

# Functional Near-Infrared Spectroscopy Feature Extraction with Application in Workload Estimation

Elisabeth R. M. Heremans  
*Microsoft Research*  
Redmond, WA, USA  
ORCID 0000-0001-6860-2356

David Johnston  
*Microsoft Research*  
Redmond, WA, USA  
davidjo@microsoft.com

Dimitra Emmanouilidou  
*Microsoft Research*  
Redmond, WA, USA  
ORCID 0000-0002-1789-9054

Andre Golard  
*Microsoft Research*  
Redmond, WA, USA  
golard@uw.edu

Ivan Tashev  
*Microsoft Research*  
Redmond, WA, USA  
ORCID 0000-0002-2263-2047

Ryen White  
*Microsoft Research*  
Redmond, WA, USA  
ryenw@microsoft.com

**Abstract**—Functional near-infrared spectroscopy (fNIRS) is a brain imaging technique used to estimate neuronal activity by measuring blood oxygenation. In this paper, we develop and evaluate an extensive set of fNIRS features for workload estimation, combining them with respiration and heartbeat signals. Our subject- and session-independent workload estimator is validated in a virtual flight simulator, where workload is objectively assessed based on task performance. We experiment with various regression models and feature ablations, identifying the most effective fNIRS features. The best fNIRS-based model achieves a correlation of 0.3188 with objective workload labels, improving to 0.3268 when incorporating breathing signals. This study demonstrates the value of our novel fNIRS feature set for workload estimation.

**Index Terms**—Functional near-infrared spectroscopy, workload estimation, adaptive training system.

## I. INTRODUCTION

Functional near-infrared spectroscopy (fNIRS) is a non-invasive brain imaging technique that monitors brain activity by measuring blood oxygenation in the cerebral cortex. It provides insights into cognitive processes like mental workload, making it valuable in fields such as education, healthcare, aviation, and transport. fNIRS offers advantages over traditional methods like electroencephalography (EEG), including greater robustness to movement artifacts [1], [2], which is crucial for real-time monitoring in dynamic settings.

While fNIRS shows promise for mental workload estimation [3], the literature is still developing, with most previous studies relying on basic statistical measures from fNIRS signals like mean, skewness, standard deviation (STD), kurtosis, and slope [4]–[9]. Moreover, workload estimation faces broader challenges as well. First, most workload estimation approaches tend to classify data into broad workload categories, typically ranging from two to four levels, which fail to capture the continuous nature of mental workload during tasks of varying difficulty levels. Second, defining workload is inherently difficult. The labels used to train workload predictors are often noisy and imperfect, either determined by predefined task difficulties or subjectively assessed by participants. Lastly, limited efforts have been made to develop subject-independent workload prediction models, which is crucial for broader applicability.

This study aims to advance fNIRS processing by exploring a wide range of novel and established features. Moreover, it investigates integrating additional physiological signals such as electrocardiography (ECG) and respiration to enhance performance in fNIRS-based workload estimation [10]. Building on our previous work that utilized EEG and eye gaze in flight simulation tasks [11], we now focus on fNIRS as the primary modality due to its better resilience to movement artifacts. Unlike traditional methods relying on subjective questionnaires or coarse workload categories, our approach generates fine-grained workload scores (0–100) based on objective performance metrics. By expanding fNIRS features and combining them with complementary signals, we aim to develop a robust, neural network-based workload estimator capable of accurate, real-time monitoring of cognitive workload across different individuals and settings.

## II. METHODS

### A. Experimental Setup

The flight simulation experiments for this study were performed with the Prepar3D flight simulator in virtual reality (VR) mode. The participants sat on a 6 degrees of freedom (DoF) motion platform, wearing Varjo VR3 glasses with a custom-made headset that had 20 fNIRS channels (shown in Fig. 1). The fNIRS system was adapted from the NIRSport2 of NIRx Medical Technologies LLC, and consisted of 11 light emitting diodes (wavelengths 760/850nm) and 11 detectors. The Varjo VR3 glasses also had built-in eye-tracking. In addition, one-channel ECG, galvanic skin resistance, and breathing signals were measured from each participant.

### B. Adaptive Training System for Pilot Training

The experiments involved individual trials in a virtual reality (VR) flight simulator where pilots performed one of two tasks: (a) straight-line level flight, maintaining constant speed, course, and altitude, or (b) glideslope flight, maintaining speed and course while bringing the plane to the runway. Each trial lasted 2–3 minutes, and the plane’s position was logged throughout. The root mean squared error (RMSE) between the real and ideal trajectory was rescaled to generate a performance score (0–100) [12].

