i.e. better than 33.3%. The accuracies ranged for different classifiers between 52.08% and 58.85% using only fNIRS data, and between 55.73% and 62.50% using only ECG data (Table I). The combination of both modalities improved all the classification models to accuracy of 63.54% for SVM classifier, 61.46% for KNN classifier, and 67.71% for LDA classifier. Using only features of average PFC activation, a study by Liu *et al.* [38] demonstrated the 3-level LDA model classification accuracy on *n*-back task of 53.9%. Also, a classification between resting periods and task periods in a study by Herff *et al.* [27] based only on fNIRS data, showed that workload induced even by relatively simple tasks can be robustly discriminated from a resting period. The same study also showed that the hemodynamic responses measured in the PFC are consistent enough to be used to discriminate between three levels of workload, which has been demonstrated in our study as well. With the addition of other modality, i.e. ECG, into our study, the mean accuracy of 3-level classification increased in all of the applied machine learning methods.

When analysing the classification accuracies in 2-level classifications (1-back task vs. 2-back task, 2-back task vs. 3-back task, 1-back task vs. 3-back task), the accuracies vary between 54.69% for fNIRS data and KNN classifier in 2-back task vs. 3-back task classification, up to 84.38% for the combination of fNIRS and ECG features for both SVM and LDA classifier in 1-back task vs. 3-back task. The fNIRS data, as well as the ECG data, collected during this experimental protocol allowed a better than chance classification accuracy on all of the two cognitive load conditions combination. The results in Table I show that the ECG features allow better discrimination in 2-back task vs. 3-back task compared with the only fNIRS data in all the performed machine learning methods, while in 1-back task vs. 3-back task, the results show better classification accuracies using fNIRS data compared with using only ECG data in all the performed methods. This finding illustrates that fNIRS-based features, related to the central nervous system, and ECG-based features, related to the autonomic nervous system, might differentially contribute to the accuracy of cognitive load classification at different difficulty levels.

In our study, the best classifier for 3-level classification is the LDA classifier, with the mean accuracy of 67.71%, following with the SVM classifier with the mean accuracy of 63.54%, and lastly the KNN classifier with the mean accuracy of 61.46%. It is worth noting that using only the fNIRS data allowed for the best classification with the SVM classifier (58.85%), and using only the ECG data, the highest accuracy was obtained with the LDA classifier (62.50%) on a 3-level classification problem. A few studies have already compared different classifiers on various tasks based on fNIRS data: Naseer *et al.* [49] compared LDA, quadratic discriminant analysis, KNN, Naive Bayes, SVM, and artificial neural networks (ANNs) classifiers on 2-level classification of mental arithmetic task and rest, while Qureshi *et al.* [50] compared KNN and ANNs on 2-level classification of mental arithmetic task and rest. The most accurate models in both studies were the ANNs. Rahman *et al.* [51] compared three classifiers (LDA, KNN, and SVM) on 2-class motor imagery task, concluding that the LDA classifier provided the highest classification accuracy among these three classifiers, which is in line with our study. A study by Keshmiri *et al.* [40] compared different machine learning classifiers on 2-level *n*-back task based on fNIRS features, obtaining the highest accuracies with their new suggested non-parametric approach. Using ECG combined with the EDA data, Posada-Quintero and Bolkhovsky [39] analysed KNN, SVM, decision trees, and discriminant analysis in cognitive load classification, concluding that the KNN classifier allowed the most accurate classification.

The multimodal combination of features improved the classification accuracies in both 2-level as well as 3-level classification models in all of the three applied machine learning methods. Although the introduction of new features has in some research led to decreasing the classification model accuracy [38], [52], by using the feature selection algorithm during the model training process in our study, the improvement in the classification accuracies is present when adding new features.

Regarding the feature ranking, the combination of fNIRS and ECG features is present in the 10 most important features from our 3-level classification models (Fig. 4) which emphasizes the improvement of the classification model based on multiple modalities. The 10 most important features in our 3-level classification models were: HbO_{mean}, $HbO_{dlPFCmean}$, $HbO_{LdlPFCmean}$, $HbO_{dlPFCratio}$, HbO_{peak}, and HbR_{mean} from the fNIRS feature set, and IBI_{ratio}, IBI_{mean}, $RMSSD_{ratio}$, and RMSSD from the ECG feature set (Fig. 4). Among the six most important features from the fNIRS feature set, three of them are related to the dlPFC activation, meaning that these features enhance the classification accuracies, which is an important conclusion since the features extracted solely from the dlPFC are more rarely used in the cognitive load research when compared with the features of the overall activation in the PFC [53].

