

Article

# Towards Intelligent Data Analytics: A Case Study in Driver Cognitive Load Classification

Shaibal Barua , Mobyen Uddin Ahmed  and Shahina Begum

School of Innovation, Design and Engineering, Mälardalen University, Högscoleplan 1, 72220 Västerås, Sweden; mobyen.ahmed@mdh.se (M.U.A.); shahina.begum@mdh.se (S.B.)

\* Correspondence: shaibal.barua@mdh.se

Received: 15 June 2020; Accepted: 29 July 2020; Published: 6 August 2020



**Abstract:** One debatable issue in traffic safety research is that the cognitive load by secondary tasks reduces primary task performance, i.e., driving. In this paper, the study adopted a version of the n-back task as a cognitively loading secondary task on the primary task, i.e., driving; where drivers drove in three different simulated driving scenarios. This paper has taken a multimodal approach to perform ‘intelligent multivariate data analytics’ based on machine learning (ML). Here, the k-nearest neighbour (k-NN), support vector machine (SVM), and random forest (RF) are used for driver cognitive load classification. Moreover, physiological measures have proven to be sophisticated in cognitive load identification, yet it suffers from confounding factors and noise. Therefore, this work uses multi-component signals, i.e., physiological measures and vehicular features to overcome that problem. Both multiclass and binary classifications have been performed to distinguish normal driving from cognitive load tasks. To identify the optimal feature set, two feature selection algorithms, i.e., sequential forward floating selection (SFFS) and random forest have been applied where out of 323 features, a subset of 42 features has been selected as the best feature subset. For the classification, RF has shown better performance with  $F_1$ -score of 0.75 and 0.80 than two other algorithms. Moreover, the result shows that using multicomponent features classifiers could classify better than using features from a single source.

**Keywords:** cognitive load; machine learning; multimodal data analytics; multicomponent signals

## 1. Introduction

Driving a vehicle requires dynamic adjustment of cognitive control, here, both visual and physical tasks are crucial to keep the driving performance to an acceptable level within a comfortable effort [1]. While driving a vehicle, drivers are often occupied with many other activities such as using a mobile phone, listening to the radio, or having a conversation with a passenger, etc. Moreover, new advanced in-vehicle information systems embedded in the modern vehicles could create distracted driving scenarios and may affect the driving performance [2–4]. Thus, these secondary activities, i.e., activities not related to driving require extra cognitive processes in ways that the driver can still keep their eyes on the road and hands on the steering wheel while being involved in other activities at the same time, and this refers to the ‘cognitive load activities’. It is reported that more than 90% of traffic crashes are assigned to the driver’s error, whereas 41% of them are due to inattention, distraction, and cognitive load activities [5]. Further, the risk concerning traffic safety and driving performance anticipating cognitive load activities have been addressed in [6,7].

Many studies have been pursued to understand the consequences of secondary task or dual-task demands while driving and different types of data such as physiological, driving behavioural, and subjective measures have been used to evaluate the driver’s mental effort [1,8–10]. In this paper, the attention selection model (ASM) [11], based on the n-back task has been employed to impose

the cognitively loading secondary task while driving. The ASM is a conceptual model of attention selection and multitasking in everyday natural driving situations. The n-back task is a continuous performance task commonly used as an assessment of working memory load [12,13]. The n-back task can vary on the task difficulty or complexity, such as very mild task demand (0-back), moderate task demand (1-back), and a high level of task demand (2-back) [14]. Since it is difficult to characterise the task demand from the invested effort on the secondary task by the drivers [1], the alternative is to use physiological measures and this work aims to explore the classification of cognitive load activities i.e., the n-back task events.

Several physiological signals such as electroencephalography (EEG), electrooculography (EOG), electrocardiogram (ECG), heart rate (HR), and heart rate variability (HRV) (the variation in the beat to beat intervals of the heart), galvanic skin response (GSR), and respiration rate (RR) signal have become object measures of cognitive load. Several studies have also considered the variations in drivers' behavioural data obtained from vehicular signals for drivers' cognitive load classification [15,16]. Machine learning methods that have been used for detecting the driver state, e.g., cognitive load include SVM [17–25], artificial neural network [2,26–29], random forest [30,31], deep learning [32,33], and case-based reasoning [34–37]. The performance of cognitive load classification is often poor when there are uncertainties—such as participants failing to perform some task or, in a real-time system. The accuracy could be improved, for example, using a suitable window size, which influences the delay that occurs between the onset of a cognitive load task and when changes are detected in the driver's performance due to the higher cognitive load [2,21]. Several studies use multimodal data in machine learning models to classify driver states [2,38,39].

This paper aims for intelligent data analytics using ML to discriminate the cognitive load task from normal driving in different driving situations such as visual cue in traffic, incoming traffic, and traffic environment. The cognitive load task classification based on physiological measures have become practical and they have shown much potential because of their granularity and high degree of responsiveness. However, in many occasions physiological measures suffer from both confounding factors and noise [40]. Hence, this paper focuses on a multimodal approach based on multicomponent signals to measure the cognitive load and to classify different cognitive load tasks. Here, several ML algorithms, e.g., k-nearest neighbour (k-NN), support vector machine (SVM), and random forest (RF) are applied for the classification. Again, two feature selection algorithms, i.e., sequential forward floating selection (SFFS) and random forest with mean decrease accuracy (MDA) are used to identify an optimal feature set. In this paper, physiological signals, i.e., EEG, EOG, GSR, ECG, and RR are fused as a driver behavioural data and combined with driving context, i.e., driving conditions in the scenarios (see Section 2). Again, a number of vehicular features, e.g., lateral speed, steering wheel angle (SWA), yaw and yaw rate, and lateral position are also being considered. Considering the study design, both multiclass and binary classification have been performed, where classifiers are trained using 5-fold cross-validation on the training datasets and finally, evaluated with the test datasets.

## 2. Materials and Methods

### 2.1. Study Design and Data Set

The experimental study took place at the Swedish National Road and Transport Research Institute (VTI), Linköping, Sweden, using a high-fidelity moving-base driving simulator (VTI Driving Simulator III (<https://www.vti.se/en/research-areas/vtis-driving-simulators/>)), see Figure 1. The study was approved by the regional ethics committee at Linköping University, (Dnr 2014/309-31) and each participant signed an informed consent form. The simulator was a car cabin consisting of front seats of a SAAB 9-3 with automatic transmission. It could simulate the movements and forces by moving, rotating, or tilting the part of the simulator with projector screens. A vibration table enables the simulation of the road surface contact. It had three liquid-crystal display systems for rear mirrors and six projectors for visualization of the frontal view with a horizontal field of view of 120 degrees.

The study that collected the cognitive load dataset consisted of two test series that contained recordings from 66 participants (33 in test series 1 and 33 in test series 2). All the participants were male with no known diseases or medications, aged between 35 and 50 ( $42.47 \pm 4.39$  years), and had held a valid driver's license for more than ten years. To obtain homogeneity, only males were chosen with the aforementioned criteria. Further, participants were not professional drivers (e.g., taxi and heavy vehicle driver), no extremes in terms of self-reported personalities (extrovert or introvert), and self-reported normal sensitivity to stressful situations. To assess stress tolerance, each participant had to fill up a questionnaire after the end of each driving session. The self-reported questionnaire used a scale of 0–6, where '0' means low-stress tolerance and '6' means the high-stress tolerance; whereas for anxiety '0' indicates low and '6' indicates high anxiety. However, the personality and stress sensitivity have not been taken into consideration in this paper.

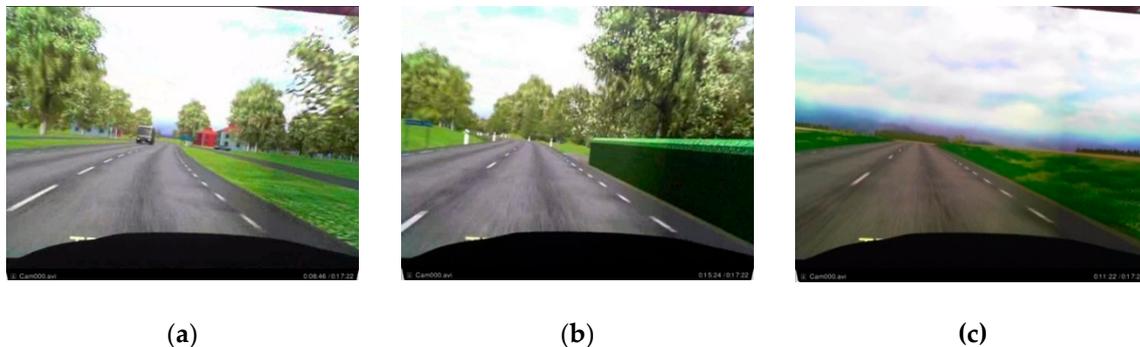


**Figure 1.** VTI simulator III and electroencephalography (EEG) electrodes setup on a participant.

The driving environment in the simulator consisted of three recurring scenarios in which the simulated road was a rural road with one lane in each direction, some curves and slopes, and a speed limit of 80 km/h. The three scenarios were (1) four-way crossing with an incoming bus and a car approaching the crossing from the right (CR), (2) a hidden exit on the right side of the road with a warning sign (HE), and (3) a strong side wind in open terrain (SW). Figure 2 represents examples of these study scenarios. In Figure 2, the image in the left shows the CR scenario, the middle one shows the HE scenario, and the rightmost image shows the SW scenario. Thus, these scenarios implied threats in off-path locations without requiring the drivers to change their responses. As a within-measure study, each scenario was repeated four times during the approximately 40 min driving session where the participants were involved either in a cognitive load task, i.e., a 1-back or 2-back task, or were driving to pass a scenario (baseline or no-task). In the first test series, participants performed the normal driving and 1-back task while driving. However, in the second test series, the participants performed all three task conditions in the hidden exit and four-way crossing scenarios. The no-task and 2-back tasks were only performed under the side wind in the open field scenario. The 1-back and 2-back tasks are considered as the secondary auditory tasks, where a number is orally presented through the simulator's speakers at an interval of 2 s. The participants had to respond whenever the last presented number was the same as the previous one (1-back) or two steps earlier (2-back).