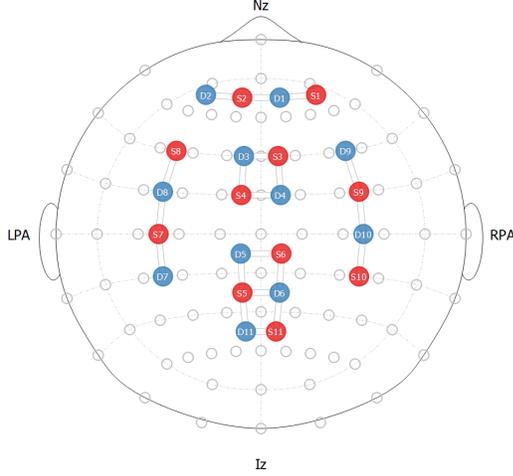


Fig. 1. Configuration of fNIRS optodes and channels.

The adaptive training system (ATS) [13] adjusts the difficulty of each trial based on past performance. It estimates pilots' skill levels and learning rates [14], recommending scenarios that optimize skill improvement. The difficulty of a scenario can be changed by incorporating factors such as wind gusts, thermals, and reducing visibility from clear conditions to fog. Taking into account both the difficulty of the scenario and the pilot's current estimated skill level, the ATS then calculates an expected score for the given task. These ATS (expected) scores serve as a strong proxy for workload, as they are derived from the two key factors influencing workload, i.e. task difficulty and skill level. Unlike raw, trial-to-trial performance scores, which can be noisy, ATS scores provide a smoother and more reliable representation of the workload associated with a task. Higher ATS scores (near 100) indicate low workload, while lower scores (near 0) signal high workload. These scores were therefore used as labels for training our workload estimator.

### C. Data Collection

Fifteen participants were initially enrolled for data collection over five consecutive days. The data collection procedure was reviewed and approved by Microsoft Research Ethical Committee Review Board. On the first day, participants completed 11 trials. Over the next four days, each completed 22 trials daily, resulting in up to 99 trials per participant. Of these 99 trials, some were lost due to signal quality or participant absence.

### D. Pre-Processing and Feature Extraction

Based on the flight logs, the performance scores were computed for each run. Then, the expected scores estimated by the ATS were extracted [11], [14] to be used as labels for training the workload estimator. To avoid bias from near-zero scores, the final training dataset included the seven participants with more than 27.5% non-zero scores. For model validation and testing, only the four participants with at least 40% non-zero scores were included, as correlations between estimated and real scores are only meaningful with a fair percentage of non-zero scores.

Artifacts were removed from the raw fNIRS signals. Each of the 11 channels consists of two signals: the HbO signal (oxygenated hemoglobin concentration) and HbR signal (deoxygenated

hemoglobin). The characteristic heartbeat pattern in the HbO signal indicates a good optode-scalp coupling and can thus be utilized as an indication of signal quality for the fNIRS channel [15]. The HbO signals were thus segmented into 10-second epochs, and each segment's quality was checked by computing its correlation with the corresponding ECG segment, and the total cardiac power using spectral analysis. Channels with over 50% bad segments were discarded. After cleaning the signals, features were extracted to facilitate further analysis.

The first set of features comprises time-domain statistics derived from the HbO signal, including the mean, variance, minimum, maximum, kurtosis, skewness, slope, power, and peak (the maximum absolute value of the signal). These features, commonly used in previous fNIRS studies [4]–[9], were calculated from the low-frequency components of the HbO signal. These components were first isolated by applying a low-pass filter below 0.5 Hz to remove the heartbeat from the signal [4], [16].

The second set of features consists of the heart rate power of the HbO signal, specifically the signal power within the frequency band corresponding to the heart rate.

The third set of features is based on correlations between fNIRS and the other physiological signals, namely breathing and ECG. We computed the correlation and delay between the two fNIRS signals themselves (HbO correlated with HbR) [6], [17], between breathing and both HbR and HbO, and ECG and both HbR and HbO.

The fourth set of features was based on the fNIRS signal segmented by using the heartbeat timings. After segmenting an fNIRS channel of a trial using these timings, all segments were shifted to start at time zero at the R-peak of the ECG QRS complex. This process, illustrated in Fig. 2 for a single fNIRS channel of a trial, resulted in a signal with a much higher (though irregular) sampling frequency, as the overlaid segments were combined. We applied a moving average strategy to this heartbeat-segmented fNIRS signal, allowing for detailed analysis of how the fNIRS signal behaves in relation to each heartbeat. From these heartbeat-segmented fNIRS signals, the following features were calculated:

- 1) the peak-to-peak amplitude of the segmented HbO and HbR signals,
- 2) the delay between the segmented HbO and HbR signal,
- 3) the upward and downward slopes of the HbO and HbR signal after each heartbeat,
- 4) the exponential decay and increase factors of the HbO and HbR signal after each heartbeat,
- 5) the delay of the upward slope and of the downward slope of both HbO and HbR signals, with respect to each heartbeat in the ECG,
- 6) the absolute rise time of both the HbO and HbR signal, and their relative rise time with respect to the heartbeat length,
- 7) the STD of the HbO and HbR signals around their respective averages (the deviation of the individual points around the moving-average curves seen in Fig. 2).