Regarding the fNIRS modality, a variety of features can be computed and applied to cognitive load classification. To this point, there is no commonly used feature set derived from the fNIRS signals and peripheral physiology in cognitive load classification. Many features derived from the signals obtained with the fNIRS have been proposed but there is still a lack of consensus on the ideal feature set combination [17]. For example, a study by Naseer *et al.* [52] investigated the optimal feature set for LDA classification from the fNIRS signals. Furthermore, a study by Hong *et al.* [54] discussed the most commonly used features in fNIRS research, while Hwang *et al.* [55] indicated that the most commonly used features in the fNIRS research are the signal mean, slope and peak derived from the HbO and the HbR.

From the ECG features, the connection between IBI_{mean} and cognitive function is already reported in literature [56], [57], and used in the cognitive load classification on various tasks [24], [58]. The peripheral physiology features, in this case the ECG features, can also be accompanied with additional features obtained from other peripheral physiology signals, such as the EMG, EDA, and the respiration signals. Such combination of features computed from the peripheral physiology would allow more accurate assessment of cognitive load.

To summarize, the main contributions of our work, when compared to all the other research on cognitive load classification based on fNIRS and ECG data, are: the addition and comparison of different machine learning algorithms, which strengthen the robustness of our finding that a combination of fNIRS and ECG features increases the accuracy of cognitive load classification; identification of the most relevant fNIRS and ECG features responsible for obtained 3-level classification accuracies from a broader integrated set of features; demonstration of the importance of not only global PFC fNIRS features but also region-specific (dlPFC) fNIRS features in cognitive load classification; as well as the initial empirical evidence regarding differential contribution of central nervous system's fNIRS and autonomic nervous system's ECG features to the accuracy of cognitive load classification at different difficulty levels.

A few possible future directions in the classification of cognitive load based on the multimodal data can be pointed out. As already stated, adequate feature selection, which is a common research topic [52], [54], [55], is the prerequisite for cognitive load classification. In the fNIRS data analysis, the feature sets are mostly based on mean activation during the task, since that feature is marked as the most important feature in cognitive load classification in previous research [55]. However, in this research, the comparison of mean signal value between task periods and resting periods of the protocol was another feature derived from the mean signal activation value, which was, to the best of our knowledge, still not used as the feature for cognitive load classification. Another approach for increasing classification accuracies might be the addition of other modalities to the experimental protocol, particularly when aiming to classify multiple levels of cognitive load, judging from the results obtained in this article. Additionaly, another condition, i.e. 0-back task, could be added into the *n*-back task. Such protocol design would enable training and evaluation of a more sensitive classifier to different levels of cognitive load, as well as facilitate the study of the impact of different baseline periods (rest vs. 0-back task) on the contribution of baseline-normalised features to the classification accuracy.

This study demonstrated the potential of a multimodal neurophysiological approach to cognitive load assessment and classification. The presented results show the feasibility of using a multimodal approach in classifying multiple levels of cognitive load, induced by various difficulty levels of the same task. Such classification between task difficulties is of special interest in continually highly cognitively demanding and safety critical professions, and implementation of such integrated system in real time may allow more precise and accurate cognitive load monitoring. This article thus presents a good foundation for future translation of research results toward on-the-job application of cognitive load classification techniques. However, translation of laboratory research results into practice needs further studies demonstrating attainable accuracies of multimodal cognitive load classification in more realistic job-specific settings, while taking into account the obtrusiveness, cost and overall suitability of selected data collection methods for each targeted real-world occupational environment.

REFERENCES

[1] F. G. W. C. Paas and J. J. G. Van Merriánboer, "Instructional control of cognitive load in the training of complex cognitive tasks," *Educ. Psychol. Rev.*, vol. 6, no. 4, pp. 351–371, Dec. 1994.