The physiological signals were acquired using a multi-channel amplifier with active electrodes (g.HIamp, g.tec Medical Engineering GmbH, Austria). The electroencephalography (EEG) electrodes were positioned based on the 10–20 system providing a 30-channel recording. The EEG signals were band-pass filtered between 0.5 and 60 Hz using an 8th order Butterworth filter, and frequencies between 48 and 52 Hz were removed using a 4th order Butterworth notch filter. In addition, electrooculography (EOG) (horizontal with electrodes at the outer canthi and vertical with electrodes above/below the left eye) were also acquired. ECG was measured using disposable ECG electrodes with a snap connection to the wiring. The respiration rate (RR) was measured using a SleepSense chest strap which was connected to the upper body. The skin conductance was measured using reusable gold-plated cup

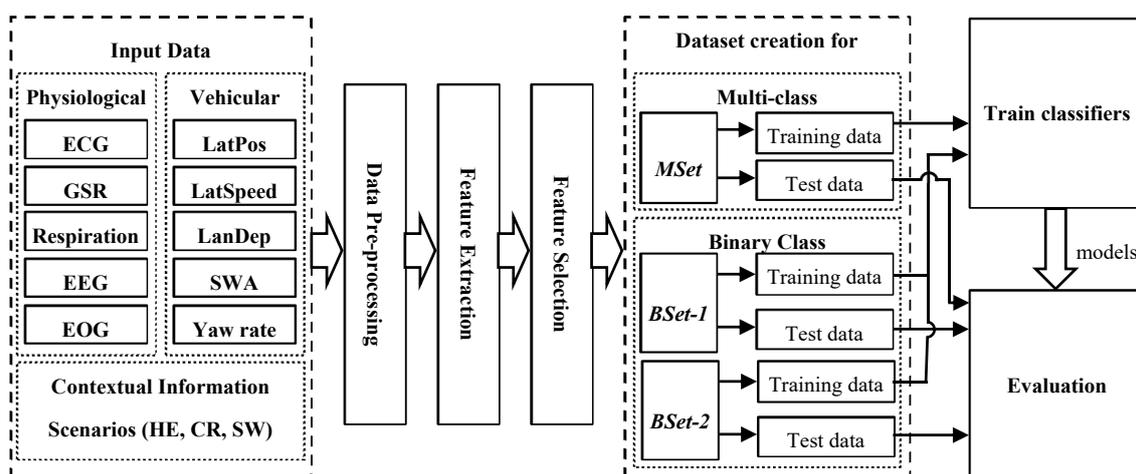
electrodes with conductive cream; the electrodes were connected to a GSR sensor (g.tec g.GSRsensor). Vehicular parameters such as lateral position (LatPos), lateral speed (LatSpeed), steering wheel angle (SWA), lane departure (LanDep), and yaw rate were recorded in the simulator control computer.



**Figure 2.** Three of the study scenarios, (a): Car approaching the crossing from the right (CR); (b): Hidden exit on the right side of the road with a warning sign (HE); and (c): Strong side wind in open terrain (SW) scenario.

## 2.2. Classification Approach

The aim of the classification task was to differentiate the driving events with the cognitive load task from normal driving. The influence of the scenarios on classification is evaluated by classifying cognitive load tasks for all individual situations. Each scenario had a duration of 60 s where the first 10 s of the data were discarded to adjust the stability of the driver with the cognitive load task. Hence, a 50 s recording of each scenario was used for feature extraction. Figure 3 shows the overall schematic diagram of the classification task. The steps include data gathering, data pre-processing, feature extraction, feature selection, dataset creation, training classifiers, and finally evaluation of each of the classifiers using the test dataset. Here, the data were gathered through a study with 66 participants (33 in test series 1 and 33 in test series 2) presented in Section 2.1.



**Figure 3.** Block diagram of the steps for driver cognitive load classification using multivariate data.

### 2.2.1. Data Pre-Processing

The driving task involves activities such as looking at the side and rear-view mirror, shifting gear, and changing body position that naturally causes muscle and ocular artifacts in the EEG signals. Therefore, it requires cleaning the EEG signal before extracting frequency component features from the EEG signal. Hence, EEG signals were artifacts handled using an in-house developed tool called ARTE (Automated aRTifacts handling in EEG) [41]. The median filter was used to handle noise in

the vehicular data, respiration, and GSR signals. The median filter is particularly useful for removing spiky noise and can separate peaks from a slowly changing signal disturbed by unknown noise distribution [42]. A QRS detection algorithm proposed by [43,44] was used to extract inter-beat-interval (IBI) data from the ECG signal. The obtained IBI data were filtered using the ARTiiFACT tool [45]. The collected raw dataset was in the European data format, which was converted into MATLAB (Matlab 2017b version. [https://se.mathworks.com/products/new\\_products/release2017b.html](https://se.mathworks.com/products/new_products/release2017b.html)) data format and all the works were done in MATLAB 2017b version.

### 2.2.2. Feature Extraction

Various features are extracted from both the physiological and vehicular parameters as presented in Table 1. Here, the feature vector consists of 323 extracted features with total observations of 721, where 306 of them are baseline or no-task, 237 observations are the 1-back task, and 178 observations are the 2-back task.

**EEG Features:** From each of the 30 channel EEG signal power, the spectral density (PSD) of the  $\delta$  (<4 Hz),  $\theta$  (4–7 Hz),  $\alpha$  (8–12 Hz),  $\beta$  (12–30 Hz), and  $\gamma$  (31–50 Hz) frequency bands are extracted as features. The Welch's method [46] is used with 50% overlapping with the Blackman window function. In addition, four different ratios of the PSDs,  $(\theta + \alpha)/\beta$ ,  $\alpha/\beta$ ,  $(\theta + \alpha)/(\alpha + \beta)$ , and  $\theta/\beta$  [47], are also estimated as features. These four ratios indicate the change of slow wave to fast wave of EEG activities over time. According to [47], an increase in the ratio is a good indicator of EEG activity compared to  $\alpha$  and  $\theta$  alone. Moreover, the authors found that  $\alpha$  and  $\theta$ , combined with the ratios, could better assess the fatigue condition of the drivers. Hence, nine features from each EEG channel resulted in 270 EEG features for each 50 s time segment driving event. The motivation of using EEG is that as the cognitive load increases, changes in alpha and theta powers in EEG have been observed in various studies [48–50]. It is reported that alpha and theta powers increase as the cognitive load increases [48–50]. Another common approach in the Brain-Computer interface is to apply the independent component analysis (ICA) to extract features from the PSDs of ICA components [51,52]. EEG classifications for different mental workload activities have been performed in [53,54]. However, depending on the study design and the type of cognitive load under scrutiny, the results are often ambiguous [48,50].

**EOG Features:** The EOG features derived from the vertical EOG using an automatic blink detection algorithm based on derivatives and thresholding was developed by Jammes and Sharabty [55]. The average spontaneous eye blink rate of a person is 15–20 per min [56]. The eye blink frequency increases as the cognitive load increases [48,57,58], whereas the decrease in blink duration is observed by [59].

**ECG Features:** Heart rate (HR) and heart rate variability (HRV), i.e., measure of the variations in time between each heartbeat, are two measures that can vary with the increasing cognitive load. HRV measures beat-to-beat (R–R interval) variations in terms of consecutive heartbeats articulated in the normal sinus rhythm from electrocardiogram (ECG) recordings [60,61]. HR and HRV features are obtained from the pre-processed interbeat interval (IBI) data. In time domain, statistical methods are applied to extract the time domain features. To obtain frequency domain features, the IBI data are transformed via FFT transformation. The PSDs of low frequency (LF) (0.04 to 0.15 Hz) and high frequency (HF) (0.15 to 0.40 Hz), LF/HR ratio, and total power are estimated. The time and frequency domain measures quantify the variability of the heart rate fluctuation characteristic in time scales. On the other hand, the non-linear measures quantify the structure or complexity of the R-R intervals, i.e., IBI data. Non-linear measures such as detrended fluctuation analysis, sample entropy, approximate entropy, and permutation entropy methods were applied to extract complexity from the IBI data [62]. An increased HR with respect to the increasing cognitive load has been reported in several studies; in contrast, the time domain measures of HRV such as mean RR, SDNN, RMSDD, pNN50, and HF power band (0.15–0.50 Hz) of HRV in the frequency domain decrease [14,63,64]. An increase in the LF power (0.04–0.15 Hz) and the LF/FH ratio of HRV have been associated with higher mental workloads [64–66].

