Preprints.org

Article

Real Time Classification of Cognitive Load Using fNIRS and EEGNet in a Driving Simulation Task

Mehshan Ahmed Khan [*] , Houshyar Asadi , Mohammad Reza Chalak Qazani , Ghazal Bargshady , Sam Oladazimi , Zoran Najdovski , Lei Wei , Hirash Moradi , Saeid Nahavandi

Article

Real Time Classification of Cognitive Load Using fNIRS and EEGNet in a Driving Simulation Task

Mehshan Ahmed Khan [1,*], Houshyar Asadi [1], Mohammad Reza Chalak Qazani [2], Ghazal Bargshady [3], Sam Oladazimi [4], Zoran Najdovski [1], Lei Wei [1], Hirash Moradi [5] and Saeid Nahavandi [6]

[1] Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Geelong, Victoria

[2] Facuaty of Computing and Information, Sohar University, Albatinah North Governorate 311, Oman

[3] Faculty of Science and Technology, University of Canberra, Australia

[4] Hermann-Hesse Str. 17 72221 Haiterbach, Germany

[5] Building Controls GmbH, Schießmauer 18, 92283 Lauterhofen, Deutschland

[6] Swinburne University of Technonology, Hawthorn, Victoria, 3122, Australia

* Correspondence: mehshan.khan@deakin.edu.au

Abstract

Understanding and monitoring cognitive workload is essential for ensuring safety and performance in cognitively demanding environments such as driving. In this study, we investigate the effectiveness of Functional Near-Infrared Spectroscopy (fNIRS) in measuring and classifying cognitive workload during a realistic simulated driving task. Participants were required to perform a dual-task paradigm, which combined a primary driving task with a secondary auditory n-back task at three levels of difficulty: 0-back (low workload), 1-back (moderate workload), and 2-back (high workload). This multi-level workload design allowed us to capture a wide range of cognitive demands reflective of real-world multitasking scenarios. To analyze the hemodynamic responses recorded from the prefrontal cortex, we employed the EEGNet deep learning model, adapted for use with fNIRS signals. A Principal Component Analysis (PCA) feature selection method was applied to reduce dimensionality and isolate the most informative features, improving model performance and computational efficiency. We evaluated the model's performance using both overlapping and non-overlapping signal segmentation strategies across different window lengths (10s, 20s, and 30s). Our results show that a learning rate of 0.001 consistently produced the highest classification accuracy. Notably, a 30-second overlapping window achieved 100% accuracy, while 10-second non-overlapping segments yielded 97% accuracy, highlighting the influence of temporal segmentation on workload classification performance.

Keywords: cognitive workload; functional near-infrared spectroscopy (fNIRS); driving simulation; deep learning; EEGNet

Introduction

Cognitive workload defined as the mental effort required to process information and execute tasks, plays a critical role in performance, safety, and well-being across a range of high-stakes environments including aviation, healthcare, transportation, and human-computer interaction [1]. Understanding and accurately assessing cognitive workload is essential for optimizing task efficiency, preventing mental fatigue, and designing adaptive systems that respond intelligently to human limitations. In mixed-initiative systems, where humans and artificial agents collaborate dynamically, it is especially important that the artificial agents are cognitively aware that is, capable of recognizing the mental demands placed on their human counterparts. In such settings, intelligent systems should avoid imposing unnecessary cognitive strain, such as through excessive alerts, irrelevant information delivery, or poorly timed verbal interactions. Designing such adaptive systems

requires reliable, real-time measures of cognitive load. In this study, we explore the application of Functional Near Infrared Spectroscopy (fNIRS) as a neuroimaging tool to monitor and assess human cognitive workload in a multi-modal driving simulation environment [2,3]. fNIRS offers a non-invasive means of measuring cortical hemodynamic responses, particularly within the prefrontal cortex, which is associated with executive functions such as attention, working memory, and decision-making.

