



# Mental Workload Estimation from EEG Signals Using Machine Learning Algorithms

Baljeet Singh Cheema<sup>1</sup>, Shabnam Samima<sup>2</sup>, Monalisa Sarma<sup>2</sup>,  
and Debasis Samanta<sup>3</sup>(✉)

<sup>1</sup> Directorate General of Information System, Integrated Headquarters,  
Ministry of Defence (Army) Government of India, Kharagpur, India  
balema.155h@gov.in

<sup>2</sup> Subir Chowdhury School of Quality and Reliability,  
Indian Institute of Technology Kharagpur, Kharagpur, India  
{shabnam.samima,monalisa}@iitkgp.ac.in

<sup>3</sup> Department of Computer Science and Engineering,  
Indian Institute of Technology Kharagpur, Kharagpur, India  
dsamanta@sit.iitkgp.ernet.in

**Abstract.** Multitasking conditions prevalent in many environments such as critical operations in defense activities, evaluating user interfaces in man-machine interaction, etc. require assessment of mental workload of operators. However, mental workload (MWL) cannot be perceived directly as it is a complex and abstract property of human physiology. The techniques available in the literature for its assessment usually depend on *subjective analysis*, *performance analysis* and *psycho-physiological measurements*. But, these approaches despite being followed often prove to be inadequate due to high inter-personal variations and inconvenient procedures and subjective to the bias of evaluators. With the recent advancements of proliferation of Brain Computer Interface (BCI) devices and machine learning algorithms, it is possible to estimate MWL automatically. Nevertheless, there is a need to address issue like managing a high dimension and high volume data in real time. In this work, we propose an approach to estimate the mental workload using electroencephalogram (EEG) signals of an operator while in operation. A thorough investigation of different features, optimization of features and selecting an optimal number of channels are the some of the crucial steps have been addressed in this work. We propose a novel feature engineering method to extract a reduced set of features and utilize only a sub-set of channels for the purpose of classification of workload into different levels with the help of supervised machine learning techniques. Further, we investigate the performance of different classifiers and compare their results. It can be inferred from the observed results that mental workload estimation using machine learning algorithms is a better solution compared to the existing approaches.

**Keywords:** Electroencephalography  
Psycho-physiological measurement · Brain-computer interface  
Mental workload estimation · Machine learning algorithms  
 $n$ -back task · Dual  $n$ -back task

## 1 Introduction

Rapidly increasing growth and development in various industrial sectors like aviation, transport, military or space requires multitasking and continuous vigilance from operators to perform various jobs. This often over burdens the operators by placing huge mental workload upon them and leads to work-related stress and possibility of human errors. According to [7], Mental workload (MWL) refers to the amount of resources needed for processing of a certain task. It depends on characteristics of the task, the situation and the person. It is an abstract property of human-machine interaction which is not directly observable as there exists inherent difficulties in defining MWL and in understanding the factors which describe it in the best possible manner. It also poses difficulty in building a general/robust model for predicting performance. However, in the literature the level of workload has been inferred through three prime approaches, namely (1) subjective measures, (2) performance-based measures and (3) physiological measures [7]. Subjective approaches rely on the self assessment from the subjects about the difficulty of various tasks; performance-based measures depend on user performance for determining and assessing the cognitive state [12] and physiological methods attempt to interpret the cognitive workload with the help of invasive, semi-invasive and non-invasive physiological techniques.

Out of the above-mentioned categories, physiological measurements are comparatively better as they provide continuous and objective measurement of operator state. These measurements attempt to interpret the psychological processes through their effect on the body state, rather than through task performance or perceptual ratings. There are a number of diverse techniques available in the literature under this category [8]; however, each one of them is associated with some merits and demerits. In this regard, measurement technique such as ECoG (electrocorticography) provides better spatial and temporal resolution and better signal quality. But, it is a semi-invasive technique, which requires risky surgery. On the other hand, MEG (magnetoencephalography) is a non-invasive measurement technique, but incurs huge equipment cost and is not suited for everyday applications. fNRIS (functional magnetic resonance imaging) is relatively inexpensive and portable, but provides shallow spatial resolution of the order of few centimeters, while the time resolution of around 200 ms. An EEG (electroencephalography) based mental workload assessment which in earlier days utilized costly, wired and bulky devices posed serious limitations for application in real world applications. However, recent developments in brain-computer interfaces targeting real-life applications include *wireless EEG acquisition systems* that a person can easily wear while performing everyday activities. Of late, such a low cost, portable and wireless EEG devices have gained immense popularity for

studying cognitive workload [19] and vigilance task [17], as they allow for direct mental state assessment and because of their high temporal resolution, which is in the order of milliseconds. This makes EEG an appropriate tool for capturing fast and dynamically changing brain wave patterns in complex cognitive tasks. Besides, it seems that the use of wireless data acquisition systems to assess mental workload can enable more novel applications of mental workload measurement. This development supports exploring the feasibility of wireless data acquisition devices in MWL assessment [1,2]. Therefore, in this work we aim to:

- Explore the feasibility of wireless data acquisition devices in MWL assessment.
- Estimate MWL induced via various  $n$ -back and Dual  $n$ -back tasks and extract desired features using feature engineering.
- Study the effect of channel selection and feature optimization on classification performance.
- Investigate the capability of supervised machine learning algorithms for efficient classification of mental workload.
- Study the performance accuracy obtained using item-class classification.

In this work, we have used *Emotiv Epoc+* device to explore the feasibility of inexpensive EEG devices for assessing MWL and for collecting data. We performed a variety of  $n$ -back and Dual  $n$ -back tasks for inducing different levels of mental load on participants brain. Moreover, we utilized feature engineering step for extracting and selecting most effective features. Next, for classifying the MWL we have resorted to machine learning as it is a popular research field devoted to the development of inductive models, algorithms and procedures that can learn from data, extract trends and make predictions. The choice of machine learning algorithms is done from a pool of such algorithms, so as to have an optimal configuration of these algorithms. We have chosen seven different types of machine learning approaches, namely:

- Similarity based:  $K$ -nearest Neighbours
- Information based:
  - Random Forest
  - Decision Tree
- Error based:
  - Support Vector Machines
    - \* Linear
    - \* Radial Basis Function (RBF) Kernel
  - Multi Layer Perceptron
- Statistics based: Linear Discriminant Analysis

The rest of the paper is organized as follows: Sect. 2 presents the literature survey of works related with the classification of mental workload. Section 3 describes materials and methodology. Section 4 elaborates the process of EEG signal analysis. Next, we discuss about the obtained results in Sect. 5. Finally, Sect. 6 concludes the paper.