**Table 1.** List of features extracted from each signal. # represent features.

Signal	# Features	Extracted Features
EEG	270	Frequency bands: $\delta$ (<4 Hz), $\theta$ (4–7 Hz), $\alpha$ (8–12 Hz), $\beta$ (12–30 Hz), $\gamma$ (31–50 Hz), and the ratio $(\theta + \alpha)/\beta$ , $\alpha/\beta$ , $(\theta + \alpha)/(\alpha + \beta)$ , and $\theta/\beta$
EOG	9	Start position of blink, blink duration calculated from the start position of blink to the end value of blink, lid closure speed, PCV (peak closing velocity), delay of eye lid reopening, duration at 80%, PERCLOS, blink rate, blink count.
ECCG	14	Time: Mean heart rate (meanHR), standard deviation of heart rate (sdHR), standard deviations of normal to normal RR intervals (SDNN), root mean square of successive differences between adjacent NN intervals (RMSSD), number of pairs of successive NN intervals with more than 50 ms (NN50), percentage of NN50 (pNN50). Frequency: Low frequency power (0.04–0.15 Hz), high frequency power (0.15–0.4 Hz), total power, LF/HF ratio. Non-linear: Alpha value of detrended fluctuation analysis (dfaAlpha), sample entropy (SampEn), approximate entropy (ApEn), and permutation entropy (PeEn).
GSR	10	Time: Number of peaks, the amplitude of the peaks (maxima-minima), duration of the rise time of each peak, index of the detected peaks in the GSR signal, mean value, standard deviation, first quartile value, third quartile value, slope value between peak and valley. Frequency: Average power of the signal under 1 Hz.
Respiration rate (RR)	9	Time: Mean value, standard deviation, kurtosis. Frequency: Power spectra power between the frequency ranges [0, 0.1], [0.1, 0.2], [0.2, 0.3], [0.3, 0.4], [0.4, 0.7], and [0.7, 1].
Vehicular parameters	11	Standard deviation of lateral position, mean squared error of lateral position. Standard deviation of steering wheel angle, steering wheel entropy, steering wheel reversal rate, high frequency component (0.3 Hz), number of zero crossings. Lanex or fraction of lane exit from lane departure. Standard deviation of lateral speed, yaw and yaw rate.

**GSR Features:** GSR measures the electrical conductivity of the skin and can provide changes in the human sympathetic nervous system [67]. GSR is significantly correlated with the cognitive load task demand and usually used for the level of cognitive load classification [40,67,68]. In time domain several estimations, i.e., number of peaks, the amplitude of the peaks (maxima-minima), duration of the rise time of each peak, index of the detected peaks in the GSR signal, mean value, standard deviation, first quartile value, third quartile value, slope value between peak and valley are extracted as features [67]. One feature which is the average power of the signal under 1 Hz is extracted in frequency domain. A comprehensive review of GSR signal interpretation can be found in [69]. Further, relations between cognitive load and GSR features have been discussed in several studies [68,70,71].

**Respiration Features:** From the respiration rate (RR) signal, arithmetic mean, standard deviation, and kurtosis are calculated as features in time domain. Similar to EEG, Welch's method has been used to estimate the PSDs from the frequency ranges [0, 0.1], [0.1, 0.2], [0.2, 0.3], [0.3, 0.4], [0.4, 0.7], and [0.7, 1] [72,73]. The cognitive load has a distinct effect on the respiratory behaviour that can differ in sensitivity in the parameters obtained from respiratory signals [74]. According to Hidalgo-Muñoz et al. [72] significant increases in the respiration rate are observed while driving in comparison to the base line condition. Moreover, the RR showed variations with a different level task difficulty and RR accelerated with an increasing cognitive workload.

**Vehicular Features:** The standard deviation from five time series data namely, lateral speed, steering wheel angle (SWA), yaw and yaw rate [75], and lateral position [15], are extracted as features. The steering wheel reversal rate (SWRR) [15], is defined as the absolute difference between maximum and minimum of the SWA signal. The SWRR is the number of reversals in a time period. Firstly, the raw SWA is smoothed using the Lowess method where the linear model is used for local fitting [76]. In this case, 110 points have been used for the moving average in the linear model. Steering wheel entropy [77,78], high frequency component (0.3 Hz), and number of zero crossings are the other features that are obtained from the SWA signal. Lanex or the fraction of lane exit feature is extracted from the lane departure signal which indicates the driver's tendency to exit the driving lane. Lanex is defined as the fraction of a given time interval spent outside driving [75]. In several studies, drivers' behavioural data in relation to vehicular signals such as speed, lateral position, steering wheel angle, etc. have been used to detect and classify drivers' cognitive load [15,16]. For example, driving performance relies on a right speed [79]. A reduced speed as a compensatory action due to the increased cognitive load is more often used as an indication of behaviour adaption rather than a change in driving performance [80,81]. Östlund and Nilsson [82] presented a few other parameters such as lateral position and steering wheel reversal rate that contribute to the driver's cognitive load. Wilschut [82] used the steering wheel angle and lane positioning to measure the driving performance. A lane change task can be used to investigate the effects of cognitive load on driving performance [83].

### 2.2.3. Feature Selection

Feature selection is conducted only on the EEG signals since 270 EEG features are extracted from the 30 channels and many of them were neighbouring electrodes. Some overlapping and redundant features might exist. Hence, sequential forward floating selection (SFFS) [84–86] was used to also investigate the intra-feature relationships. SFFS is a successor of the sequential forward selection (SFS) method, which does not suffer from the 'nesting effect', and is computationally more efficient than other branch and bound methods [86]. SFFS was wrapped with an SVM classifier to obtain an optimal feature subset. Further, the SVM classification was evaluated using 5-fold cross-validation. For other features, random forest with the mean decrease accuracy (MDA) [87] approach was used in the feature selection process. The idea of using MDA is to find the direct impact of each feature on the performance of the random forest model. Here, a permutation of each feature measures the decreasing accuracy of the model and for the unimportant features the permutation has little effect on the model accuracy. On the other hand, removing important features should drastically decrease the accuracy.

### 2.2.4. Cognitive Load Classification

For cognitive load classification, data from both the test series are combined, and both multi-class (*MSet* data) and binary class (*BSet* data) classification are defined based on the n-back task and normal driving events. The binary class is defined as the task group and baseline group. For the binary classification, two data sets are created such that the first set (*BSet-1*) baseline consists of normal driving and 1-back task, and the task group contains data of 2-back task. In the second set (*BSet-2*), the baseline includes data from normal driving only, and the task group consists of data from both 1-back and 2-back tasks. These two binary datasets preparation was motivated by the assumption that the 1-back task did not have much influence on the driver (e.g., on working memory) compared to the 2-back task [88]. The *MSet*, *BSet-1*, and *BSet-2* datasets are split into training and test datasets, where the training set contains 70% and the test dataset contains 30% of the data sets. Three separate classifiers k-NN, SVM, and RF are developed and trained using 5-fold cross-validation with the training datasets and later evaluated with the test datasets. In addition, the binary classification was performed for both scenario-wise and task-wise to discriminate the effect of scenarios on the cognitive load task. Here, only the training dataset was used in the feature selection step. The training set is further divided into two sets, where 80% of the training data is used for SFFS and MDA, and 20% of the training data is used as a validation set.

k-NN is a simple memory-based algorithm that uses the observations in the training set to find the most similar properties of the test dataset [89]. In this work, the Euclidean distance function is used with a 'squared inverse' distance weight and  $K = 5$  was considered. SVM finds the hyperplane that not only minimizes the empirical classification error but also maximizes the geometric margin in the classification [90]. SVM can map the original data points from the input space to a high dimensional feature space such that the classification problem becomes simple in this feature space. In this study, an SVM with a Gaussian kernel was used for the classification task. A popular ensemble algorithm in machine learning is RF, that consists of a series of randomizing decision-trees, where the output is the majority vote of all these decision-trees [91]. One important aspect of RF is that it does not assume independence of features. In the driving context, data is often noisy and rarely linearly separable into a different mental state [92]. RF is implemented using bagging, which is the process of bootstrapping the data plus using the aggregate to make a decision. During classification, MATLAB's *fitcknn* function is used for k-NN, the *fitcecoc* function with an SVM template is used for SVM, and the *fitcensemble* function with 4357 tree splits is used for RF. The three classifiers were evaluated considering confusion matrices, accuracy, balanced accuracy (BACC), Matthews correlation coefficient (MCC),  $F_1$ -score, sensitivity, and specificity.