Researchers have made significant efforts to develop mental load classification models using various physiological signals, including Electroencephalogram EEG [4], eye-tracking [5], and, more recently, fNIRS [6], particularly within the domain of cognitive load assessment. Numerous machine learning models, such as Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and Random Forests [2,7–9], as well as deep learning architectures like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks [10,11], have been employed for this purpose. These models offer distinct advantages in detecting and interpreting cognitive states: they leverage advanced pattern recognition capabilities through optimized function fitting methods, enabling the identification of complex, non-obvious relationships in the data that traditional statistical techniques might overlook [12,13]. Furthermore, cross-validation strategies provide a systematic mechanism to verify that cognitive states can be reliably induced and measured across different participants. However, despite these advancements, several limitations remain. The cognitive workload states induced by benchmark tasks may not accurately replicate the natural cognitive demands experienced during real-world activities, potentially limiting the generalizability of findings. Additionally, many studies have relied on proprietary experimental designs, making it difficult to establish standard benchmarks or compare results across different investigations. It also remains uncertain whether methodologies developed for other brain imaging modalities, such as EEG or Functional magnetic resonance imaging   fMRI, are directly applicable or optimal for fNIRS data, given its distinct spatial and temporal characteristics.

To investigate varying levels of cognitive workload, we developed a dual-task experimental paradigm in which participants were required to perform a primary driving task concurrently with a secondary cognitive task a modified version of the n-back task delivered through the auditory modality. This setup was designed to replicate the cognitive demands encountered in real-world multitasking situations, such as managing complex navigation while processing verbal information. The n-back task was implemented at three distinct levels of difficulty: 0-back (representing low cognitive workload), 1-back (moderate workload), and 2-back (high workload). By varying the task difficulty, we aimed to elicit clearly differentiated cognitive states that could be measured through physiological responses during the driving simulation. This approach allows us to evaluate the sensitivity and effectiveness of fNIRS in detecting subtle changes in cognitive load under realistic conditions. While previous studies have primarily focused on two-level workload comparisons (e.g., low vs. high), our three-tiered workload design introduces a more granular framework for understanding how cognitive demand span across multiple task intensities. In addition to the experimental design, we employed the EEGNet deep learning architecture to analyze the recorded fNIRS signals. EEGNet, originally developed for EEG data classification, has been adapted in our study to process and classify hemodynamic responses captured by fNIRS, enabling automated workload detection. Importantly, we explored the model's performance using both overlapping and non-overlapping window segments of the fNIRS signal, an aspect that has been rarely addressed in existing literature. By incorporating these two segmentation strategies, we provide a comparative evaluation of how temporal segmentation influences classification accuracy and model generalizability.

Materials and Methods

Thirty-eight drivers participated in the study. Inclusion criteria required a valid driver's license and the absence of mental health disorders, neurological conditions, or physical impairments that

could affect cognitive function. This ensured consistent cognitive performance across participants, improving the reliability and comparability of the results.

Apparatus

For this study, we employed a high-fidelity driving simulator specifically designed to deliver an immersive and realistic driving experience. Central to the system was the Next Level Racing Motion Platform V3, mounted securely on the Traction Plus Platform. This configuration was selected to maximize the simulation of real-world vehicle dynamics by providing physical feedback for acceleration, braking, road surface texture, and lateral forces. Such detailed motion feedback is critical for enhancing the ecological validity of the simulated driving environment, allowing participants to respond more naturally to vehicle behavior. The visual interface consisted of three 32-inch Samsung monitors arranged in a panoramic setup to create an expansive field of view, closely simulating the peripheral vision critical for real-world driving tasks. This multi-monitor arrangement aimed to enhance situational awareness, depth perception, and hazard anticipation among participants. To further increase realism, the simulator's control system was calibrated to emulate the driving dynamics and interior cabin layout of a Toyota Fortuner SUV. This ensured consistency across participants in terms of vehicle handling expectations, control ergonomics, and spatial orientation within the virtual environment. The driving scenarios were developed and implemented using Euro Truck Simulator 2 (ETS2), a simulation platform widely recognized for its accurate driving physics and flexible environmental settings. ETS2 enabled the replication of challenging driving conditions, including nighttime driving and heavy rainfall, to systematically induce cognitive workload.