[2] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. M. Van Gerven, "Cognitive load measurement as a means to advance cognitive load theory," *Educ. Psychol.*, vol. 38, no. 1, pp. 63–71, Mar. 2003.

[3] C. S. Jordan and S. D. Brennen, "Instantaneous self-assessment of workload technique (ISA)," Defence Research Agency, Portsmouth, U.K., Tech. Rep. DRA/TM (CAD5)/92011, 1992.

[4] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," *Adv. Psychol.*, vol. 52, pp. 139–183, Apr. 1988.

[5] H. Woo, B. Willems, J. Harrison, K. Izzetoglu, S. Hah, and H. Ayaz, "The use of functional near infrared spectroscopy (FNIRS) to assess cognitive workload of air traffic controllers," in *Proc. 17th Int. Symp. Aviation Psychol.*, 2013, p. 615.

[6] K. Cosic *et al.*, "New tools and methods in selection of air traffic controllers based on multimodal psychophysiological measurements," *IEEE Access*, vol. 7, pp. 174873–174888, 2019.

[7] G. Derosière, K. Mandrick, G. Dray, T. E. Ward, and S. Perrey, "NIRS-measured prefrontal cortex activity in neuroergonomics: Strengths and weaknesses," *Frontiers Hum. Neurosci.*, vol. 7, p. 583, Dec. 2013.

[8] G. E. Strangman, V. Ivkovic, and Q. Zhang, "Wearable brain imaging with multimodal physiological monitoring," *J. Appl. Physiol.*, vol. 124, no. 3, pp. 564–572, Mar. 2018.

[9] F. A. Fishburn, M. E. Norr, A. V. Medvedev, and C. J. Vaidya, "Sensitivity of fNIRS to cognitive state and load," *Frontiers Hum. Neurosci.*, vol. 8, p. 76, Feb. 2014.

[10] M. Causse, Z. Chua, V. Peysakhovich, N. Del Campo, and N. Matton, "Mental workload and neural efficiency quantified in the prefrontal cortex using fNIRS," *Sci. Rep.*, vol. 7, no. 1, pp. 1–15, Jul. 2017.

[11] G. Strangman, D. A. Boas, and J. P. Sutton, "Non-invasive neuroimaging using near-infrared light," *Biol. Psychiatry*, vol. 52, no. 7, pp. 679–693, Oct. 2002.

[12] W. B. Baker, A. B. Parthasarathy, D. R. Busch, R. C. Mesquita, J. H. Greenberg, and A. G. Yodh, "Modified beer-lambert law for blood flow," *Biomed. Opt. Express*, vol. 5, no. 11, pp. 4053–4075, 2014.

[13] N. Naseer and K.-S. Hong, "FNIRS-based brain-computer interfaces: A review," *Frontiers Hum. Neurosci.*, vol. 9, p. 3, Jan. 2015.

[14] G. Strangman, J. P. Culver, J. H. Thompson, and D. A. Boas, "A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation," *NeuroImage*, vol. 17, no. 2, pp. 719–731, Oct. 2002.

[15] X. Cui, S. Bray, D. M. Bryant, G. H. Glover, and A. L. Reiss, "A quantitative comparison of NIRS and fMRI across multiple cognitive tasks," *NeuroImage*, vol. 54, no. 4, pp. 2808–2821, Feb. 2011.

[16] F. Al-Shargie, M. Kiguchi, N. Badruddin, S. C. Dass, A. F. M. Hani, and T. B. Tang, "Mental stress assessment using simultaneous measurement of EEG and fNIRS," *Biomed. Opt. Express*, vol. 7, no. 10, pp. 3882–3898, Oct. 2016.

[17] T. Gateau, G. Durantin, F. Lancelot, S. Scannella, and F. Dehais, "Real-time state estimation in a flight simulator using fNIRS," *PLoS ONE*, vol. 10, no. 3, Mar. 2015, Art. no. e0121279.

[18] W. K. Kirchner, "Age differences in short-term retention of rapidly changing information," *J. Express Psychol.*, vol. 55, no. 4, p. 352, Apr. 1958.

[19] A. Meule, "Reporting and interpreting working memory performance in n-back tasks," *Frontiers Psychol.*, vol. 8, p. 352, Mar. 2017.