## 2 Literature Survey

In [5], authors estimated the mental workload, by using EEG features, for designing the intelligent learning systems. The developed workload index uses a Gaussian Process Regression model for predicting the workload of the individuals. The potential of both fNIRS and EEG (in combination) for classification of users' mental workload has been explored by the authors in [8]. In the recent years, rigorous efforts are being made to classify the mental workload into different levels. For example, in [14] classification of workload has been done using EEG based features. In [12], stepwise regression and multi-class linear classification has been utilized to extract statistical EEG features and to classify the workload into four levels. Authors in [20] have classified workload in seven levels by applying discrete wavelet transform and using artificial neural network (ANN). Further, in [9] EEG features that are sensitive enough to detect workload changes were identified. Variation of workload in different tasks has been found to be correlated with the EEG patterns in [4]. In [11], authors utilized cross-task performance based feature selection and regression model to classify mental workload. Binary classification of mental workload has successfully been done with the help of Fisher LDA and ERP based EEG features in [16]. Besides, the traditional mental workload assessment techniques were compared against the classification models built using machine learning approaches in [13].

## 3 Materials and Methodology

### 3.1 Subjects

Five healthy male and five healthy female volunteers participated in the experiment. The participants were between 20 and 24 years old, and except one, all were right handed. The participants were under-graduate and post-graduate students studying at the Indian Institute of Technology, Kharagpur. The participants had normal or corrected-to-normal vision. Further, participants were not on any medication and had no psychiatric or neurological disorders. Informed consent was taken from each participant before beginning of the experiment and were given liberty to select a time for the experiment in which they would feel *alert*. Moreover, the participants were also instructed to refrain from ingesting alcohol and/or sedative medications 24 h prior to the experiment and from caffeine and/or nicotine two hours prior to the experiment.

### 3.2 Data Acquisition and Experiment Protocol

The data collection has been carried out using the bluetooth enabled *Emotiv EPOC+* EEG device, having sampling rate of 128 Hz. The device comprises of 14 channels, namely AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8 and AF4 plus two references (P3/P4) and follows the international 10–20 standard for locations of electrodes. A minimum distance of five meters was maintained

between all power sources and the place of experiment. Further, the use of mobile phones was prohibited inside the laboratory during the experimentation. Two dedicated systems were used for the purpose of data collection, one for data recording and the other one for running the workload generating tasks. The tasks were run on a 21" all-in-one PC kept at a distance of 75 cm from the subjects. To minimize the artifacts originating from muscular movement, that is, due to electromyographic (EMG) activity, subjects were asked to avoid unnecessary physical movements during data recording. Furthermore, their hands were placed in a fixed position such that they could easily tap their fingers on the keyboard in response to the correct answer.

### 3.3 Workload Generating Task

We used the open source application, namely "Brain Workshop" [10] for generating the mental workload (MWL). The  $n$ -back task available in this application is a cognitive task which is mostly used as an assessment tool in cognitive neuroscience. The main advantage of  $n$ -back task is that it does not introduce any bias due to experience of an individual participating in the experiment. In other words, repetitive experiments with same participant introduces seldom bias. In this work, we have considered two variants of  $n$ -back task ( $n$ -back and dual  $n$ -back task) for MWL generation. The  $n$ -back/dual  $n$ -back tasks have all three ingredients of cognitive load namely:

- *Intrinsic load*, which is the load induced by the inherent nature of the task being processed. The inherent difficulty of task can be increased firstly by increasing the value of 'n' from 1 to 2 and secondly by migrating from  $n$ -back to dual  $n$ -back task.
- *Extraneous load*, which is induced by external factors like time pressure, noise, situation, work organization, etc. This type of load can be increased by reducing the time between the stimuli. In our experimental setup, we kept it constant at 3 s.
- *Germane load*, which is the load placed on working memory during schema formation and automation. Such a kind of load can be increased by increasing the value of 'n' which leads to increased amount of information required to be stored and processed in the working memory.

Further, we used five different tasks to generate five different load levels, namely idle, 1-back, 2-back, dual 1-back and dual 2-back in our experimentation. During the idle task, the participants were asked to remain still with eyes closed. In the 1-back and 2-back task scenarios, a  $3 \times 3$  grid was shown with stimuli appearing randomly at one of the grid locations on the screen. On the appearance of a stimuli, or trial, the participants were asked to respond whether or not the current stimulus is the same as the one that they saw  $n$  (that is, 1 or 2) presentations ago. Hence, for each trial, participants needed to memorize the previous  $n$  sequence of stimuli and perform a matching task mentally. Successfully, the dual  $n$ -back task involves remembering a sequence of spoken alphabet

and a sequence of positions of the stimuli at the same time, and identifying when an alphabet or position matches the one that appeared  $n$  trials back. Each task in the experiment had a total of 60 audio/visual stimuli (depending on various tasks) appearing after every three seconds.

### 3.4 Procedure

The experiments for data collection were conducted in an electrically isolated BCI laboratory under controlled environmental conditions so as to ensure adequate comfort to the participants. Here, we have performed experiments and tried to develop a method to classify mental workload not only when training and testing is done on the same task, but also when training and testing is done on different tasks. We utilized five distinct task levels in this experiment. Each participant performed all five levels of experiment successively.

Before beginning the experiment, each participant first filled the consent form and personal details form containing information about their age, gender, sleep duration, medication, status of mental health, education background, etc. Next, the experiment was started with the minimum load task, that is, idle task which was followed by the 1-back, 2-back, dual 1-back and dual 2-back tasks, respectively. In the  $n$ -back ( $n = 1$  or  $2$ ) task, the participants responded to the ‘position matching’ of the stimuli by pressing the alphabet ‘A’ from the keyboard if the position of the current stimulus matched to the position of stimulus presented  $n$ -trials back. While in the dual  $n$ -back ( $n = 1$  or  $2$ ), the participant pressed ‘A’ key for ‘position match’ and ‘L’ for ‘sound match’, respectively (refer Figs. 1 and 2). Switching from one task level to other was marked by a rest period of one minute. In each task level, a total of 60 trials/stimuli were presented, wherein each one appeared after every three seconds. EEG data recording for every load level of  $n$ -back task has been done for three minutes. Thus in total, for all levels, the duration of experiment was 20 min. The entire experiment protocol is graphically shown in Fig. 3.