To monitor brain activity during the experiments, we utilized the NIRSIT (OBELAB Inc., South Korea), a high-density, wearable fNIRS device. The NIRSIT system incorporates 24 laser diode light sources and 32 photodetectors, operating at two near-infrared wavelengths (780 nm and 850 nm), optimized for detecting cerebral hemodynamic responses. The overall experimental setup, including the driving simulator and fNIRS device configuration, is illustrated in Figure 1. a and   Figure 1. b respectively.
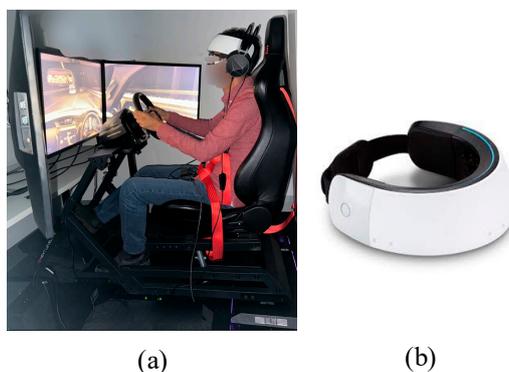

(a)   (b)

Figure 1. (a) Experimental driving simulator setup featuring the Next Level Racing Motion Platform V3 and Traction Plus Platform with a triple-monitor configuration to replicate realistic driving dynamics and immersive environmental conditions. (b) fNIRS neuroimaging device (NIRSIT, OBELAB Inc.) used for recording oxygenated and deoxygenated hemoglobin signals with a sampling frequency of 8.138 Hz.

Secondary Task

To simulate realistic multitasking and increase cognitive load during dual-task driving, an auditory-modified n-back task was used. This task targeted working memory and executive function as participants drove in a dynamic scenario. It combined elements of the n-back and digit-span tasks, creating three difficulty levels: 0-back (baseline, minimal effort), 1-back, and 2-back (increasing memory load). Participants responded to spoken digits (0–9) delivered via the simulator's speakers, using red and green buttons mounted on the steering wheel as shown in Figure 2. 0A green button

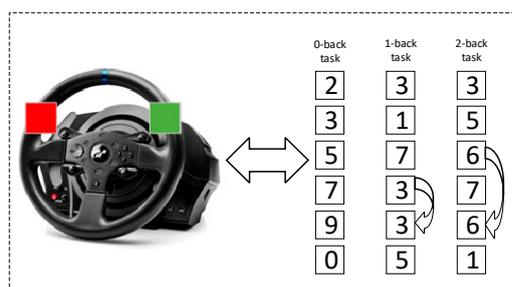indicated a match with the target digit; a red button indicated a non-match, minimizing physical distraction.

Figure 2. Steering wheel interface used during the auditory modified n-back task.

The auditory n-back tasks were developed using PsychoPy [14], which also recorded button responses. PsychoPy operated in parallel with the driving simulator to maintain an immersive, ecologically valid dual-task environment.

Experimental Procedure

The study will begin with the collection of written informed consent from all participants, ensuring they understand their rights and the voluntary nature of their involvement. Following consent, participants will attend a comprehensive briefing session, where they will receive both verbal and written instructions outlining the study's objectives, the structure of the experimental tasks, and all relevant safety procedures. After the briefing, participants will undergo a sensor fitting session. They will be equipped with the necessary physiological monitoring devices, including a high-density fNIRS system to continuously record cerebral hemodynamic responses. Once fitted, participants will complete a simulator familiarization phase, allowing them to practice basic driving maneuvers and become comfortable with the controls and environment before testing begins. The experimental phase will involve a series of structured driving tasks performed under challenging simulated conditions, including nighttime driving at approximately 1:00 AM and heavy rainfall, both intended to increase visual and attentional demands. While navigating this environment, participants will simultaneously perform the auditory-modified n-back task at varying levels of difficulty (0-back, 1-back, and 2-back), using red and green buttons mounted on the steering wheel to register their responses. This dual-task setup is designed to elicit working memory and attention, providing a realistic measure of cognitive workload under multitasking conditions. Throughout the session, fNIRS data will be collected to monitor changes in prefrontal cortex oxygenation, capturing neural responses to variations in task difficulty and environmental complexity.