[20] I. Mijic, M. Sarlija, and D. Petrinovic, "MMOD-COG: A database for multimodal cognitive load classification," in *Proc. 11th Int. Symp. Image Signal Process. Anal. (ISPA)*, Sep. 2019, pp. 15–20.

[21] C. Mãhl, C. Jeunet, and F. Lotte, "EEG-based workload estimation across affective contexts," *Frontiers Neurosci.*, vol. 8, p. 114, Jun. 2014.

[22] J.-C. Croizet, G. Després, M.-E. Gauzins, P. Huguet, J.-P. Leyens, and A. Méot, "Stereotype threat undermines intellectual performance by triggering a disruptive mental load," *Personality Social Psychol. Bull.*, vol. 30, no. 6, pp. 721–731, Jun. 2004.

[23] A. S. Le, H. Aoki, F. Murase, and K. Ishida, "A novel method for classifying driver mental workload under naturalistic conditions with information from near-infrared spectroscopy," *Frontiers Human Neurosci.*, vol. 12, p. 431, Oct. 2018.

[24] A. Henelius, K. Hirvonen, A. Holm, J. Korpela, and K. Muller, "Mental workload classification using heart rate metrics," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Sep. 2009, pp. 1836–1839.

[25] L. Angell, M. A. Perez, and S. A. Soccolich, *Identification of Cognitive Load in Naturalistic Driving*. Blacksburg, VA, USA: Virginia Tech Transportation Institute, 2015.

[26] K. Khaksari, E. Condy, J. B. Millerhagen, A. A. Anderson, H. Dashtestani, and A. H. Gandjbakhche, "Effects of performance and task duration on mental workload during working memory task," *Photonics*, vol. 6, no. 3, p. 94, Aug. 2019.

[27] C. Herff, D. Heger, O. Fortmann, J. Hennrich, F. Putze, and T. Schultz, "Mental workload during n-back task—Quantified in the prefrontal cortex using fNIRS," *Frontiers Hum. Neurosci.*, vol. 7, p. 935, Jan. 2014.

[28] E. Molteni, M. Butti, A. M. Bianchi, and G. Reni, "Activation of the prefrontal cortex during a visual n-back working memory task with varying memory load: A near infrared spectroscopy study," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2008, pp. 4024–4027.

[29] A. A. Anderson *et al.*, "Exploring the role of task performance and learning style on prefrontal hemodynamics during a working memory task," *PLoS ONE*, vol. 13, no. 6, Jun. 2018, Art. no. e0198257.

[30] E. Molteni *et al.*, "Load-dependent brain activation assessed by time-domain functional near-infrared spectroscopy during a working memory task with graded levels of difficulty," *J. Biomed. Opt.*, vol. 17, no. 5, 2012, Art. no. 056005.

[31] H. Ayaz, P. A. Shewokis, S. Bunce, K. Izzetoglu, B. Willems, and B. Onaral, "Optical brain monitoring for operator training and mental workload assessment," *NeuroImage*, vol. 59, no. 1, pp. 36–47, Jan. 2012.

[32] T. Li, Q. Luo, and H. Gong, "Gender-specific hemodynamics in prefrontal cortex during a verbal working memory task by near-infrared spectroscopy," *Behavioural Brain Res.*, vol. 209, no. 1, pp. 148–153, May 2010.

[33] Y. Hoshi *et al.*, "Spatiotemporal characteristics of hemodynamic changes in the human lateral prefrontal cortex during working memory tasks," *NeuroImage*, vol. 20, no. 3, pp. 1493–1504, Nov. 2003.

[34] Y. Ogawa, K. Kotani, and Y. Jimbo, "Relationship between working memory performance and neural activation measured using near infrared spectroscopy," *Brain Behav.*, vol. 4, no. 4, pp. 544–551, Jul. 2014.

[35] M. Fallahi, R. Heidarimoghadam, M. Motamedzade, and M. Farhadian, "Psycho physiological and subjective responses to mental workload levels during N-back task," *J. Ergonom.*, vol. 06, no. 06, p. 12, 2016.

[36] K. Mandrick, V. Peysakhovich, F. Rémy, E. Lepron, and M. Causse, "Neural and psychophysiological correlates of human performance under stress and high mental workload," *Biol. Psychol.*, vol. 121, pp. 62–73, Dec. 2016.