## 3. Results

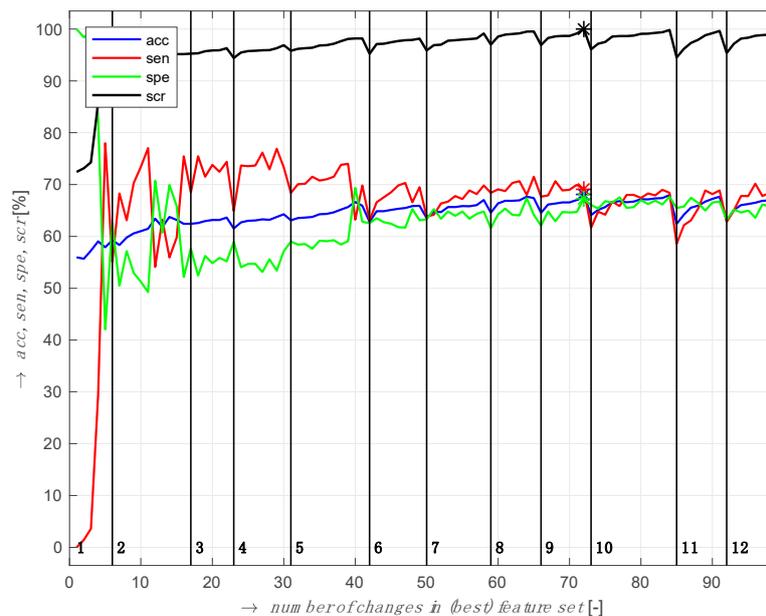
### 3.1. Feature Selection

The SFFS selected  $\theta/\beta$ ,  $\alpha/\beta$ ,  $(\theta + \alpha)/\beta$ ,  $\theta$ ,  $\beta$ , and  $\alpha$  features from seven frontal channels namely, FP1, FP2, F7, F4, FPz, FC2, and FC5. The best classification accuracy was 66% with 11 EEG features. Figure 4 shows the accuracy (acc), sensitivity (*sen*), specificity (*spe*), and classification score (*scr*) defined as  $2^{\sin(\frac{\pi \cdot \text{sen}}{2})} \cdot \sin(\frac{\pi \cdot \text{spe}}{2})$  [85].

The feature subset after feature selection using SFFS and MDA is listed in Table 2. In total, from all the signals, 42 features were selected out of 323 features.

**Table 2.** List of selected features from each signal. # represent features.

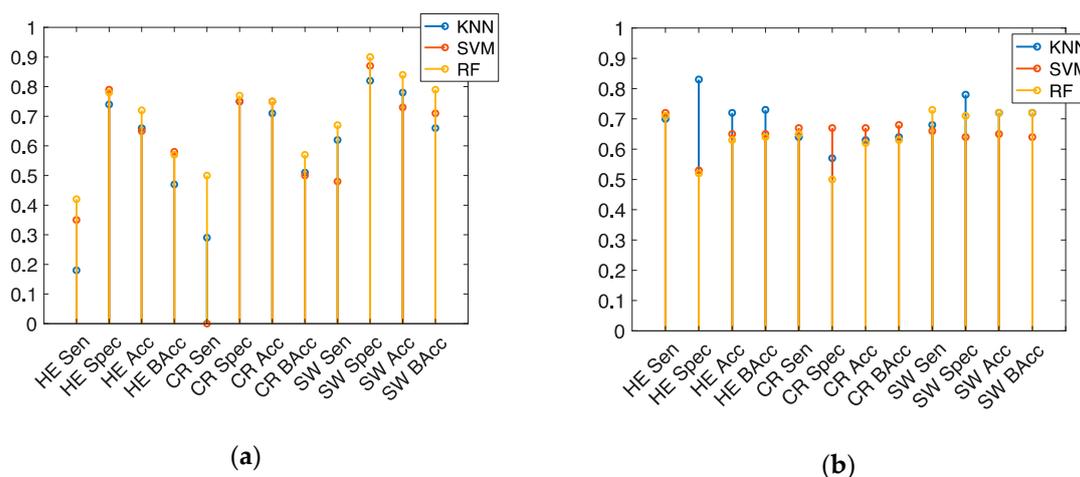
Data	# Extracted Features	# Selected Features	Features
EEG	270	11	FP1: $\beta$ , FP2: $\theta$ , $(\theta + \alpha)/(\alpha + \beta)$ FP2: $\theta$ , $(\theta + \alpha)/(\alpha + \beta)$ FPz: $\beta$ , $\theta/\beta$ , $(\theta + \alpha)/\beta$ F4: $\theta$ F7: $\theta$ FC2: $\theta/\beta$ , $\alpha/\beta$
EOG	9	5	Start position of blink, blink duration calculated from the start position of blink to the end value of blink, PERCLOS, blink rate, blink count.
ECG	14	9	<b>Time:</b> sdHR, SDNN, NN50, pNN50. <b>Frequency:</b> LF, HF, LF/HF ratio. <b>Non-linear:</b> dfaAlpha, SampEn.
GSR	10	4	<b>Time:</b> The amplitude of the peaks, duration of the rise time of each peak, mean value. <b>Frequency:</b> Average power of the signal under 1 Hz.
RR	9	7	<b>Time:</b> Mean value, standard deviation, kurtosis. <b>Frequency:</b> Power spectra power between the frequency ranges [0, 0.1], [0.2, 0.3], [0.4, 0.7], and [0.7, 1].
Vehicular data	11	6	Standard deviation of lateral speed. Standard deviation of lateral speed yaw. Steering wheel entropy, high frequency component (0.3 Hz), and number of zero crossings. Lanex or fraction of lane exit from lane departure.
All	323	42	Best subset of features after feature selection.



**Figure 4.** EEG feature selection using sequential forward floating selection (SFFS) on the training dataset, validated using 10-fold cross-validation.

### 3.2. Classification Evaluation

The scenario wise binary classification was performed to see if there are any effects of the scenarios in the classification performance. Figure 5 shows the performance of binary classification for each scenario using both test datasets of *BSet-1* and *BSet-2*. It can be seen that the balanced accuracies (BAcc) of HE and CR scenarios are lower and higher only for the SW scenario using RF on the data of *BSet-1*. It is important to mention that the ratio of the baseline and task groups in the *BSet-1* is much imbalanced than the *BSet-2*. For both *BSet-1* and *BSet-2*, the BAcc is higher for the side wind scenario. Using the test dataset of *BSet-1*, in the HE scenario, BAcc(s) are 47%, 58%, and 57% for k-NN, SVM, and RF, respectively; in the CR scenario, BAcc(s) are 51% for k-NN, 50% for SVM, and 57% for RF; in the SW scenario, BAcc(s) are 66% for k-NN, 71% for SVM, and 79% for RF.



**Figure 5.** Scenario wise classification performance on the test dataset. (a) Shows results for *BSet-1*; (b) Shows results for *BSet-2*.

On the other hand, using the test dataset of *BSet-2*, in the HE scenario, BAcc(s) are 73%, 65%, and 64% for k-NN, SVM, and RF, respectively; in the CR scenario, BAcc(s) are 64% for k-NN, 68% for SVM, and 63% for RF; in the SW scenario, BAcc(s) are 72% for k-NN, 64% for SVM, and 72% for RF.

As observed in Figure 5, the classification may have some influence on the driving scenario, hence a categorical feature is incorporated with the existing features presented in Table 2. Afterwards the multiclass classification was performed using the *MSet* dataset and the binary classification was performed using both *BSet-1* and *BSet-2* datasets.

Multiclass classifications with k-NN, SVM, and RF are performed to investigate how each class contributed to the classification performance. On the training dataset of *MSet*, using 5-fold cross-validation, k-NN achieved 53% classification accuracies, whereas both SVM and RF achieved 59% classification accuracies. Table 3 shows the confusion matrices for the test dataset. RF shows better performance than k-NN and SVM considering the number of correct classifications of each target group.

**Table 3.** Confusion matrix of k-nearest neighbour (k-NN), support vector machine (SVM), and random forest (RF) multiclass classification on the test dataset. The grey cells represent the true positive (TP) value. TP represents the number of observations that were correctly classified, and the precision value in percentage.

Predicted Class	Actual Class								
	k-NN			SVM			RF		
	Baseline	1-back	2-back	Baseline	1-back	2-back	Baseline	1-back	2-back
Baseline	60 (65%)	18 (20%)	14 (15%)	66 (72%)	13 (14%)	13 (14%)	70 (76%)	15 (16%)	7 (8%)
1-back	24 (34%)	39 (56%)	7 (10%)	22 (31%)	39 (56%)	9 (13%)	24 (34%)	36 (51%)	10 (14%)
2-back	13 (25%)	19 (35%)	21 (40%)	16 (30%)	12 (23%)	25 (47%)	13 (25%)	8 (15%)	32 (60%)

Table 4 represents the classification summary on the test dataset of *MSet* considering true positive (TP), true negative (TN), false positive (FP), false negative (FN), precision, sensitivity, specificity, and balanced accuracy (BACC). Here, one-vs.-rest was used to determine the target groups in the positive (P) and negative (N) classes. The positive class is the target group that corresponds to either baseline, 1-back, or 2-back task in each column. The negative (N) class consists of the other two target groups, i.e., 1-back + 2-back, baseline + 2-back, and baseline + 1-back. Overall, RF shows better performance considering the balanced accuracy.

**Table 4.** Classification summary of multiclass classification for k-NN, SVM, and RF on the test dataset. Where the target classes are baseline (BL) or no-task, 1-back, and 2-back task. SEN: Sensitivity; SPE: Specificity; PRE: Precision; ACC: Accuracy; and BACC: Balanced accuracy.