Research Methodology

fNIRS data were acquired at a sampling frequency of 8.138 Hz, providing continuous, high-resolution monitoring of cortical oxygenation dynamics during task performance. The raw optical density signals collected by the fNIRS device were subjected to an initial preprocessing stage using OBELAB's integrated Digital Signal Processing (DSP) toolkit. This preprocessing pipeline included several critical steps: noise reduction to minimize ambient and physiological interference, motion artifact correction to address signal distortions caused by participant movement, and baseline drift removal to correct for slow fluctuations unrelated to task activity, thereby enhancing the overall quality and interpretability of the signals. Following preprocessing, the concentrations of HbO2 and HbR were computed using the Modified Beer–Lambert Law [15]. This standard analytical technique translates changes in optical density into quantitative measures of hemodynamic activity, offering insights into underlying neural processes.

Data Pre-Processing

In this study, Principal Component Analysis (PCA) was employed to reduce the dimensionality of the full set of 204 fNIRS channel features and to extract the most informative patterns associated with cognitive load during simulated driving scenarios. The primary goal of using PCA was to transform the original high-dimensional dataset into a lower-dimensional space while retaining the most critical information relevant to variance in the data. PCA works by identifying a new set of orthogonal axes, known as principal components, which are derived as linear combinations of the original features. These components are ordered in such a way that the first principal component accounts for the maximum variance in the dataset. The second principal component captures the maximum amount of remaining variance while being orthogonal to the first, and each subsequent component follows this principle of orthogonality and decreasing variance. By projecting the fNIRS signals onto these principal components, we aimed to highlight the dominant patterns in brain activity that correspond to changes in cognitive workload. This transformation not only improves the interpretability of the data but also enhances the efficiency and performance of downstream machine learning models by removing redundancy and noise.

Following PCA, the selected hemodynamic features were segmented into temporal windows of 10, 20, and 30 seconds, using both overlapping and non-overlapping strategies. This segmentation step was designed to systematically assess the impact of varying temporal resolutions and windowing methods on classification performance. The segmented datasets were subsequently used as input to EEGNet, a lightweight convolutional neural network initially developed for EEG signal classification. For this study, EEGNet was adapted to analyze fNIRS-derived hemodynamic patterns, aiming to classify varying levels of cognitive workload. An overview of the full data processing and classification workflow including fNIRS signal acquisition, preprocessing, windowing methodology, and EEGNet-based classification is presented in Figure 3.
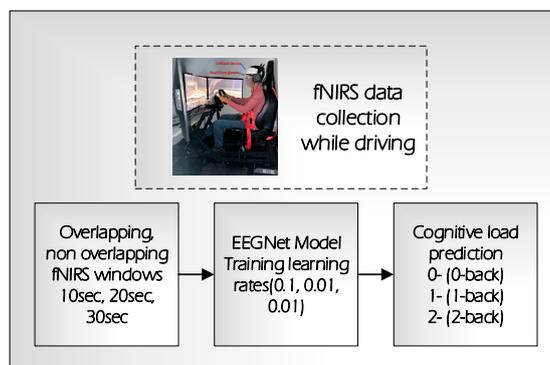

Figure 3. Schematic representation of the fNIRS-based cognitive workload classification pipeline.

EEGNet Model

To classify cognitive load across three distinct levels—namely, 0-back, 1-back, and 2-back from the fNIRS signals, this study utilized the EEGNet architecture, originally proposed by Lawhern et al. (2018) [16]. The structure and configuration details of the EEGNet model employed in this study are summarized in Table 1. The EEGNet model was selected due to its lightweight, compact design, which is particularly well-suited for learning from relatively small physiological datasets while effectively capturing spatiotemporal dependencies in neural signals. For optimization, the model was trained using the Adam optimizer, an adaptive learning rate method known for its robust convergence behavior. The EEGNet architecture is composed of three sequential convolutional blocks, each carefully designed to extract and integrate increasingly abstract spatial and temporal features from the fNIRS time-series data.

Table 1. EEGNet Model Configuration for Multilevel Cognitive Workload Detection.