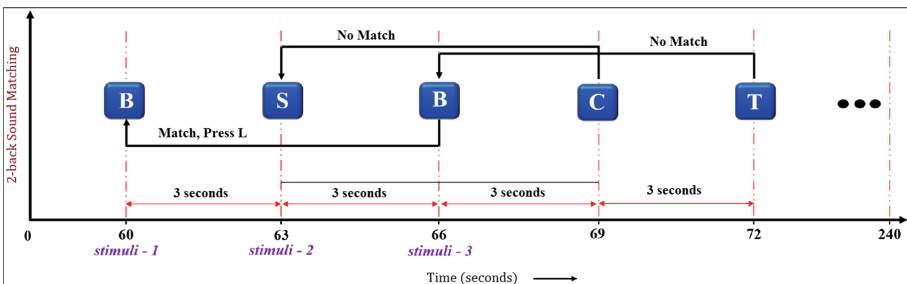
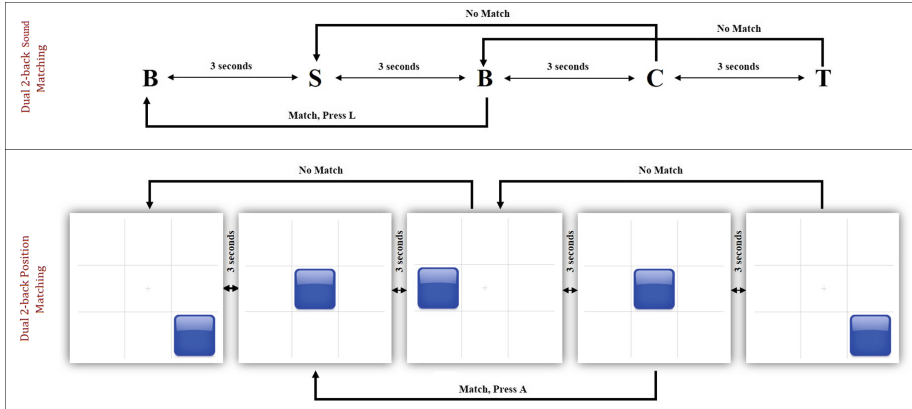
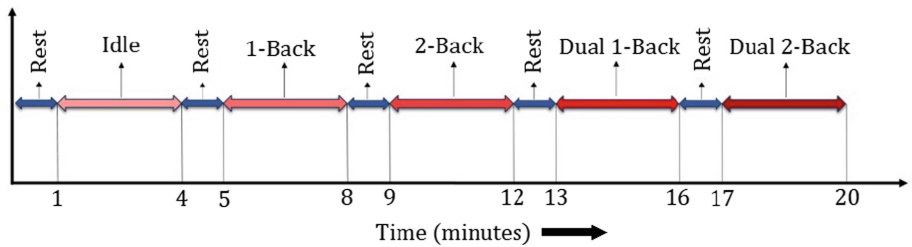


Fig. 1. An illustration to represent a 2-back task



**Fig. 2.** An illustration to represent a dual 2-back task



**Fig. 3.** Complete protocol

## 4 EEG Signal Analysis

The raw EEG signals captured through the scalp are contaminated with electrical signals and other undesired cerebral activities, which makes them unsuitable for feature extraction. These artifacts cause changes in the EEG measurements and severely degrade the useful signal of interest. Thus, it is necessary to process the EEG signals before we extract features from it. To process EEG signals, we begin with signal pre-processing phase which is followed by channel selection and feature extraction. Finally, we classify the data using the machine learning algorithms.

### 4.1 Signal Pre-processing

The recorded EEG signals are mostly, severely contaminated signals and are not actual brain signals. The contaminants are also known as artifacts. There are different kinds of artifacts such as power line noise, muscle contraction or electromyogram (EMG), heart activity or electrocardiogram (ECG), and eye movement or electrooculogram (EOG) [3]. These artifacts can be orders of magnitude

larger than the EEG signal. Therefore, the removal of artifacts is necessary to obtain the desired brain signals.

Many automated artifact removal methods have been proposed in the literature to remove artifacts from EEG recordings [18]. However, most of these methods either works well with additional EOG and EMG recordings or were designed to remove a single artifact. Hence, out of available artifact removal methods, we have used fully online and automated artifact removal tool for brain computer interfacing method (FORCe) [6] to remove all types of source generated artifacts. The clean data thus obtained (after the artifact removal phase) is further processed for baseline removal.

## 4.2 Channel Selection

Channel selection is done to choose the optimal subset of channels from the complete set of available channels. It is done to improve the model performance, provide faster processing, remove dimensionality curse, and to efficiently locate brain area that is responsible for neural activity.

In this work, we have used a very simple non-linear approach of channel selection which is called *Mutual Information (MI)*. It helps to evaluate non-linear dependencies between two or more random variables. Let  $X$  and  $Y$  be two random variables. Then, the *MI* between  $X$  and  $Y$  is the measure of amount of knowledge about  $Y$  which is provided by  $X$  and vice-versa. The *MI* between two random variables  $X$  and  $Y$  can be defined as:

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) \\ I(Y; X) &= H(Y) - H(Y|X) \\ I(X; Y) &= H(X) + H(Y) - H(X; Y) \end{aligned} \quad (1)$$

where,  $H(X)$  and  $H(Y)$  are the entropies of random variables  $X$  and  $Y$ , and  $H(X; Y)$  is their joint entropy. Their respective formulas are given here under.

$$\begin{aligned} H(X) &= - \int_X p_X(x) \log p_X(x) dx \\ H(Y) &= - \int_Y p_Y(y) \log p_Y(y) dy \\ H(X; Y) &= - \int_X \int_Y p_{X,Y}(x, y) \log p_{X,Y}(x, y) dx dy \end{aligned} \quad (2)$$

If *MI* between  $H(X)$  and  $H(Y)$  is zero, then  $X$  contains no information about random variable  $Y$  and vice-versa, which implies they are independent.

Based on *MI*, channels are either selected or rejected. We observed that AF3, F3, FC5, F7, F8, FC6, F4 and AF4 channels are confined to frontal lobe, which verifies the theory of cognition, according to which neuron activity related with cognitive workload is observed in the frontal lobe of human brain.

## 4.3 Feature Extraction

The feature extraction step involves extraction/selection of some distinctive components from the EEG signals. It is an extremely important step after signal



preprocessing and channel selection, as extraction of useful features is needed for classification of different levels of mental workload.

In this work, before extracting features, we divided the EEG data into epochs of length 3 s. Thus, for each epoch we obtained:  $14 \text{ channels} \times 128 \text{ Hz} \times 3 \text{ s} = 14 \times 384 = 5376$  samples. Further, to classify the mental workload, we have calculated six different categories of features from the EEG signals, which are briefly discussed below:

- *Statistical features:* As EEG signal is a time-series signal and it can easily be characterized by the distribution of the amplitude and its statistical features. Therefore, for each epoch of an EEG signal, we calculated different statistical features and tabulated them in Table 1.

**Table 1.** Statistical features

Features	Description
MEAN	Mean value
STD	Standard deviation
MAX_VALUE	Maximum positive amplitude
MIN_VALUE	Maximum negative amplitude
SKEWNESS	A measure of symmetry of the distribution
MEDIAN	The middle value of the set of ordered data
FD	Fractal dimension
AR	Auto regression

- *Derivative features:* Derivative features are obtained by calculating the mean of first and second derivative of EEG signals and the maximum value of the first and second derivative of EEG signals. The extracted features are shown in Table 2.