Criteria	k-NN			SVM			RF		
	BL	1-back	2-back	BL	1-back	2-back	BL	1-back	2-back
TP	60	39	21	66	39	25	70	36	32
FP	32	31	32	26	31	28	22	34	21
FN	37	37	21	38	25	22	37	23	17
TN	86	108	141	85	120	140	86	122	145
PRE	0.65	0.56	0.40	0.72	0.56	0.47	0.76	0.51	0.60
SEN	0.62	0.51	0.50	0.63	0.61	0.53	0.65	0.61	0.65
SPE	0.73	0.78	0.82	0.77	0.79	0.83	0.80	0.78	0.87
BACC	0.68	0.65	0.63	0.70	0.69	0.67	0.73	0.68	0.75
$F_1$ -score	0.63	0.53	0.13	0.67	0.58	0.50	0.70	0.56	0.63
MCC	0.35	0.30	0.44	0.40	0.39	0.35	0.46	0.37	0.51

Binary classifications were performed using *BSet-1* and *BSet-2*. The observed classification accuracies for k-NN, SVM, and RF with 5-fold cross-validation on the training dataset of *BSet-1* are 79%, 81%, and 82%, respectively. On the training dataset of *BSet-2*, the achieved classification accuracies are 67% for k-NN, 72% for SVM, and 75% for RF. The prediction performance of k-NN, SVM, and RF, on the test dataset of each of *BSet-1* and *BSet-2* is presented in Table 5.

**Table 5.** Performance summary of the classifiers for binary classification on the test dataset.

Criteria	BSet-1			BSet-2		
	k-NN	SVM	RF	k-NN	SVM	RF
Task group (P)	52	52	52	126	126	126
Baseline group (N)	163	163	163	89	89	89
TP	20	27	38	116	104	107
FP	8	10	11	47	43	36
FN	32	25	14	10	22	19
TN	155	153	152	42	46	53
Sensitivity	0.38	0.52	0.78	0.92	0.83	0.84
Specificity	0.95	0.86	0.91	0.47	0.52	0.60
Accuracy	0.81	0.84	0.88	0.73	0.70	0.74
BACC	0.67	0.73	0.85	0.70	0.67	0.72
$F_1$ -score	0.50	0.61	0.75	0.80	0.76	0.80
MCC	0.43	0.52	0.68	0.45	0.36	0.46

#### 4. Discussion

Cognitive loading activities on traffic safety and its relation to driving performance has drawn an increasing attention to the traffic safety research issue. Here, the cognitive load dataset was acquired and analysed to understand the effect of cognitive load on traffic safety. Driving a vehicle is an anticipatory task where a driver needs adaptation concerning the road users' behaviours and their actions which are dynamic in nature. Driving is often considered as a process that is nearly automated, partially self-paced, and a satisficing task [93]. A driver can somewhat distribute the load of the driving task by deciding when, where, and what they do. This holds true not only for driving-related tasks but also for secondary tasks such as talking on a mobile phone or conversing with a passenger while driving. Most of the time this works well, but sometimes it does not [6,7,94–96]. In the cognitive load theory, working memory is considered as an executive function that holds information and mentally processes that information [97]. Hence, in this paper the cognitive load is considered as the amount of cognitive resources (i.e., mechanisms necessary for cognitive control) used at a certain time [11]. The effect of cognitive load on traffic safety is considered utilizing the attention selection model (ASM) [11]. According to the ASM model, the cognitive load does not affect the automatic performance but impairs subtasks that rely on cognitive control.

Among the physiological signals, the EEG is one accessible technique to measure cognitive load and the EEG signal analysis can detect changes in an instantaneous load and the effects of cognitively loading secondary tasks. The EEG feature selection in cognitive load classification showed the best feature subset selected by the SFFS algorithm, containing  $\theta/\beta$ ,  $\alpha/\beta$ ,  $(\theta + \alpha)/\beta$ ,  $\theta$ ,  $\beta$ , and  $\alpha$  features from only the frontal electrode. Features from the frontal region might suggest only motor function, and attention affected the cognitive loading activities. HRV from ECG, GSR, and RR features might be better indicators for cognitive load classification, a finding also supported by other studies. HRV features can be an important indicator for classifying cognitive load because cognitive load modulates the sympathetic and parasympathetic nervous systems inversely to driver sleepiness [98]. The time domain GSR, i.e., the peak amplitude, the duration of the rise time of each peak, and the mean GSR value were found to be useful indicators for cognitive load detection when a person is under the influence of different stress levels [99]. Furthermore, the states depend on the experimental design, driving environment, confounding factors, etc., and hence, multi-variate data and data fusion considering the driving context are needed to accurately assess the cognitive load. It should be noted that subjective measures, for example, the NASA-TLX [100] or the DALI (driving activity load index) [101], require understanding the importance of physiological features and vehicular features.

In this paper, the cognitive load classification was performed based on the baseline (just driving) and n-back task (1-back and 2-back). This approach could have affected the classification performance

because the influence of a cognitive loading task (e.g., on working memory) might not be the same for everyone, especially for the 1-back task. It is noteworthy to mention that the cognitive load classification distinguishes among different levels of cognitive-level tasks and does not imply how cognitively loaded participants are performed during the n-back task. In terms of classification, the problem lies in the class noise in the dataset. Apart from the analysis presented in this paper, several other classification experiments [102] have been conducted considering features according to (1) cerebral activities recorded via EEG, (2) cerebral activities recorded via EEG and eye blink waveform via EOG, (3) non-cerebral physiological signals recorded via HRV, GSR, and respiration, and (4) driving behavioural data based on vehicular parameters obtained from the control computer. The results showed poor performance than combining all features as the results presented in this paper. By using only the EEG features from *BSet-1* (i.e., baseline = normal + 1-back task, and the task group = 2-back task) dataset, the highest accuracy of 74%, 46% sensitivity, and 78% specificity was obtained by the RF algorithm. The performance was decreased using only the EEG features from the *BSet-2* (i.e., baseline = normal, and the task group = 1-back task + 2-back task) dataset. Again, RF showed the best performance with 57% accuracy, 61% sensitivity, and 49% specificity. When features from both the EEG and EOG signals were combined, a slight improvement was observed in the classification performance using both the *BSet-1* and *BSet-2* datasets. Here, k-NN showed the best performance for both *BSet-1* and *BSet-2* datasets and the accuracy, sensitivity, and specificity were around 75%, 59%, and 81%, respectively. It has been observed that using only the vehicular features classification perform similarly as using the EEG features only. However, a combination of features from non-cerebral physiological signals, i.e., HRV, GSR, and respiration, was found to perform better for the classification compared to using only EEG, vehicular, and a combination of EEG and EOG features. The RF algorithm obtained the best performance using the *BSet-1* dataset considering 78% accuracy, 70% sensitivity, and 82% specificity. Similarly, RF showed the best performance using the *BSet-2* dataset considering 73% accuracy, 75% sensitivity, and 69% specificity. In all the cases, i.e., using only the EEG feature, a combination of features from EEG and EOG, features from vehicular signal, and combination of features from ECG, GSR, and RR signals the obtained classification accuracy was not more than 50%, and the sensitivity and specificity were around 55% and 60%, respectively.

Overall, a 10% improvement in the classification performance was observed by using a combination of all multivariate features compared to the performance observed when using only the features from the EEG signals. A 20% improvement in the classification performance for multiclass classification was observed by using a combination of all multivariate features compared to that observed using only the feature based on the vehicular data. The current classification approach implies that it is not individualised; that is, the response pattern is assumed to be the same for all drivers. The scenario-wise classification shows that there is an effect of driving condition on the cognitive load. Thus, integrating contextual information as features can be beneficial to the classification. However, in this work, it was not fully comprehending the consequence of adding contextual features. The limitation of this approach can be overcome by incorporating subjective measures into the study design and adding a wide range of contextual information.

## 5. Conclusions

The objective of this paper was to provide analytics on multivariate data for driver cognitive load classification. The multiclass classification results portray the difficulty to correctly classify when there are imbalance classes in the dataset, which leads to performing the binary classification. These analytics emphasize the study design with a wide range of contextual information and subjective measure to predict or identify the level of cognitive load during driving. It is also found that multicomponent features could improve the overall classification performance. Another important issue of this study was the imbalanced class in the dataset. Hence, in this study, BACC, MCC, and  $F_1$ -score were considered along with accuracy, sensitivity, and specificity. It should be noted that though for some occasions  $F_1$ -score, sensitivity, and specificity showed reasonable measures but looking at MCC it is

evident that the models tend to bias towards the class with higher observations. Though the inclusion of contextual feature is inconclusive, yet it is believed that contextual information not only can improve the classification performance but also can provide insights when it requires interpretation of the ML model. It is argued that the n-back task is an efficient task to measure the individual working memory capacity [103]. The scenario wise classification with a *BSet-2* could better discriminate between normal driving and n-back task compared to the binary classification with *BSet-1*. It can be concluded from the result of the scenario wise classification that the cognitive load impairs the driving subtask that depends on cognitive control which is also the suggestion by ASM. Although studies [103,104] found more discriminatory EEG activity patterns between the n-back tasks, those studies only considered the n-back task as the main discriminatory factor. However, in this study the n-back task is adapted for ASM that may influence the classification performance and supports the idea of ASM that the automatic performances of the driving task are unaffected by the cognitive load.