Layer Type	Key Settings
2D Convolution	In-channels: 1; Out-channels: F1; Kernel size: (1, kernel length); Padding: (0, kernel length//2); No bias
Batch Normalization	Number of features: F1
Constrained Conv2D	In-channels: F1; Out-channels: F1×D; Kernel size: (Channels, 1); Max-norm constraint: 1; No bias
Batch Normalization	Number of features: F1×D
ELU Activation	-
2D Average Pooling	Kernel size: (1, 4); Stride: (1, 4)
Dropout	Drop rate: p
Depthwise 2D Convolution	In-channels: F1×D; Out-channels: F1×D; Kernel size: (1, 16); Groups: F1×D; Padding: (0,8); No bias
Pointwise 2D Convolution	In-channels: F1×D; Out-channels: F2; Kernel size: (1,1); No padding; No bias
Batch Normalization	Number of features: F2
ELU Activation	-
2D Pooling (Avg or Max)	Kernel size: (1,8); Stride: (1,8)
Dropout	Drop rate: p
Final 2D Convolution	In-channels: F2; Out-channels: N (classes); Kernel size: (1, final_conv_length); Bias: Yes
Log-Softmax Activation	Along dimension 1, [Batch, N, 1, 1]
Output Reshaping	Squeeze operation, [Batch, N]

The model architecture begins with an input layer followed by two key convolutional operations. Initially, a 2D convolution is applied to capture low-level features across both the temporal and channel dimensions. This is followed by a depthwise convolution, where each channel is processed independently with its own filter. Unlike conventional convolutions, this method substantially reduces the number of trainable parameters while preserving the ability to learn meaningful spatial patterns. Batch normalization is applied after each convolution to standardize feature distributions and promote stable learning. The use of depthwise convolution notably enhances training efficiency and reduces the risk of overfitting, which is particularly important when working with smaller neuroimaging datasets.

Building on this, the next stage incorporates a separable convolution to independently model temporal and spatial information. A depthwise convolution first extracts temporal dynamics within each feature map, followed by a pointwise (1×1) convolution that fuses information across channels, capturing cross-channel dependencies. This two-step process not only reduces computational complexity but also ensures that the model effectively captures distributed brain activity patterns across time and space. By explicitly decoupling spatial and temporal processing, the model becomes more sensitive to subtle shifts in cognitive workload. Finally, the high-level features extracted through these operations are flattened into a one-dimensional vector, making them suitable for classification. This vector is passed through a dense layer projecting the features onto three output nodes, each representing a cognitive load level (0-back, 1-back, or 2-back). A SoftMax activation function generates a probability distribution across these classes, allowing the model to express its confidence in each prediction. The training process relies on the categorical cross-entropy loss function, which measures how well the predicted probabilities match the true labels.

Results and Discussions

Given the high dimensionality of the dataset, which contains signals from 204 fNIRS channels, a feature selection step was essential to reduce computational complexity and isolate the most informative signals associated with cognitive load during simulated driving. PCA was employed to select the top 50 features that retained the highest variance, thereby preserving the most critical information relevant to differentiating cognitive states. Figure 4. illustrates the correlation matrix of these top 50 features. The matrix indicates a low level of inter-feature correlation, suggesting that the selected features are largely independent and thus contribute unique information to the predictive model. Notably, the selected features include an equal representation of HbO2 and HbR signals. This balanced distribution highlights the complementary roles of both chromophores in capturing hemodynamic responses related to varying levels of cognitive load.

Previous research has consistently shown that HbO2 signals demonstrate greater sensitivity to task-related neural activation, especially in regions such as the prefrontal cortex, which is heavily involved in executive functions and decision-making. This high sensitivity has led many studies to prioritize HbO2 signals when analyzing cognitive workload. However, our findings suggest a more balanced contribution from both HbO2 and HbR signals in predicting cognitive load during simulated driving. This indicates that while HbO_2 may be more prominent in some contexts, the inclusion of HbR provides complementary information that enhances the overall sensitivity and robustness of cognitive state detection in dynamic, real-world-like environments.
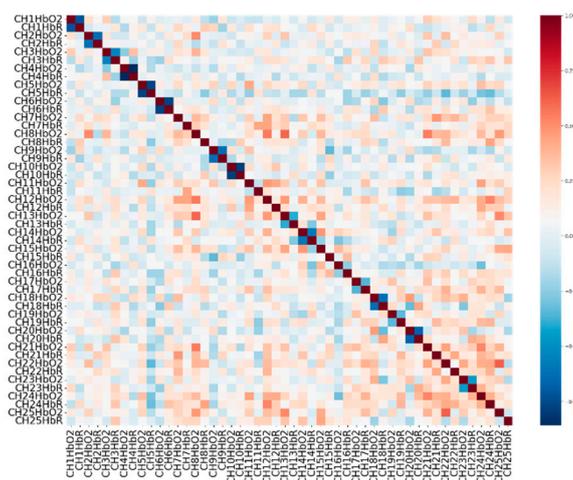

Figure 4. Heatmap of the correlation matrix for the 50 most informative fNIRS features, selected using PCA. The low correlation values suggest independence across features, supporting robust dimensionality reduction.