**Table 2.** Derivative features

Features	Description
1 <sup>st</sup> DIFF_MEAN	Mean value of the first derivative of the signal
1 <sup>st</sup> DIFF_MAX	Maximum value of the first derivative of the signal
2 <sup>nd</sup> DIFF_MEAN	Mean value of the second derivative of the signal
2 <sup>nd</sup> DIFF_MAX	Maximum value of the second derivative of the signal

- *Interval or period features:* EEG signals can also be analyzed by measuring the distribution of the intervals between zero and other level crossings or between maxima and minima. The calculated interval features are listed in Table 3.

**Table 3.** Interval or period features

Features	Description
LINE_LENGTH	Line length
MEAN_VV_AMPL	Mean of vertex to vertex amplitudes
VAR_VV_AMPL	Variance of vertex to vertex amplitudes
MEAN_VV_TIME	Mean of vertex to vertex times
MEAN_VV_SLOPE	Mean of vertex to vertex slope
VAR_VV_SLOPE	Variance of vertex to vertex slope
ZERO_CROSSING	Number of zero crossings in the signal
MIN_MAX_NUMBER	Number of local minima and maxima
COEFF_OF_VARIATION	A statistical measure of the deviation of a variable from its mean, standard deviation divided by mean
AMPL_RANGE	The difference between maximum positive and maximum negative amplitude values

- *Hjorth parameters:* Hjorth parameters gives an idea about the complexity of a time-series EEG signals. These values are very useful in EEG analysis and prove to be of great importance for its quantitative description. Refer Table 4 for the parameters.

**Table 4.** Hjorth parameters

Features	Description
HJORTH1	Ability
HJORTH2	Mobility $(\frac{\sigma_{x'}}{\sigma_x})$
HJORTH3	Complexity $(\frac{\frac{\sigma_{x''}}{\sigma_x'}}{\frac{\sigma_x}{\sigma_x'}})$

- *Frequency-domain features:* These features are one of the most important features for the analysis of EEG Signals. Based on the frequency content of the EEG signals, we extracted the features shown in Table 5 by applying the Fast Fourier Transform (FFT) to various EEG wave bands. Further, we also calculated other important ratios of FFT from various bands.
- *Wavelet features:* The wavelet transform (WT) is capable of distinguishing very small and delicate differences between time-series signals even from short signal epochs. It can easily identify highly irregular and non-stationary signals. Further, WT based methods can localize the signal components in time-frequency space in a better way than FFT analysis. Therefore, we evaluated the features listed in Table 6 using WT.

**Table 5.** Frequency-domain features

Features	Description
FFT_DELTA	0.1–4 Hz
FFT_THETA	4–8 Hz
FFT_ALPHA	8–13 Hz
FFT_BETA	13–30 Hz
FFT_GAMMA	30–40 Hz
FFT_WHOLE	.1–40 Hz
FFT_DT_RATIO	DELTA/THETA
FFT_DA_RATIO	DELTA/ALPHA
FFT_TA_RATIO	THETA/ALPHA
FFT_DTA_RATIO	(DELTA+THETA)/ALPHA
FFT_SEF	Spectral edge frequency
FFT_SP_ROLL_OFF	Below which 85% of the total spectral power resides

**Table 6.** Wavelet features

Features	Description
MIN_WAV_VALUE	Minimum value
MAX_WAV_VALUE	Maximum value
MEAN_WAV_VALUE	Mean value
MEDIAN_WAV_VALUE	Median value
STD_WAV_VALUE	Standard deviation
SKEWNESS_WAV_VALUE	Skewness
KURTOSIS_WAV_VALUE	Kurtosis
WAV_BAND	Relative energy
ENTROPY_SPECTRAL_WAV	The spectral entropy
1 <sup>st</sup> _DIFF_WAV_MEAN	Mean value of the 1 <sup>st</sup> derivative
1 <sup>st</sup> _DIFF_WAV_MAX	Maximum value of the 1 <sup>st</sup> derivative
2 <sup>nd</sup> _DIFF_WAV_MEAN	Mean value of the 2 <sup>nd</sup> derivative
2 <sup>nd</sup> _DIFF_WAV_MAX	Maximum value of the 2 <sup>nd</sup> derivative
ENERGY_PERCENT_WAV	Percentage of the total energy of a detail/approximation
WAV_ZERO_CROSSING	Zero crossing
WAV_COEFF_OF_VARIATION	Coefficient of variation
WAV_TOTAL_ENERGY	Total energy

#### 4.4 Feature Normalization and Optimization

The extracted features are normalized to bring them within a common range. This helps in feature optimization and reduces the inter-subject variability. Here, we have mean-normalized the extracted features using Eq. 3.

$$x_{new} = \frac{x - \mu}{\sigma} \tag{3}$$

where,  $\mu$  and  $\sigma$  denote *mean* and *standard deviation*, respectively.

Feature optimization also helps in minimizing the curse of dimensionality and enhanced generalization by reducing over-fitting. In feature optimization/selection, we identify data that are relevant to the selected parameters and assign them *maximum relevance*. We select those features which are strongly correlated to the classification and call this task as *maximum-relevance selection*. Besides, features which are mutually separated but have high degree of correlation to the classification are also selected and this task is referred to as *minimum-redundant selection*. These parameters are sometimes redundant and can be easily suppressed using maximum Relevance Minimum Redundancy (mRMR) algorithm [15]. Therefore, we have applied the mRMR algorithm to the extracted feature set to obtain the most optimized set of features. The features obtained after applying the optimization algorithm are tabulated in Table 7.