**Author Contributions:** S.B. (Shaibal Barua) conceived the proposed approach and performed the data analysis. He is the main author of the manuscript with support from M.U.A. and S.B. (Shahina Begum) M.U.A. provided advice to formulate the data analysis and experiment of the proposed approach. He also provided support for writing the manuscript. S.B. (Shahina Begum) is the project leader of the Vehicle Driver Monitoring (VDM) project. She received the grant for supporting the VDM research project from the VINNOVA (Swedish Governmental Agency for Innovation Systems). She supervised the idea and prepared the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by VINNOVA (Swedish Governmental Agency for Innovation Systems).

**Acknowledgments:** The authors would like to acknowledge VINNOVA (Swedish Governmental Agency for Innovation Systems) for supporting the Vehicle Driver Monitoring (VDM) research project. The authors would also like to acknowledge our project partners Emma Nilsson, Per Lindén, and Bo Svanberg (Volvo Car Corporation) as well as Anna Anund, Christer Ahlström, and Carina Fors (VTI).

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

- Gabaude, C.; Baracat, B.; Jallais, C.; Bonniaud, M.; Fort, A. Cognitive load measurement while driving. In *Human Factors: A View from an Integrative Perspective, on the Occasion of the Human Factors and Ergonomics Society Europe Chapter Annual Meeting in Toulouse*; De Waard, D., Ed.; HFES: Toulouse, France, 2012; pp. 67–80.
- Solovey, E.T.; Zec, M.; Perez, E.A.G.; Reimer, B.; Mehler, B. Classifying driver workload using physiological and driving performance data: Two field studies. In *Proceedings of the 32nd Annual ACM Conference on Hum. Factors in Computing Systems*, Toronto, ON, Canada, 26 April–1 May 2014; pp. 4057–4066.
- Bennakhi, A.; Safar, M. Ambient Technology in Vehicles: The Benefits and Risks. *Procedia Comput. Sci.* **2016**, *83*, 1056–1063. [[CrossRef](#)]
- Lee, J.D. Driving Safety. *Rev. Hum. Factors Ergon.* **2005**, *1*, 172–218. [[CrossRef](#)]
- Singh, S. Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey. In *Traffic Safety Facts Crash*; Report No. DOT HS 812 115; National Highway Traffic Safety Administration: Washington, DC, USA, 2015.
- Lee, J.D.; Boyle, L.N. Is Talking to Your Car Dangerous? It Depends. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2015**, *57*, 1297–1299. [[CrossRef](#)] [[PubMed](#)]
- Caird, J.K.; Willness, C.R.; Steel, P.; Scialfa, C. A meta-analysis of the effects of cell phones on driver performance. *Accid. Anal. Prev.* **2008**, *40*, 1282–1293. [[CrossRef](#)] [[PubMed](#)]
- Cooper, J.M.; Medeiros-Ward, N.; Strayer, D.L. The impact of eye movements and cognitive workload on lateral position variability in driving. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2013**, *55*, 1001–1014. [[CrossRef](#)]
- Healey, J.; Picard, R. Detecting Stress During Real-World Driving Tasks Using Physiological Sensors. *IEEE Trans. Intell. Transp. Syst.* **2005**, *6*, 156–166. [[CrossRef](#)]
- Victor, T.; Dozza, M.; Bärgrman, J.; Boda, C.-N.; Engström, J.; Flannagan, C.; Lee, J.D.; Markkula, G. *Analysis of Naturalistic Driving Study Data: Safer Glances, Driver Inattention, and Crash Risk*; Victor, T., Ed.; The National Academy Press: Washington, DC, USA, 2014.

11. Engström, J.; Markkula, G.; Victor, T.; Merat, N. Effects of Cognitive Load on Driving Performance: The Cognitive Control Hypothesis. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2017**, *59*, 734–764. [[CrossRef](#)]
12. Jaeggi, S.M.; Buschkuhl, M.; Perrig, W.J.; Meier, B. The concurrent validity of the N-back task as a working memory measure. *Memory* **2010**, *18*, 394–412. [[CrossRef](#)]
13. Kane, M.J.; Conway, A.R.A.; Miura, T.K.; Colflesh, G.J.H. Working memory, attention control, and the n-back task: A question of construct validity. *J. Exp. Psychol. Learn. Mem. Cogn.* **2007**, *33*, 615–622. [[CrossRef](#)]
14. Mehler, B.; Reimer, B.; Wang, Y. A comparison of heart rate and heart rate variability indices in distinguishing single-task driving and driving under secondary cognitive workload. In Proceedings of the 6th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Olympic Valley-Lake Tahoe, CA, USA, 27–30 June 2011; pp. 590–597.
15. Kountouriotis, G.K.; Spyridakos, P.; Carsten, O.M.; Merat, N. Identifying cognitive distraction using steering wheel reversal rates. *Accid. Anal. Prev.* **2016**, *96*, 39–45. [[CrossRef](#)]
16. Chakraborty, B.; Nakano, K. Automatic detection of driver's awareness with cognitive task from driving behavior. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary, 9–12 October 2016; pp. 003630–003633.
17. Yeo, M.V.; Li, X.; Shen, K.; Wilder-Smith, E. Can SVM be used for automatic EEG detection of drowsiness during car driving? *Saf. Sci.* **2009**, *47*, 115–124. [[CrossRef](#)]
18. Chen, L.-L.; Zhao, Y.; Ye, P.-F.; Zhang, J.; Zou, J.-Z. Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers. *Expert Syst. Appl.* **2017**, *85*, 279–291. [[CrossRef](#)]
19. Hu, S.; Zheng, G. Driver drowsiness detection with eyelid related parameters by Support Vector Machine. *Expert Syst. Appl.* **2009**, *36*, 7651–7658. [[CrossRef](#)]
20. Chui, K.T.; Tsang, K.-F.; Chi, H.R.; Wu, C.K.; Ling, B.W.-K. Electrocardiogram based classifier for driver drowsiness detection. In Proceedings of the IEEE 13th International Conference on Industrial Informatics (INDIN), Cambridge, UK, 22–24 July 2015.
21. Liang, Y.; Reyes, M.L.; Lee, J.D. Real-Time Detection of Driver Cognitive Distraction Using Support Vector Machines. *IEEE Trans. Intell. Transp. Syst.* **2007**, *8*, 340–350. [[CrossRef](#)]
22. Yoshizawa, A.; Nishiyama, H.; Iwasaki, H.; Mizoguchi, F. Machine-learning approach to analysis of driving simulation data. In Proceedings of the IEEE 15th International Conference on Cognitive Informatics & Cognitive Computing (ICCI\*CC), Stanford, CA, USA, 22–23 August 2016.
23. Liao, Y.; Li, S.E.; Li, G.; Wang, W.; Cheng, B.; Chen, F. Detection of driver cognitive distraction: An SVM based real-time algorithm and its comparison study in typical driving scenarios. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Gothenburg, Sweden, 19–22 June 2016.
24. Soman, K.; Sathiya, A.; Suganthi, N. Classification of stress of automobile drivers using Radial Basis Function Kernel Support Vector Machine. In Proceedings of the International Conference on Information Communication and Embedded Systems (ICICES2014), Chennai, India, 27–28 February 2014.
25. Munla, N.; Khalil, M.; Shahin, A.; Mourad, A. Driver stress level detection using HRV analysis. In Proceedings of the International Conference on Advances in Biomedical Engineering (ICABME), Beirut, Lebanon, 16–18 September 2015.
26. Correa, A.G.; Orosco, L.; Laciari, E. Automatic detection of drowsiness in EEG records based on multimodal analysis. *Med. Eng. Phys.* **2014**, *36*, 244–249. [[CrossRef](#)] [[PubMed](#)]
27. Ma, J.; Murphey, Y.L.; Zhao, H. Real Time Drowsiness Detection Based on Lateral Distance Using Wavelet Transform and Neural Network. In Proceedings of the IEEE Symposium Series on Computational Intelligence, Xiamen, China, 6–9 December 2015.
28. Dwivedi, K.; Biswaranjan, K.; Sethi, A.; Dwivedi, K.; Sethi, A. Drowsy driver detection using representation learning. In Proceedings of the IEEE International Advance Computing Conference (IACC), New Delhi, India, 21–22 February 2014.
29. Manawadu, U.E.; Kawano, T.; Murata, S.; Kamezaki, M.; Muramatsu, J.; Sugano, S. Multiclass Classification of Driver Perceived Workload Using Long Short-Term Memory based Recurrent Neural Network. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Suzhou, China, 26–30 June 2018.
30. El Haouij, N.; Poggi, J.-M.; Ghazi, R.; Sevestre-Ghalila, S.; Jaidane, M. Random forest-based approach for physiological functional variable selection for driver's stress level classification. *J. Ital. Stat. Soc.* **2018**, *28*, 157–185. [[CrossRef](#)]