To train and evaluate the EEGNet model, we employed a k-fold cross-validation approach, specifically using a 5-fold cross-validation strategy, to ensure robust and generalizable performance across different subsets of the data. For model optimization, we employed the Adam optimizer due to its adaptive learning capabilities and proven effectiveness in training deep neural networks. To investigate the impact of learning rate on model convergence and classification accuracy, we experimented with three different learning rates: 0.1, 0.01, and 0.001. The model was trained for 200 epochs using a batch size of 64, providing a sufficient number of iterations for learning while maintaining computational efficiency.

To further explore the temporal sensitivity of cognitive load classification, we segmented the fNIRS time-series data into both overlapping and non-overlapping windows of 10 seconds, 20 seconds, and 30 seconds. This segmentation allowed us to evaluate how different temporal resolutions affect the model's ability to detect cognitive states. The classification results for overlapping window segmentation are summarized in Table 2., whereas the outcomes for non-overlapping segmentation are reported in Table 3.

Table 2. Comparative analysis of EEGNet performance with overlapping time segments, showing the influence of segment length on model sensitivity to cognitive load variations.

Window size	Learning rate	Accuracy		AUC		Recall		Precision		F1-score	
10s	0.1	0.5897	± 0.0476	0.7587	± 0.0416	0.5897	± 0.0476	0.6249	± 0.0397	0.6249	± 0.0397
20s	0.1	0.5029	± 0.0349	0.6958	± 0.0428	0.5029	± 0.0349	0.5952	± 0.0513	0.5952	± 0.0513
30s	0.1	0.4856	± 0.0598	0.6657	± 0.0824	0.4856	± 0.0598	0.5075	± 0.1229	0.5075	± 0.1229
10s	0.01	0.9043	± 0.0148	0.9543	± 0.0148	0.9043	± 0.0148	0.9087	± 0.0181	0.9087	± 0.0181
20s	0.01	0.9391	± 0.0071	0.9787	± 0.0048	0.9391	± 0.0071	0.9395	± 0.0072	0.9395	± 0.0072
30s	0.01	0.9048	± 0.0186	0.9498	± 0.0131	0.9048	± 0.0186	0.9113	± 0.0172	0.9113	± 0.0172
10s	0.001	0.9994	± 0.0002	0.9996	± 0.0002	0.9994	± 0.0002	0.9994	± 0.0002	0.9994	± 0.0002
20s	0.001	0.9999	± 0.0001	0.9999	± 0.0001	0.9999	± 0.0001	0.9999	± 0.0001	0.9999	± 0.0001
30s	**0.001**	**1.0000**	**± 0.0000**	**1.0000**	**± 0.0000**	**1.0000**	**± 0.0000**	**1.0000**	**± 0.0000**	**1.0000**	**± 0.0000**

Table 3. Results of EEGNet performance with non-overlapping time segments, illustrating the trade-offs in temporal granularity for cognitive load detection.

Window size	Learning rate	Accuracy	AUC	Recall	Precision	F1-score
10s	0.1	0.7079 ± 0.0202	0.8594 ± 0.0251	0.7079 ± 0.0202	0.7180 ± 0.0200	0.7180 ± 0.0200
20s	0.1	0.6700 ± 0.0173	0.8379 ± 0.0113	0.6700 ± 0.0173	0.6810 ± 0.0186	0.6810 ± 0.0186
30s	0.1	0.6548 ± 0.0257	0.7987 ± 0.0476	0.6548 ± 0.0257	0.6832 ± 0.0254	0.6832 ± 0.0254
10s	0.01	0.9508 ± 0.0089	0.9805 ± 0.0067	0.9508 ± 0.0089	0.9517 ± 0.0092	0.9517 ± 0.0092
20s	0.01	0.9470 ± 0.0115	0.9694 ± 0.0135	0.9470 ± 0.0115	0.9484 ± 0.0113	0.9484 ± 0.0113
30s	0.01	0.9100 ± 0.0302	0.9441 ± 0.0411	0.9100 ± 0.0302	0.9152 ± 0.0275	0.9152 ± 0.0275
10s	**0.001**	**0.9722 ± 0.0025**	**0.9888 ± 0.0010**	**0.9722 ± 0.0025**	**0.9726 ± 0.0024**	**0.9726 ± 0.0024**