[37] T. C. Hankins and G. F. Wilson, "A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight," *Aviat. Space Environ. Med.*, vol. 69, no. 4, pp. 360–367, Apr. 1998.

[38] Y. Liu, H. Ayaz, and P. A. Shewokis, "Multisubject Learning for mental workload classification using concurrent EEG, fNIRS, and physiological measures," *Frontiers Human Neurosci.*, vol. 11, p. 389, Jul. 2017.

[39] H. F. Posada-Quintero and J. B. Bolkhovsky, "Machine learning models for the identification of cognitive tasks using autonomic reactions from heart rate variability and electrodermal activity," *Behav. Sci.*, vol. 9, no. 4, p. 45, Apr. 2019.

[40] S. Keshmiri, H. Sumioka, R. Yamazaki, and H. Ishiguro, "A non-parametric approach to the overall estimate of cognitive load using NIRS time series," *Frontiers Hum. Neurosci.*, vol. 11, p. 15, Feb. 2017.

[41] Y. Liu, H. Ayaz, and P. A. Shewokis, "Mental workload classification with concurrent electroencephalography and functional near-infrared spectroscopy," *Brain-Comput. Interface*, vol. 4, no. 3, pp. 175–185, Mar. 2017.

[42] N. Z. Gurel, H. Jung, S. Hersek, and O. T. Inan, "Fusing near-infrared spectroscopy with wearable hemodynamic measurements improves classification of mental stress," *IEEE Sensors J.*, vol. 19, no. 19, pp. 8522–8531, Oct. 2019.

[43] R. C. Oldfield, "The assessment and analysis of handedness: The edinburgh inventory," *Neuropsychologia*, vol. 9, no. 1, pp. 97–113, Mar. 1971.

[44] H. Ayaz and B. Onaral, "Analytical software and stimulus-presentation platform to utilize, visualize and analyze near-infrared spectroscopy measures," M.S. thesis, School Biomed. Eng., Sci. Health Syst., Drexel Univ., Philadelphia, PA, USA, 2005.

[45] H. Ayaz, M. Izzetoglu, P. A. Shewokis, and B. Onaral, "Sliding-window motion artifact rejection for functional near-infrared spectroscopy," in *Proc. Annu. Int. Conf.*, Aug. 2010, pp. 6567–6570.

[46] M. Sarlija, F. Jurisic, and S. Popovic, "A convolutional neural network based approach to QRS detection," in *Proc. 10th Int. Symp. Image Signal Process. Anal.*, Sep. 2017, pp. 121–125.

[47] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers Public Health*, vol. 5, no. 258, Sep. 2017.

[48] A. Porta *et al.*, "Measuring regularity by means of a corrected conditional entropy in sympathetic outflow," *Biol. Cybern.*, vol. 78, no. 1, pp. 71–78, Jan. 1998.

[49] N. Naseer, N. K. Qureshi, F. M. Noori, and K. S. Hong, "Analysis of different classification techniques for two-class functional near-infrared spectroscopy-based brain-bomputer interface," *Comput. Intel. Neurosc.*, vol. 5, Sep. 2016, Art. no. 5480760.

[50] N. K. Qureshi, F. M. Noori, A. Abdullah, and N. Naseer, "Comparison of classification performance for fNIRS-BCI system," in *Proc. 2nd Int. Conf. Robot. Artif. Intell. (ICRAI)*, Nov. 2016, pp. 54–57.

[51] M. A. Rahman, F. Khanam, and M. Ahmad, "Detection of effective temporal window for classification of motor imagery events from prefrontal hemodynamics," in *Proc. Int. Conf. Electr., Comput. Commun. Eng. (ECCE)*, Feb. 2019, pp. 1–6.

[52] N. Naseer, F. M. Noori, N. K. Qureshi, and K.-S. Hong, "Determining optimal feature-combination for LDA classification of functional near-infrared spectroscopy signals in brain-computer interface application," *Frontiers Hum. Neurosci.*, vol. 10, p. 237, May 2016.

[53] R. McKendrick and A. Harwood, "Cognitive workload and workload transitions elicit curvilinear hemodynamics during spatial working memory," *Frontiers Hum. Neurosci.*, vol. 13, p. 405, Nov. 2019.