31. Yoshida, Y.; Ohwada, H.; Mizoguchi, F. Extracting tendency and stability from time series and random forest for classifying a car driver's cognitive load. In Proceedings of the IEEE 13th International Conference on Cognitive Informatics and Cognitive Computing, London, UK, 18–20 August 2014; pp. 258–265.
32. Sarkar, P.; Ross, K.; Ruberto, A.J.; Rodenburg, D.; Hungler, P.; Etemad, A. Classification of Cognitive Load and Expertise for Adaptive Simulation using Deep Multitask Learning. In Proceedings of the 8th International Conference on Affective Computing and Intelligent Interaction (ACII), Cambridge, UK, 3–6 September 2019.
33. Saha, A.; Minz, V.; Bonela, S.; Sreeja, S.R.; Chowdhury, R.; Samanta, D. *Classification of EEG Signals for Cognitive Load Estimation Using Deep Learning Architectures*; Springer International Publishing: Cham, Switzerland, 2018.
34. Begum, S.; Ahmed, M.; Funk, P.; Xiong, N.; Von Schéele, B. *Using Calibration and Fuzzification of Cases for Improved Diagnosis and Treatment of Stress*; ECCBR: Vasteras, Sweden, 2006.
35. Begum, S.; Barua, S.; Filla, R.; Ahmed, M.U. Classification of physiological signals for wheel loader operators using Multi-scale Entropy analysis and case-based reasoning. *Expert Syst. Appl.* **2014**, *41*, 295–305. [[CrossRef](#)]
36. Begum, S.; Ahmed, M.U.; Funk, P.; Filla, R. Mental state monitoring system for the professional drivers based on Heart Rate Variability analysis and Case-Based Reasoning. In Proceedings of the Computer Science and Information Systems Federated Conference (FedCSIS), Wroclaw, Poland, 9–12 September 2012; pp. 35–42.
37. Barua, S.; Ahmed, M.U.; Begum, S. Classifying Drivers' Cognitive Load Using EEG Signals. *Stud. Health Technol. Inform.* **2017**, *237*, 99–106.
38. De Naurois, C.J.; Bourdin, C.; Stratulat, A.; Diaz, E.; Vercher, J.-L. Detection and prediction of driver drowsiness using artificial neural network models. *Accid. Anal. Prev.* **2019**, *126*, 95–104. [[CrossRef](#)]
39. Kartsch, V.J.; Benatti, S.; Schiavone, P.D.; Rossi, D.; Benini, L. A sensor fusion approach for drowsiness detection in wearable ultra-low-power systems. *Inf. Fusion* **2018**, *43*, 66–76. [[CrossRef](#)]
40. Chen, F.; Zhou, J.; Wang, Y.; Yu, K.; Arshad, S.Z.; Khawaji, A.; Conway, D. Galvanic Skin Response-Based Measures. In *Creativity and Rationale*; Springer International Publishing: Cham, Switzerland, 2016; pp. 87–99.
41. Barua, S.; Ahmed, M.U.; Ahlstrom, C.; Begum, S.; Funk, P. Automated EEG Artifact Handling With Application in Driver Monitoring. *IEEE J. Biomed. Health Inform.* **2017**, *22*, 1350–1361. [[CrossRef](#)] [[PubMed](#)]
42. Stone, D.C. Application of median filtering to noisy data. *Can. J. Chem.* **1995**, *73*, 1573–1581. [[CrossRef](#)]
43. Nygård, M.-E.; Sörnmo, L. Delineation of the QRS complex using the envelope of the e.c.g. *Med. Biol. Eng.* **1983**, *21*, 538–547. [[CrossRef](#)] [[PubMed](#)]
44. Afonso, V.; Tompkins, W.J.; Nguyen, T.; Luo, S. ECG beat detection using filter banks. *IEEE Trans. Biomed. Eng.* **1999**, *46*, 192–202. [[CrossRef](#)] [[PubMed](#)]
45. Kaufmann, T.; Sütterlin, S.; Schulz, S.M.; Vögele, C. ARTiiFACT: A tool for heart rate artifact processing and heart rate variability analysis. *Behav. Res. Methods* **2011**, *43*, 1161–1170. [[CrossRef](#)]
46. Welch, P. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* **1967**, *15*, 70–73. [[CrossRef](#)]
47. Jap, B.T.; Lal, S.; Fischer, P.; Bekiaris, E. Using EEG spectral components to assess algorithms for detecting fatigue. *Expert Syst. Appl.* **2009**, *36*, 2352–2359. [[CrossRef](#)]
48. Borghini, G.; Astolfi, L.; Vecchiato, G.; Mattia, D.; Babiloni, C. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.* **2014**, *44*, 58–75. [[CrossRef](#)]
49. Gevins, A.; E Smith, M. Neurophysiological measures of cognitive workload during human-computer interaction. *Theor. Issues Ergon. Sci.* **2003**, *4*, 113–131. [[CrossRef](#)]
50. Hagemann, K. The alpha band as an electrophysiological indicator for internalized attention and high mental workload in real traffic driving. In *Mathematics and Natural Sciences*; Heinrich-Heine University of Dusseldorf: Dusseldorf, Germany, 2008.
51. Demanuele, C.; Broyd, S.J.; Sonuga-Barke, E.J.; James, C. Neuronal oscillations in the EEG under varying cognitive load: A comparative study between slow waves and faster oscillations. *Clin. Neurophysiol.* **2013**, *124*, 247–262. [[CrossRef](#)]
52. Dasari, D.; Shou, G.; Ding, L. ICA-Derived EEG Correlates to Mental Fatigue, Effort, and Workload in a Realistically Simulated Air Traffic Control Task. *Front. Mol. Neurosci.* **2017**, *11*, 297. [[CrossRef](#)] [[PubMed](#)]
53. Gupta, A.; Parameswaran, S.; Lee, C.-H. Classification of electroencephalography (EEG) signals for different mental activities using Kullback Leibler (KL) divergence. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Taipei, Taiwan, 19–24 April 2009.

54. Wang, Z.; Hope, R.M.; Wang, Z.; Ji, Q.; Gray, W.D. An EEG workload classifier for multiple subjects. In Proceedings of the IEEE International Conference on Engineering in Medicine, Biology and Society (EMBC), Boston, MA, USA, 30 August–3 September 2011.
55. Jammes, B.; Sharabaty, H.; Estève, D. Automatic EOG analysis: A first step toward automatic drowsiness scoring during wake-sleep transitions. *Somnol. Schlaforschung Schlafmed.* **2008**, *12*, 227–232. [[CrossRef](#)]
56. Nakano, T.; Kato, M.; Morito, Y.; Itoi, S.; Kitazawa, S. Blink-related momentary activation of the default mode network while viewing videos. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 702–706. [[CrossRef](#)] [[PubMed](#)]
57. Recarte, M.A.; Pérez, E.; Conchillo, A.; Nunes, L.M. Mental workload and visual impairment: Differences between pupil, blink, and subjective rating. *Span. J. Psychol.* **2008**, *11*, 374–385. [[CrossRef](#)]
58. Wascher, E.; Heppner, H.; Möckel, T.; Kobald, S.O.; Getzmann, S. Eye-blinks in choice response tasks uncover hidden aspects of information processing. *EXCLI J.* **2015**, *14*, 1207–1218.
59. Benedetto, S.; Pedrotti, M.; Minin, L.; Baccino, T.; Re, A.; Montanari, R. Driver workload and eye blink duration. *Transp. Res. Part F Traffic Psychol. Behav.* **2011**, *14*, 199–208. [[CrossRef](#)]
60. Föhr, T.; Tolvanen, A.; Myllymäki, T.; Järvelä-Reijonen, E.; Rantala, S.; A Korpela, R.; Peuhkuri, K.; Kolehmainen, M.; Puttonen, S.; Lappalainen, R.; et al. Subjective stress, objective heart rate variability-based stress, and recovery on workdays among overweight and psychologically distressed individuals: A cross-sectional study. *J. Occup. Med. Toxicol.* **2015**, *10*, 39. [[CrossRef](#)]
61. Reisman, S. Measurement of physiological stress. In Proceedings of the IEEE 23rd Northeast Bioengineering Conference, Durham, NH, USA, 21–22 May 1997.
62. Schumacher, A. Linear and Nonlinear Approaches to the Analysis of R-R Interval Variability. *Boil. Res. Nurs.* **2004**, *5*, 211–221. [[CrossRef](#)]
63. Brookhuis, K.A.; De Waard, D. Monitoring drivers' mental workload in driving simulators using physiological measures. *Accid. Anal. Prev.* **2010**, *42*, 898–903. [[CrossRef](#)]
64. Cinaz, B.; Arnrich, B.; La Marca, R.; Tröster, G. Monitoring of mental workload levels during an everyday life office-work scenario. *Pers. Ubiquitous Comput.* **2011**, *17*, 229–239. [[CrossRef](#)]
65. Muthukrishnan, S.; Gurja, J.; Sharma, R. Does heart rate variability predict human cognitive performance at higher memory loads? *Ind. J. Physiol. Pharmacol.* **2017**, *61*, 14–22.
66. Togo, F.; Takahashi, M. Heart Rate Variability in Occupational Health—A Systematic Review. *Ind. Health* **2009**, *47*, 589–602. [[CrossRef](#)] [[PubMed](#)]
67. Nourbakhsh, N.; Wang, Y.; Chen, F. GSR and Blink Features for Cognitive Load Classification. In *Human-Computer Interaction—INTERACT 2013*; Kotzé, P., Ed.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 159–166.
68. Nourbakhsh, N.; Wang, Y.; Chen, F.; Calvo, R.A. Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. In Proceedings of the 24th Australian Computer-Human Interaction Conference (ACM), Melbourne, Australia, 26–30 November 2012; pp. 420–423.
69. Boucsein, W. *Electrodermal Activity*; Springer Science & Business Media: New York, NY, USA, 2012.
70. Ikehara, C.; Crosby, M. Assessing cognitive load with physiological sensors. In Proceedings of the 38th IEEE Annual Hawaii International Conference on System Sciences, Big Island, HI, USA, 3–6 January 2005.
71. Shi, Y.; Ruiz, N.; Taib, R.; Choi, E.; Chen, F. Galvanic skin response (GSR) as an index of cognitive load. *CHI '07 Ext. Abstr.* **2007**, 2651–2656. [[CrossRef](#)]
72. Hidalgo-Muñoz, A.R.; Béquet, A.J.; Astier-Juvenon, M.; Pépin, G.; Fort, A.; Jallais, C.; Tattegrain, H.; Gabaude, C. Respiration and Heart Rate Modulation Due to Competing Cognitive Tasks While Driving. *Front. Hum. Neurosci.* **2019**, *12*, 525. [[CrossRef](#)]
73. Veltman, J.; Gaillard, A.W.K. Physiological workload reactions to increasing levels of task difficulty. *Ergonomics* **1998**, *41*, 656–669. [[CrossRef](#)] [[PubMed](#)]
74. Grassmann, M.; Vlemincx, E.; Von Leupoldt, A.; Mittelstaedt, J.; Bergh, O.V.D. Respiratory Changes in Response to Cognitive Load: A Systematic Review. *Neural Plast.* **2016**, *2016*, 8146809. [[CrossRef](#)]
75. Sandberg, D.; Kecklund, G.; Wahde, M.; Åkerstedt, T.; Anund, A. Detecting Driver Sleepiness Using Optimized Nonlinear Combinations of Sleepiness Indicators. *IEEE Trans. Intell. Transp. Syst.* **2010**, *12*, 97–108. [[CrossRef](#)]
76. Cleveland, W.S. Robust Locally Weighted Regression and Smoothing Scatterplots. *J. Am. Stat. Assoc.* **1979**, *74*, 829–836. [[CrossRef](#)]
77. Nakayama, O.; Futami, T.; Nakamura, T.; Boer, E.R. Development of a Steering Entropy Method for Evaluating Driver Workload. *SAE Tech. Pap. Ser.* **1999**, *1*, 1686–1695. [[CrossRef](#)]