**Table 7.** Optimized features

Feature	Description
FD	Fractal dimension
AR	Auto regression
1 <sup>st</sup> _DIFF_MEAN	Mean value of the first derivative of the signal
1 <sup>st</sup> _DIFF_MAX	Maximum value of the first derivative of the signal
2 <sup>nd</sup> _DIFF_MEAN	Mean value of the second derivative of the signal
2 <sup>nd</sup> _DIFF_MAX	Maximum value of the second derivative of the signal
HJORTH1	Ability
HJORTH2	Mobility ( $\frac{\sigma'}{\sigma_x}$ )
HJORTH3	Complexity ( $\frac{\frac{\sigma''}{\sigma'} \cdot \frac{\sigma'}{\sigma_x}}{\frac{\sigma'}{\sigma_x}}$ )
FFT_DT_RATIO	$\frac{DELTA}{THETA}$
FFT_DA_RATIO	$\frac{DELTA}{ALPHA}$
FFT_TA_RATIO	$\frac{THETA}{ALPHA}$
FFT_DTA_RATIO	$\frac{DELTA+THETA}{ALPHA}$
WAV_COEFF_OF_VARIATION	Coefficient of variation
WAV_TOTAL_ENERGY	Total energy

## 5 Results and Discussion

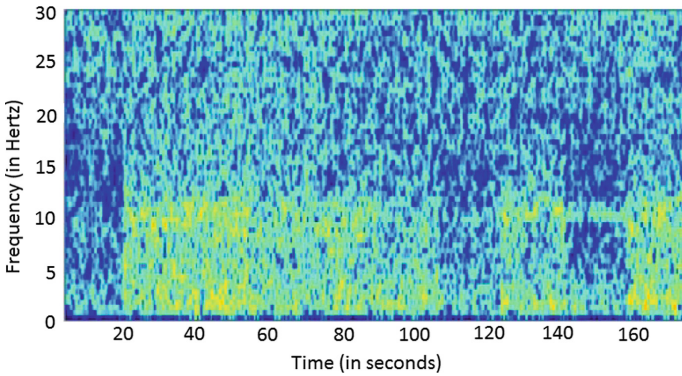
In this section, we present the spectrogram plots for the five levels of workload data. Next, we show the classification accuracy results for pre and post channel selection and feature extraction, respectively. At last, we present confusion matrix for pre and post channel selection and feature extraction, respectively.

For easy identification, we have labelled our cognitive workload level as  $C_i$ , where  $i = \{1,2,3,4,5\}$ , wherein  $C_1$  denotes *idle* task,  $C_2$  denotes 1-back task,  $C_3$  denotes 2-back task,  $C_4$  denotes dual 1-back task and  $C_5$  denotes dual 2-back task.

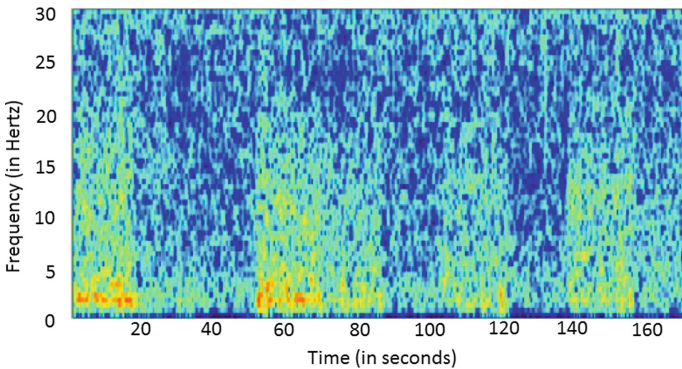
Various combinations of two-class, three-class, four-class and five-class classification for the above-mentioned categories have been summarized in the form of table and bar-chart. The obtained results are described next.

## 5.1 Spectrogram Plot

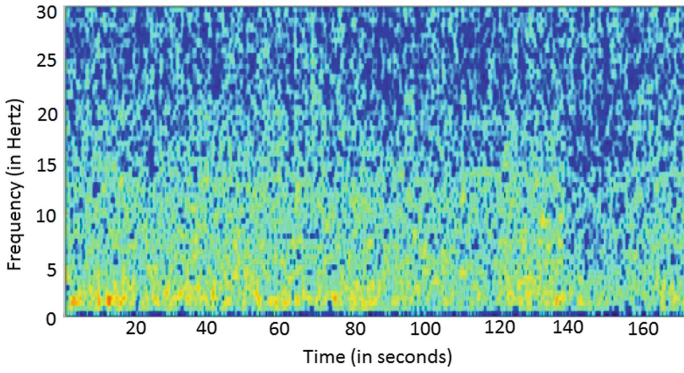
We have plotted spectrograms for all (five) levels of cognitive tasks for subject *M05*. From these plots (see Figs. 4, 5, 6, 7 and 8) we find/visualize the dominant EEG bands in a cognitive task. It can be clearly seen that *theta* and *alpha* wave activities are the most dominant in these spectrograms and possess most of the band power. Moreover, from the spectrograms for dual 2-back task one can notice that there is an increase in the beta band activity due to an increase in cognitive workload.



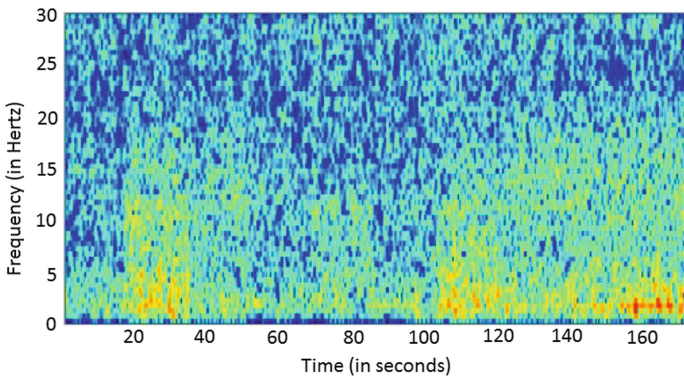
**Fig. 4.** Spectrogram plot for idle task



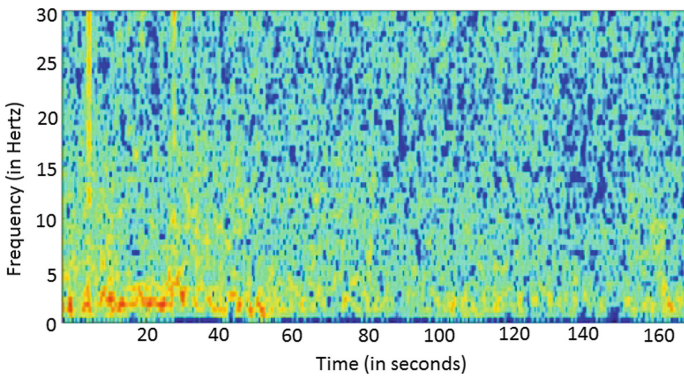
**Fig. 5.** Spectrogram plot for 1-back task



**Fig. 6.** Spectrogram plot for 2-back task



**Fig. 7.** Spectrogram plot for dual 1-back task



**Fig. 8.** Spectrogram plot for dual 2-back task

### 5.2 Classification Accuracy Using the Classifiers

Classification of the mental workload data into different levels has been done with the aid of seven supervised machine learning algorithms already named in Sect. 1. Each classifier model has been trained by dividing the complete dataset into a training set comprising of 80% data values and a test set comprising of remaining 20% data values. We have used *scikit-learn* open library for executing our machine learning algorithms. Further, for visualizing the effect of channel selection and feature optimization, we have carried out our classification in two different categories which are described next. In addition, we have also summarized various combinations of two-class, three-class, four-class and five-class classifications and summarized them in the form of Tables 8 and 9.