The features of the fNIRS channels were averaged over each of the five channel groups based on optode position (see Fig. 1): the frontal channels (S1-D1, S2-D1, S2-D2), the fronto-central channels (S3-D3, S3-D4, S4-D3, S4-D4), the back-central channels (S5-D5, S5-D6, S6-D5, S6-D6, S5-D11, S11-D6, S11-D11), the central-left channels (S7-D7, S7-D8, S8-D8), and the central-right channels (S9-D9, S9-D10, S10-D10).

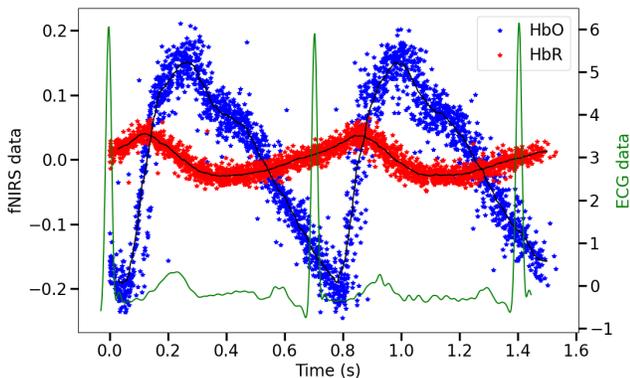


Fig. 2. An example of a heartbeat-segmented fNIRS signal (both HbO and HbR) for one channel of a trial. An example ECG segment is overlaid on top of the fNIRS to demonstrate how the two are related.

The two other physiological signals were processed as follows. The ECG signal was first band-pass filtered between 0.5 and 40 Hz, and normalized. From the pre-processed ECG signal, the heart rate and the heart rate variability were computed. The breathing signal was band-pass filtered between 0.1 and 0.5 Hz, and normalized as well. The breathing rate and breathing rate variability were computed.

### E. Machine Learning Experiments

We implemented the following estimators:

- 1) Fully connected deep neural network (FC DNN),
- 2) Long-short term memory (LSTM) neural network,
- 3) Support-vector machine (SVM) with radial basis function (RBF) kernel,
- 4) Linear regression model.

Since all estimators above used minimum mean squared error (MMSE) as cost function, we selected a different metric for evaluation: the correlation between estimated workload scores and the labels. This performance criterion evaluates the strength of the relationship and the structural fit between predictions and labels, rather than focusing solely on absolute errors. All further design decision and results are based on this evaluation criterion.

A gridsearch was performed for the FC DNN and LSTM networks for hyperparameter optimization - number of layers (1 to 3) and number of neurons (exploring powers of 2, from 4 to 128). The optimal FC DNN architecture consisted of 2 hidden layers with 64 neurons each and ReLU activation functions. The optimal LSTM network consisted of a single hidden layer of 64 neurons, and ReLU activations.

Ablations were performed to estimate the effect of the different features, and to find the best fNIRS feature set for workload estimation. Starting from the full feature set, we first removed different modalities. Then, starting from only the fNIRS features, we removed one by one the different feature groups, and fNIRS optode groups. After these ablations, we also performed an iterative ablation on the fNIRS features to find the best feature set. We started by grouping the features by channel position, with a total of five groups. With the features grouped as such, we started

TABLE I  
FEATURE ABLATIONS USING THE PHYSIOLOGICAL SIGNAL TO GROUP FEATURES.

Baseline	Ablations: modality		
All features	Heartbeat	Breathing	fNIRS
0.1073	0.1286	<b>-0.0149</b>	<b>0.0869</b>

by training our model on all features except one group. Then, permanently deleting the feature group for which the performance increase was maximal, we continued this process. After finding the best channels, we went on to group features by type. Starting with only the remaining features after the channel ablations, we removed feature group by feature group with the same iterative process until the best features were found.

We trained our models on the best fNIRS features found through iterative ablation, and we experimented with the addition of breathing and heartbeat features by early fusion and late fusion. The early fusion approach consisted of fusing the other physiological signals' features with the fNIRS features before feeding them to the model, and late fusion consisted of training a separate predictor on both fNIRS and the other features, and then training a linear regression layer to make a final prediction.