78. Kersloot, T.; Flint, A.; Parkes, A. Steering Entropy as a Measure of Impairment. In *Presented during the Young Researchers Seminar*; TRL Limited: Berkshire, UK, 2003.
79. Lewis-Evans, B.; De Waard, D.; Brookhuis, K.A. Speed maintenance under cognitive load—Implications for theories of driver behaviour. *Accid. Anal. Prev.* **2011**, *43*, 1497–1507. [[CrossRef](#)]
80. Östlund, J.; Nilsson, L.; Carsten, O.; Merat, N.; Jamson, S.; Janssen, W.; Mouta, S.; Carvalhais, J.; SANTOS, J.; Anttila, V.; et al. Deliverable 2-HMI and Safety-related Driver Performance. In *Human Machine Interface and the Safety of Traffic in Europe (HASTE) Project*; EC: Geneva, Switzerland, 2004.
81. Engström, J. *Understanding Attention Selection in Driving: From Limited Capacity to Adaptive Behaviour in Vehicle Safety Division, Department of Applied Mechanics*; Chalmers University of Technology: Gothenburg, Sweden, 2011.
82. Wilschut, E.S. *The Impact of In-Vehicle Information Systems on Simulated Driving Performance, Effects of Age, Timing and Display Characteristics*; University of Groningen: Groningen, The Netherlands, 2009.
83. Apparies, R.J.; Riniolo, T.C.; Porges, S.W. A psychophysiological investigation of the effects of driving longer-combination vehicles. *Ergonomics* **1998**, *41*, 581–592. [[CrossRef](#)] [[PubMed](#)]
84. Whitney, A. A Direct Method of Nonparametric Measurement Selection. *IEEE Trans. Comput.* **1971**, *20*, 1100–1103. [[CrossRef](#)]
85. Mekyska, J.; Galaz, Z.; Mzourek, Z.; Smekal, Z.; Rektorova, I.; Eliasova, I.; Kostalova, M.; Mrackova, M.; Berankova, D.; Faundez-Zanuy, M.; et al. Assessing progress of Parkinson’s disease using acoustic analysis of phonation. In *Proceedings of the 4th International Work Conference on Bioinspired Intelligence (IWOB)*, San Sebastian, Spain, 10–12 June 2015; pp. 111–118.
86. Pudil, P.; Novovicova, J.; Kittler, J. Floating search methods in feature selection. *Pattern Recognit. Lett.* **1994**, *15*, 1119–1125. [[CrossRef](#)]
87. Hong, H.; Xiaoling, G.; Hua, Y. Variable selection using Mean Decrease Accuracy and Mean Decrease Gini based on Random Forest. In *Proceedings of the 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, China, 26–28 August 2016.
88. Nilsson, E.; Ahlström, C.; Barua, S.; Fors, C.; Lindén, P.; Svanberg, B.; Begum, S.; Ahmed, M.U.; Anund, A. *Vehicle Driver Monitoring: Sleepiness and Cognitive Load*; VTI Rapport; Statens Väg-Och Transportforskningsinstitut: Linköping, Sweden, 2017; p. 66.
89. Larose, D.T. k-Nearest Neighbor Algorithm. In *Discovering Knowledge in Data*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2005; pp. 90–106.
90. Vapnik, V.N. Principles of Risk Minimization for Learning Theory. *Adv. Neural Inform. Process. Syst.* **1992**, *4*, 831–838.
91. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.
92. McDonald, A.D.; Lee, J.D.; Schwarz, C.; Brown, T.L. Steering in a random forest: Ensemble learning for detecting drowsiness-related lane departures. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2014**, *56*, 986–998. [[CrossRef](#)]
93. Kircher, K.; Ahlstrom, C. Minimum Required Attention: A Human-Centered Approach to Driver Inattention. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2016**, *59*, 471–484. [[CrossRef](#)] [[PubMed](#)]
94. Horrey, W.J.; Lesch, M.F. Driver-initiated distractions: Examining strategic adaptation for in-vehicle task initiation. *Accid. Anal. Prev.* **2009**, *41*, 115–122. [[CrossRef](#)]
95. Engström, J.; Johansson, E.; Östlund, J. Effects of visual and cognitive load in real and simulated motorway driving. *Transp. Res. Part F Traffic Psychol. Behav.* **2005**, *8*, 97–120. [[CrossRef](#)]
96. May, J.F.; Baldwin, C.L. Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. *Transp. Res. Part F Traffic Psychol. Behav.* **2009**, *12*, 218–224. [[CrossRef](#)]
97. Ilkowska, M.; Engle, R.W. Working memory capacity and self-regulation. In *Handbook of Personality and Self-Regulation*; APA: Washington, DC, USA, 2010; pp. 263–290.
98. Tjolleng, A.; Jung, K.; Hong, W.; Lee, W.; Lee, B.; You, H.; Son, J.; Park, S. Classification of a Driver’s cognitive workload levels using artificial neural network on ECG signals. *Appl. Ergon.* **2017**, *59*, 326–332. [[CrossRef](#)] [[PubMed](#)]
99. Conway, D.; Dick, I.; Li, Z.; Wang, Y.; Chen, F. The Effect of Stress on Cognitive Load Measurement. In *Human-Computer Interaction—INTERACT 2013*; Springer: Berlin/Heidelberg, Germany, 2013.

100. Cici, R.; Högman, L.; Patten, C. Measures of Driver Behavior and Cognitive Workload in a Driving Simulator and in a Real Traffic Environment-Experiences from Two Experimental Studies in Sweden. In Proceedings of the First International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Aspen, CO, USA, 14–17 August 2001; pp. 137–142.
101. Pauzié, A. A method to assess the driver mental workload: The driving activity load index (DALI). *IET Intell. Transp. Syst.* **2008**, *2*, 315. [[CrossRef](#)]
102. Barua, S. *Multivariate Data Analytics to Identify Driver's Sleepiness, Cognitive Load and Stress*; Mälardalen University: Västerås, Sweden, 2019.
103. Wilhelm, O.; Hildebrandt, A.H.; Oberauer, K. What is working memory capacity, and how can we measure it? *Front. Psychol.* **2013**, *4*, 4. [[CrossRef](#)] [[PubMed](#)]
104. Yang, C.-Y.; Huang, C.-K. Working-memory evaluation based on EEG signals during n-back tasks. *J. Integr. Neurosci.* **2018**, *17*, 695–707. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).