20s	0.001	0.9335 ± 0.0077	0.9662 ± 0.0056	0.9335 ± 0.0077	0.9345 ± 0.0081	0.9345 ± 0.0081
30s	0.001	0.8545 ± 0.0232	0.9037 ± 0.0364	0.8545 ± 0.0232	0.8622 ± 0.0227	0.8622 ± 0.0227

The evaluation results using overlapping window segments demonstrate that a 30-second window, paired with a learning rate of 0.001, produces the best classification performance achieving a perfect accuracy of 100%. This finding suggests that longer, overlapping temporal segments enable the EEGNet model to more effectively capture stable and nuanced patterns of hemodynamic activity linked to varying levels of cognitive workload in fNIRS signals. The overlapping strategy ensures continuity across samples, allowing the model to learn from richer and more contextually informed representations of neural dynamics.

In contrast, the non-overlapping window segmentation reveals a different trend. The highest accuracy achieved in this setting is 97.22%, also with a learning rate of 0.001, but notably with a much shorter window length of 10 seconds. This indicates that in the absence of overlapping segments which inherently provide more data redundancy and temporal context shorter windows may help preserve temporal specificity and prevent dilution of task-related signals. Interestingly, increasing the window size in the non-overlapping configuration does not lead to improved accuracy and, in fact, results in a gradual decline in performance. This divergence highlights the critical role of segmentation strategy in shaping model sensitivity and suggests that the choice between overlapping and non-overlapping windows should be guided by both tasks demands and desired temporal resolution in cognitive state classification.

While the results of this study demonstrate the promise of using fNIRS and deep learning to classify cognitive load in a simulated driving environment, several limitations must be acknowledged. First, the participant pool was relatively small, which may affect the generalizability of the findings. Second, although the driving simulator offered a safe and controlled environment for data collection, it cannot fully replicate the dynamic and unpredictable nature of real-world driving scenarios. Additionally, the cognitive load was induced through structured n-back tasks, which, while effective for experimental control, may not reflect the full spectrum of mental demands encountered during naturalistic driving. Another limitation is the use of a single physiological modality (fNIRS); multimodal approaches could provide a more relistic assessment of driver states. Moreover, the black-box nature of deep learning models such as EEGNet presents challenges in interpretability, limiting insights into the neurophysiological underpinnings of model predictions.

Future work should focus on validating the findings in real-world driving settings using wearable fNIRS systems. Incorporating additional physiological signals like EEG, ECG, or eye-tracking could improve model robustness and allow for cross-modal verification of cognitive load. Expanding the participant pool to include more varied demographics would enhance the external validity of the model. There is also a need to develop explainable AI approaches to improve transparency and interpretability of model outputs. Finally, real-time implementation of such cognitive load monitoring systems in Advanced Driver-Assistance Systems (ADAS) could help adapt vehicle behavior based on driver state, thereby enhancing safety and driving experience.

Conclusion

This study presents a comprehensive approach to assessing cognitive workload during simulated driving using fNIRS and advanced deep learning techniques. By integrating a dual-task paradigm involving a realistic driving scenario with an auditory n-back task at three workload levels (0-back, 1-back, and 2-back), we successfully created a multi-modal experimental framework that reflects real-world multitasking demands encountered by drivers. Our findings demonstrate that fNIRS is a highly effective tool for monitoring cognitive states in dynamic environments, owing to its non-invasiveness, ease of use, and robustness against motion artifacts. Importantly, this study goes