All experiments were conducted using a leave-one-subject-out cross-validation approach. In each run of cross-validation, one of the four subjects with more than 40% non-zero scores was used as the test set, and a different subject from the same group was used as the validation set. This setup ensured that the validation and test subjects were not the same in any run. The results were averaged over 12 runs, covering all possible combinations of test and validation subjects. In each cross-validation run, the training dataset consisted of the remaining two subjects (with >40% non-zero scores), along with three additional subjects with 27.5–40% non-zero scores who were not included in the validation or test sets.

### III. RESULTS

The results of the ablation studies are presented in Table I and Table II. Table I shows that the breathing features are especially informative, with fNIRS in second place. The heartbeat features proved to be the least informative, as demonstrated by the correlation increase when these features were removed. Table II shows the fNIRS feature ablations for the different groups of fNIRS features and for the different fNIRS optode positions. In terms of feature groups, the time-domain statistics from the HbO signal were the most informative, followed by the correlations of HbO and HbR with the ECG. Regarding the optode position groups, the features from the frontal channels were the most informative for workload estimation, followed by the fronto-central features. The back-central channels were the least informative.

The iterative ablation experiment used smaller feature groups, and showed more insight into which specific features were the most and least informative. Of the time-domain statistics, only the kurtosis and slope were removed, so all other features were informative. The heart rate power was retained too, as was the correlation with the ECG. The correlation with respiration and the HbO-HbR correlation were removed. Of the segmented fNIRS feature group, the features that were retained were all except three: the upslope of both the HbR and HbO signal, the exponential

TABLE II  
fNIRS FEATURE ABLATIONS FOR THE DIFFERENT GROUPS OF fNIRS FEATURES AND THE DIFFERENT OPTODE POSITIONS.

Baseline	fNIRS feature group						Optode position				
All fNIRS features	fNIRS statistics	Correlation respiration	Correlation ECG	Correlation HbO-HbR	fNIRS-segments	Heart rate power	Frontal	Fronto-central	Back-central	Central-left	Central-right
0.0728	<b>0.0249</b>	0.2071	<b>0.0240</b>	0.1388	0.0746	0.1125	<b>-0.0564</b>	<b>0.0186</b>	0.1588	0.0887	<b>0.0203</b>

TABLE III  
FINAL RESULTS OF WORKLOAD ESTIMATION WITH THE FOUR MACHINE LEARNING MODELS USING THE BEST fNIRS FEATURES, WITH AND WITHOUT THE HEARTBEAT FEATURES AND BREATHING FEATURES.

	Modalities	LIN	SVM	FC DNN	LSTM
Only fNIRS	fNIRS	0.2463	0.2259	<b>0.3188</b>	0.1857
Early fusion	fNIRS, heartbeat	0.1555	0.1609	<b>0.2049</b>	0.0748
	fNIRS, breathing	0.2516	0.2534	<b>0.3268</b>	0.0998
	fNIRS, heartbeat, breathing	0.1556	0.1872	<b>0.2703</b>	0.0862
Late fusion	fNIRS, heartbeat	0.2136	0.0811	<b>0.2036</b>	0.1899
	fNIRS, breathing	0.2655	0.2211	<b>0.3144</b>	0.2044
	fNIRS, heartbeat, breathing	0.2347	0.0910	<b>0.2796</b>	0.1721
Baseline* * results from [11]	EEG	0.1471	0.1204	0.1332	<b>0.2986</b>

factors of increase and decay, and the absolute and relative rise times. All other segmented fNIRS features were hence proven to be informative.

Table III shows the final results of the workload estimation with the best fNIRS features, with and without the addition of the other physiological signals. fNIRS managed to outperform the EEG as a modality for workload estimation [11]. The FC DNN reached the best results, with a correlation score of 0.3188 for fNIRS features alone, and a correlation score of 0.3268 for early fusion of fNIRS with the breathing features. Early fusion outperformed late fusion for the SVM and FC DNN model, while late fusion outperformed early fusion for the linear regression model and the LSTM. fNIRS outperformed the EEG baseline [11] for all models except the LSTM.

#### IV. DISCUSSION

This paper proposed novel fNIRS features and demonstrated their use in workload estimation for adaptive pilot training. Through extensive ablation experiments, we investigated which features were the most informative for this task. Our results highlight the potential of fNIRS for objective, subject-independent, and session-independent workload estimation, a relatively unexplored machine learning task.