**Table 8.** Classification accuracy in (%) without channel selection and feature optimization

Class	Classifier						
Two-class	k-NN	Random forest	SVM (RBF)	SVM (Linear)	MLP	D-Tree	LDA
C <sub>1</sub> -C <sub>2</sub>	86.55	97.47	89.00	94.00	94.95	87.39	92.01
C <sub>1</sub> -C <sub>3</sub>	90.80	<b>97.99</b>	94.00	89.00	96.39	90.80	90.40
C <sub>1</sub> -C <sub>4</sub>	92.40	96.39	94.00	92.00	94.39	90.00	91.20
C <sub>1</sub> -C <sub>5</sub>	89.55	95.58	91.00	86.00	93.57	91.16	85.14
C <sub>2</sub> -C <sub>3</sub>	81.09	89.91	83.00	86.00	89.70	77.31	81.51
C <sub>2</sub> -C <sub>4</sub>	92.85	95.37	88.00	89.00	93.69	86.55	92.43
C <sub>2</sub> -C <sub>5</sub>	84.81	92.40	84.00	91.00	92.82	81.85	88.60
C <sub>3</sub> -C <sub>4</sub>	77.60	84.39	74.00	72.00	82.39	73.99	78.40
C <sub>3</sub> -C <sub>5</sub>	81.92	86.74	75.00	75.00	82.32	73.49	79.91
C <sub>4</sub> -C <sub>5</sub>	83.13	86.34	80.00	84.00	88.35	77.91	83.93
Average	86.07	92.258	85.2	85.8	90.857	83.045	86.353
Three-class	k-NN	Random forest	SVM (RBF)	SVM (Linear)	MLP	D-Tree	LDA
C <sub>1</sub> -C <sub>2</sub> -C <sub>3</sub>	77.41	<b>91.46</b>	73.00	82.00	84.29	79.06	82.09
C <sub>1</sub> -C <sub>3</sub> -C <sub>5</sub>	76.47	87.96	73.00	75.00	82.08	68.71	71.92
C <sub>2</sub> -C <sub>3</sub> -C <sub>4</sub>	66.94	79.33	62.00	70.00	75.75	67.49	70.79
C <sub>3</sub> -C <sub>4</sub> -C <sub>5</sub>	66.84	82.62	65.00	66.00	71.65	66.04	67.11
Average	71.915	85.3425	68.25	73.25	78.4425	70.325	72.9775
Four-class	k-NN	Random forest	SVM (RBF)	SVM (Linear)	MLP	D-Tree	LDA
C <sub>1</sub> -C <sub>2</sub> -C <sub>3</sub> -C <sub>4</sub>	65.98	81.55	64.00	68.00	76.02	67.82	71.31
C <sub>2</sub> -C <sub>3</sub> -C <sub>4</sub> -C <sub>5</sub>	67.14	79.87	60.00	63.00	72.07	60.36	62.42
C <sub>1</sub> -C <sub>2</sub> -C <sub>4</sub> -C <sub>5</sub>	71.66	<b>86.44</b>	70.00	75.00	82.54	68.58	76.79
Average	68.26	82.62	64.6667	68.6667	76.87667	65.58667	70.1733
Five-class	k-NN	Random forest	SVM (RBF)	SVM (Linear)	MLP	D-Tree	LDA
C <sub>1</sub> -C <sub>2</sub> -C <sub>3</sub> -C <sub>4</sub> -C <sub>5</sub>	61.43	<b>80.22</b>	57.00	62.00	68.13	58.33	63.39

**Table 9.** Classification accuracy in (%) with channel selection and feature optimization

Class	Classifier						
	k-NN	Random forest	SVM (RBF)	SVM (Linear)	MLP	D-Tree	LDA
Two-class	k-NN	Random forest	SVM (RBF)	SVM (Linear)	MLP	D-Tree	LDA
C <sub>1</sub> -C <sub>2</sub>	95.19	<b>99.19</b>	95.00	88.00	95.99	92.00	95.19
C <sub>1</sub> -C <sub>3</sub>	93.60	95.19	96.00	95.00	94.39	91.20	96.79
C <sub>1</sub> -C <sub>4</sub>	95.19	93.60	97.00	97.00	99.19	88.00	97.59
C <sub>1</sub> -C <sub>5</sub>	94.35	95.96	94.00	92.00	96.77	87.90	91.12
C <sub>2</sub> -C <sub>3</sub>	83.19	92.80	82.00	94.00	86.39	85.59	91.20
C <sub>2</sub> -C <sub>4</sub>	95.19	96.79	90.00	92.00	92.00	91.20	90.40
C <sub>2</sub> -C <sub>5</sub>	86.29	94.35	88.00	95.00	93.54	84.67	91.12
C <sub>3</sub> -C <sub>4</sub>	81.59	83.99	75.00	76.00	80.00	74.39	80.80
C <sub>3</sub> -C <sub>5</sub>	79.03	88.70	78.00	76.00	76.61	75.00	76.61
C <sub>4</sub> -C <sub>5</sub>	80.64	87.90	82.00	87.00	83.06	81.45	83.87
Average	88.426	92.928	87.7	89.2	89.794	85.14	89.469
Three-class	k-NN	Random forest	SVM (RBF)	SVM (Linear)	MLP	D-Tree	LDA
C <sub>1</sub> -C <sub>2</sub> -C <sub>3</sub>	83.51	<b>91.48</b>	86.00	89.00	88.82	82.97	86.70
C <sub>1</sub> -C <sub>3</sub> -C <sub>5</sub>	81.28	86.63	83.00	83.00	87.16	75.40	86.09
C <sub>2</sub> -C <sub>3</sub> -C <sub>4</sub>	75.00	<b>91.48</b>	71.00	78.00	80.85	82.44	77.65
C <sub>3</sub> -C <sub>4</sub> -C <sub>5</sub>	66.31	79.67	66.00	69.00	76.47	65.77	74.33
Average	76.525	87.315	76.5	79.75	83.325	76.645	81.1925
Four-class	k-NN	Random forest	SVM (RBF)	SVM (Linear)	MLP	D-Tree	LDA
C <sub>1</sub> -C <sub>2</sub> -C <sub>3</sub> -C <sub>4</sub>	71.59	87.60	73.00	77.00	85.19	77.20	79.20
C <sub>2</sub> -C <sub>3</sub> -C <sub>4</sub> -C <sub>5</sub>	67.87	82.73	64.00	73.00	73.09	69.07	78.71
C <sub>1</sub> -C <sub>2</sub> -C <sub>4</sub> -C <sub>5</sub>	74.29	<b>90.76</b>	74.00	84.00	89.15	73.89	82.32
Average	71.25	87.03	70.3333	78	82.47667	73.3867	80.07667
Five-class	k-NN	Random forest	SVM (RBF)	SVM (Linear)	MLP	D-Tree	LDA
C <sub>1</sub> -C <sub>2</sub> -C <sub>3</sub> -C <sub>4</sub> -C <sub>5</sub>	64.42	<b>84.61</b>	63.00	73.00	79.80	69.55	78.20

**Classification without Channel Selection and Feature Optimization:**

From the obtained results (refer Table 8) it can be observed that the Random Forest algorithm gives the best classification accuracy for all combinations of classes. It can be noted that, this classifier presents highest accuracy of 97.22% for two-class classification followed by percentage accuracy of 91.46, 86.44 and 80.22 for the combination of three, four and five classes, respectively.