beyond prior work by employing three levels of cognitive load, thereby offering a deeper understanding of how workload fluctuates during complex task execution. We further employed the EEGNet deep learning model, adapted for fNIRS data, and evaluated its performance using both overlapping and non-overlapping segmentation windows an area rarely explored in current literature. Additionally, we incorporated a PCA-based feature selection method to reduce computational complexity and focus on the most informative signals from the fNIRS data. This technique proved effective in isolating relevant features and improving the model's ability to classify cognitive states. Our results revealed that a learning rate of 0.001 consistently yielded the best classification performance. Specifically, a 30-second overlapping window achieved 100% accuracy, while 10-second non-overlapping segments yielded 97% accuracy, indicating that segment length and segmentation strategy can influence performance, depending on the workload context.

This research demonstrates the potential of combining neuroimaging techniques and deep learning models to enable real-time and accurate detection of mental workload in naturalistic settings. The outcomes not only contribute to the growing body of literature on cognitive state monitoring but also hold practical implications for the development of adaptive driver-assistance systems and human-aware AI agents in mixed-initiative environments. Future work may focus on validating this framework in real-world driving conditions and expanding it to include multimodal physiological data fusion for even greater precision and generalizability.

References

1. Lapierre, A., et al., *Association between clinical simulation design features and novice healthcare professionals' cognitive load: a systematic review and meta-analysis.* Simulation & Gaming, 2022. **53**(5): p. 538-563.

2. Yu, K., et al., *Exploring cognitive load through neuropsychological features: an analysis using fNIRS-eye tracking.* Medical & Biological Engineering & Computing, 2025. **63**(1): p. 45-57.

3. Ücrak, F., et al., *The Impact of Minimally Invasive Surgical Modality and Task Complexity on Cognitive Workload: An fNIRS Study.* Brain Sciences, 2025. **15**(4): p. 387.

4. Kyriaki, K., D. Koukopoulos, and C.A. Fidas, *A comprehensive survey of EEG preprocessing methods for cognitive load assessment.* IEEE Access, 2024. **12**: p. 23466-23489.

5. Nasri, M., et al. *Exploring Eye Tracking to Detect Cognitive Load in Complex Virtual Reality Training.* in *2024 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct).* 2024. IEEE.

6. Khan, M.A., et al., *Application of artificial intelligence in cognitive load analysis using functional near-infrared spectroscopy: A systematic review.* Expert Systems with Applications, 2024: p. 123717.

7. Khan, M.A., et al. *Measuring Cognitive Load: Leveraging fNIRS and Machine Learning for Classification of Workload Levels.* in *International Conference on Neural Information Processing.* 2023. Springer.

8. Benerradi, J., et al. *Exploring machine learning approaches for classifying mental workload using fNIRS data from HCI tasks.* in *Proceedings of the Halfway to the Future Symposium 2019.* 2019.

9. Ho, T.K.K., et al., *Discrimination of mental workload levels from multi-channel fNIRS using deep leaning-based approaches.* Ieee Access, 2019. **7**: p. 24392-24403.

10. Khan, M.A., et al., *Functional near-infrared spectroscopy (fNIRS) and Eye tracking for Cognitive Load classification in a Driving Simulator Using Deep Learning.* arXiv preprint arXiv:2408.06349, 2024.

11. Wang, J., et al., *Taking a Deeper Look at the Brain: Predicting Visual Perceptual and Working Memory Load From High-Density fNIRS Data.* IEEE Journal of Biomedical and Health Informatics, 2021. **26**(5): p. 2308-2319.

12. Eastmond, C., et al., *Deep learning in fNIRS: a review.* Neurophotonics, 2022. **9**(4): p. 041411-041411.

13. Fernandez Rojas, R., et al., *Empirical comparison of deep learning models for fNIRS pain decoding.* Frontiers in Neuroinformatics, 2024. **18**: p. 1320189.

14. Peirce, J., et al., *PsychoPy2: Experiments in behavior made easy.* Behavior research methods, 2019. **51**: p. 195-203.

15. Baker, W.B., et al., *Modified Beer-Lambert law for blood flow.* Biomedical optics express, 2014. **5**(11): p. 4053-4075.

16. Lawhern, V.J., et al., *EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces.* Journal of neural engineering, 2018. **15**(5): p. 056013.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.