[54] K.-S. Hong, M. J. Khan, and M. J. Hong, "Feature extraction and classification methods for hybrid fNIRS-EEG brain-computer interfaces," *Frontiers Hum. Neurosci.*, vol. 12, Jun. 2018.

[55] H.-J. Hwang, J.-H. Lim, D.-W. Kim, and C.-H. Im, "Evaluation of various mental task combinations for near-infrared spectroscopy-based brain-computer interfaces," *J. Biomed. Opt.*, vol. 19, no. 7, Jul. 2014, Art. no. 077005.

[56] A. Luque-Casado, J. C. Perales, D. Cárdenas, and D. Sanabria, "Heart rate variability and cognitive processing: The autonomic response to task demands," *Biol. Psychol.*, vol. 113, pp. 83–90, Jan. 2016.

[57] S. Delliaux, A. Delaforge, J.-C. Deharo, and G. Chaumet, "Mental workload alters heart rate variability, lowering non-linear dynamics," *Frontiers Physiol.*, vol. 10, p. 565, May 2019.

[58] E. H. Magnusdottir, K. R. Johannsdottir, C. Bean, B. Olafsson, and J. Gudnason, "Cognitive workload classification using cardiovascular measures and dynamic features," in *Proc. 8th IEEE Int. Conf. Cognit. Infocommun.*, Sep. 2017, pp. 351–356.

Ivan Kesedžić received the B.Sc. and M.Sc. degrees in information and communication technology from the Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia, in 2014 and 2016, respectively, where he is currently pursuing the Ph.D. degree in computing.

He is a Research and Teaching Assistant with the Faculty of Electrical Engineering and Computing, University of Zagreb. His research interest includes signal and image processing. He has worked most extensively on brain-imaging analysis in cognitive tasks and resilience to stress prediction.

Marko Šarlija received the B.Sc. and M.Sc. degrees in electrical engineering and information technology from the Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia, in 2014 and 2016, respectively, where he is currently pursuing the Ph.D. degree in computer science.

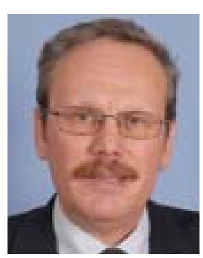
He is a Research and Teaching Assistant with the Faculty of Electrical Engineering and Computing, University of Zagreb. His research interests include system theory, physiological signal processing, affective computing, and machine learning applications, lying at the intersection with psychophysiology and neuroscience. He has worked most extensively in the area of physiology-based stress resilience prediction.

Jelena Božek (Member, IEEE) received the B.Sc. and Ph.D. degrees in electrical engineering from the Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia, in 2008 and 2013, respectively.

From 2013 to 2015, she has worked as a Postdoctoral Researcher on the Developing Human Connectome Project with the Oxford Centre for Functional MRI of the Brain (FMRIB), University of Oxford, U.K. She currently works as a Researcher with the Faculty of Electrical Engineering and Computing, University of Zagreb. Her research interests include brain MR image analysis and machine learning approaches in neuroimaging.

Siniša Popović (Member, IEEE) received the B.Sc. degree in mathematics and the M.Sc. and Ph.D. degrees in computer science from the Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia, in 2001, 2006, and 2011, respectively.

He is an Assistant Professor with the Faculty of Electrical Engineering and Computing, University of Zagreb. His research interests include affective computing, interactive simulation and virtual reality, as well as their intersections with psychology and neuroscience. He has mostly worked on applications of virtual reality and/or affective computing methods in psychotherapy of stress-related disorders and assessment of human resilience to stress.



Krešimir Ćosić received the Dipl.Ing., M.S., and Ph.D. degrees in electrical engineering from the Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia, in 1973, 1978, and 1984, respectively.

He is currently a Professor with the University of Zagreb and a Full Member with the Croatian Academy of Technical Sciences. He led a series of research projects financed by industry and different government agencies. He has published more than 100 papers in scientific journals and conference proceedings in the area of modeling and simulation, guidance and control, defense systems and technologies, interactive simulation systems, virtual reality, affective computing, and cyber psychology. He was also the Co-Director of a series of conferences in the framework of NATO Science for Peace programs.