**Classification with Channel Selection and Feature Optimization:**

After channel selection and feature optimization, it has been observed that the average classification accuracy increases for all the classifiers involved (refer Table 9). Further, it has been observed that the Random Forest classifier outperforms all other classifiers. Highest accuracy obtained with Random Forest is 99.19% in two-class classification followed by percentage accuracy of 91.48%, 90.76% and 84.61% for three, four and five classes, respectively.



### 5.3 Confusion Matrix

For efficiently depicting the accuracy of classification, we have shown the obtained results with the help of confusion matrix. Each column of the matrix represents the instances in a predicted class while each row represents the instances in a true class (or vice-versa). The diagonal elements represent the number of points for which the predicted class is same as the true class, while off-diagonal elements are those which are misclassified by the classifier. Higher diagonal values of the confusion matrix indicates better predictions. We present confusion matrix in two categories which are discussed next.

**Confusion Matrix Before Channel Selection and Feature Optimization:** Confusion matrix before channel selection and feature optimization techniques for two-class, three-class, four-class and five-class classification is shown in Fig. 9.

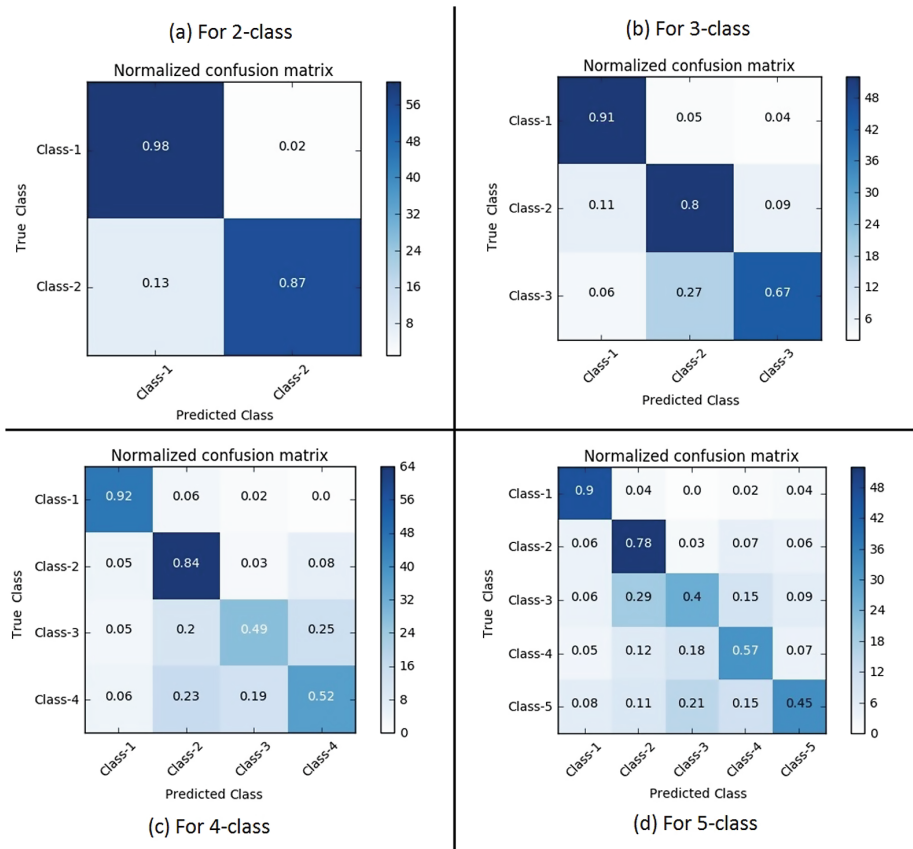
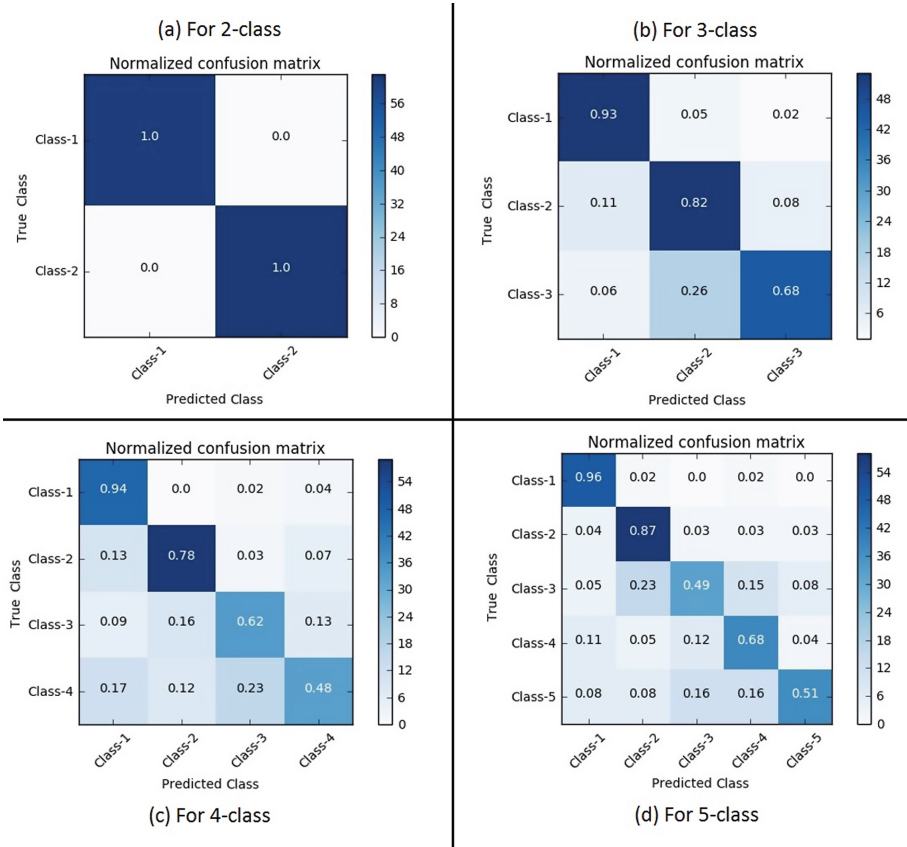


Fig. 9. Confusion matrix before channel selection and feature optimization:

**Confusion Matrix After Channel Selection and Feature Optimization:**

Confusion matrix after channel selection and feature optimization techniques for two-class, three-class, four-class and five-class classification is shown in Fig. 10.



**Fig. 10.** Confusion matrix after channel selection and feature optimization

On comparing the matrices for the two cases, we can observe that there is a substantial improvement in classification accuracy with the usage of channel selection and feature optimization for all class labels. For instance, we can note that, in class-1 to class-1 matching from both cases, the accuracy increases from 98% to 100% in two class classification, 91% to 93% in three class classification, from 92% to 94% in four class classification and from 90% to 96% in five class classification.

## 6 Conclusion

In this paper, first, we explored the feasibility of wireless data acquisition devices in mental workload assessment with the help of  $n$ -back task. From this it is evident that these devices have enormous potential which may be exploited in everyday environment and can be of utmost importance to handle critical situations such as monitoring pilots of flights, nuclear operations, driving tasks, etc. Second, we modeled and evaluated MWL induced during human-computer interaction with the help of features extracted from EEG signals. Third, we investigated the potential of machine learning to classify MWL in different levels. To accomplish this, we have used different categories of supervised machine learning algorithms that can learn from the data (about its pattern) and give predictions. Fourth, we studied the effect of channel selection and feature optimization on classification performance. From the obtained results, it can be easily observed that the Random Forest algorithm results in best accuracy in comparison to all the other compared algorithms. Further, we also studied the performance accuracy obtained due to inter-class classification. We hope that this study would be helpful in future to explore and devise new methods for studying and understanding cognitive workload.

## References

1. Anderson, E.W., Potter, K.C., Matzen, L.E., Shepherd, J.F., Preston, G.A., Silva, C.T.: A user study of visualization effectiveness using EEG and cognitive load. *Comput. Graph. Forum* **30**(3), 791–800 (2011)
2. Ayaz, H., Onaral, B., Izzetoglu, K., Shewokis, P.A., McKendrick, R., Parasuraman, R.: Continuous monitoring of brain dynamics with functional near infrared spectroscopy as a tool for neuroergonomic research: empirical examples and a technological development. *Front. Human Neurosci.* **7**, 871 (2013)
3. Benbadis, S.R.: EEG artifacts. <http://emedicine.medscape.com/article/1140247-overview#a3>. Accessed 14 Feb 2018
4. Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Olmstead, R.E., Tremoulet, P.D., Craven, P.L.: EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviat. Sp. Environ. Med.* **78**(5), B231–B244 (2007)
5. Chaouachi, M., Jraidi, I., Frasson, C.: Modeling mental workload using EEG features for intelligent systems. In: Konstan, J.A., Conejo, R., Marzo, J.L., Oliver, N. (eds.) *UMAP 2011*. LNCS, vol. 6787, pp. 50–61. Springer, Heidelberg (2011). [https://doi.org/10.1007/978-3-642-22362-4\\_5](https://doi.org/10.1007/978-3-642-22362-4_5)
6. Daly, I., Scherer, R., Billinger, M., Müller-Putz, G.: Force: fully online and automated artifact removal for brain-computer interfacing. *IEEE Trans. Neural Syst. Rehabil. Eng.* **23**(5), 725–736 (2015)
7. Heine, T., Lenis, G., Reichensperger, P., Beran, T., Doessel, O., Deml, B.: Electrocardiographic features for the measurement of drivers' mental workload. *Appl. Ergon.* **61**, 31–43 (2017)

8. Hirshfield, L.M., Chauncey, K., Gulotta, R., Girouard, A., Solovey, E.T., Jacob, R.J.K., Sassaroli, A., Fantini, S.: Combining electroencephalograph and functional near infrared spectroscopy to explore users' mental workload. In: Schmorrow, D.D., Estabrooke, I.V., Grootjen, M. (eds.) FAC 2009. LNCS (LNAI), vol. 5638, pp. 239–247. Springer, Heidelberg (2009). [https://doi.org/10.1007/978-3-642-02812-0\\_28](https://doi.org/10.1007/978-3-642-02812-0_28)
9. Holm, A., Lukander, K., Korpela, J., Sallinen, M., Müller, K.M.I.: Estimating brain load from the EEG. *Sci. World J.* **9**, 639–651 (2009)
10. Hoskinson, P.: Brain workshop - a dual n-back game. <http://brainworkshop.sourceforge.net/>. Accessed 14 Feb 2018
11. Ke, Y., Qi, H., He, F., Liu, S., Zhao, X., Zhou, P., Zhang, L., Ming, D.: An EEG-based mental workload estimator trained on working memory task can work well under simulated multi-attribute task. *Front. Hum. Neurosci.* **8**, 703 (2014)
12. Mahmoud, R., Shanableh, T., Bodala, I.P., Thakor, N., Al-Nashash, H.: Novel classification system for classifying cognitive workload levels under vague visual stimulation. *IEEE Sens. J.* **17**, 7019–7028 (2017)
13. Moustafa, K., Luz, S., Longo, L.: Assessment of mental workload: a comparison of machine learning methods and subjective assessment techniques. In: Longo, L., Leva, M.C. (eds.) H-WORKLOAD 2017. CCIS, vol. 726, pp. 30–50. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-61061-0\\_3](https://doi.org/10.1007/978-3-319-61061-0_3)
14. Mühl, C., Jeunet, C., Lotte, F.: EEG-based workload estimation across affective contexts. *Front. Neurosci.* **8**, 114 (2014)
15. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(8), 1226–1238 (2005)
16. Roy, R.N., Bonnet, S., Charbonnier, S., Campagne, A.: Efficient workload classification based on ignored auditory probes: a proof of concept. *Front. Hum. Neurosci.* **10**, 519 (2016)
17. Samima, S., Sarma, M., Samanta, D.: Correlation of P300 ERPS with visual stimuli and its application to vigilance detection. *IEEE*, July 2017
18. Urigüen, J.A., Garcia-Zapirain, B.: EEG artifact removal-state-of-the-art and guidelines. *J. Neural Eng.* **12**(3), 031001 (2015)
19. Wang, S., Gwizdka, J., Chaovaitwongse, W.A.: Using wireless EEG signals to assess memory workload in the *n*-back task. *IEEE Trans. Hum.-Mach. Syst.* **46**(3), 424–435 (2016)
20. Zarjam, P., Epps, J., Chen, F., Lovell, N.H.: Estimating cognitive workload using wavelet entropy-based features during an arithmetic task. *Comput. Biol. Med.* **43**(12), 2186–2195 (2013)