Several important findings emerged from the ablation studies. The first key finding is the dominance of frontal channels in workload estimation, aligning with previous studies that focus on the prefrontal cortex due to its role in cognitive processing [4]–[6]. Second, time-domain statistics of the HbO signal were highly effective for predicting workload, supporting their frequent use in the fNIRS literature [4]–[9]. Of our novel features, the correlation between fNIRS and ECG emerged as very informative, as did several features from the heartbeat-segmented fNIRS signal, including the STDs, delays, peak-to-peak amplitude, and downslope. This demonstrates that the limited set of frequently used fNIRS features can be expanded to obtain more information from this rich signal. Lastly, breathing-related features significantly improved the performance of our fNIRS-based workload estimator, but the

heartbeat features did not help to improve the performance. A potential explanation is that the fNIRS features already captured most of the information extracted from the ECG, such as heart rate being partly reflected in the time constants of the heartbeat-segmented fNIRS signal. Alternatively, the heartbeat estimation using the Pan-Tompkins algorithm may not have been robust enough to handle noise in the signal, as heart rate variability is highly sensitive to missed heartbeats.

Future research will explore alternative heart rate estimation methods to enhance ECG utility, and the integration of fNIRS with modalities like EEG or eye gaze for potentially improved workload estimation. While our results demonstrate that fNIRS can outperform EEG in this context, EEG and fNIRS may still offer complementary information that could enhance overall performance [4], [18]. Lastly, a crucial next step is to implement our workload estimator in real-time for adaptive pilot training.

In conclusion, fNIRS proves to be a rich signal that provides valuable information for workload estimation, with significant potential in enhancing adaptive training systems and other real-world brain-computer interface applications.

#### REFERENCES

- [1] J. Heard, C. E. Harriott, and J. A. Adams, "A survey of workload assessment algorithms," *IEEE Transactions on Human-Machine Systems*, vol. 48, pp. 434–451, 10 2018.
- [2] P. Pinti, I. Tachtsidis, A. Hamilton, J. Hirsch, C. Aichelburg, S. Gilbert, and P. W. Burgess, "The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience," *Annals of the New York Academy of Sciences*, vol. 1464, p. 5, 2020. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/3367070/>
- [3] T. K. K. Ho, J. Gwak, C. M. Park, and J.-I. Song, "Discrimination of mental workload levels from multi-channel fnirs using deep learning-based approaches," *IEEE Access*, vol. 7, pp. 24 392–24 403, 2019.

- [4] H. Chu, Y. Cao, J. Jiang, J. Yang, M. Huang, Q. Li, C. Jiang, and X. Jiao, "Optimized electroencephalogram and functional near-infrared spectroscopy-based mental workload detection method for practical applications," *BioMedical Engineering Online*, vol. 21, pp. 1–17, 12 2022. [Online]. Available: <https://link.springer.com/articles/10.1186/s12938-022-00980-1> <https://link.springer.com/article/10.1186/s12938-022-00980-1>
- [5] U. Asgher, K. Khalil, M. J. Khan, R. Ahmad, S. I. Butt, Y. Ayaz, N. Naseer, and S. Nazir, "Enhanced accuracy for multiclass mental workload detection using long short-term memory for brain-computer interface," *Frontiers in Neuroscience*, vol. 14, p. 522313, 6 2020. [Online]. Available: [www.frontiersin.org](http://www.frontiersin.org)
- [6] M. R. Siddiquee, R. Atri, J. S. Marquez, S. M. Hasan, R. Ramon, and O. Bai, "Sensor location optimization of wireless wearable fNIRS system for cognitive workload monitoring using a data-driven approach for improved wearability," *Sensors* 2020, Vol. 20, Page 5082, vol. 20, p. 5082, 9 2020. [Online]. Available: <https://www.mdpi.com/1424-8220/20/18/5082/html> <https://www.mdpi.com/1424-8220/20/18/5082>
- [7] A. S. Le, H. Aoki, F. Murase, and K. Ishida, "A novel method for classifying driver mental workload under naturalistic conditions with information from near-infrared spectroscopy," *Frontiers in Human Neuroscience*, vol. 12, p. 384642, 10 2018.
- [8] A. Unni, K. Ihme, H. Surm, L. Weber, A. Lüdtke, D. Nicklas, M. Jipp, and J. W. Rieger, "Brain activity measured with fNIRS for the prediction of cognitive workload," in *2015 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, 2015, pp. 349–354.
- [9] T. Gateau, G. Durantin, F. Lancelot, S. Scannella, and F. Dehais, "Real-time state estimation in a flight simulator using fNIRS," *PLOS ONE*, vol. 10, no. 3, p. e0121279, mar 2015.
- [10] T. Nagasawa and H. Hagiwara, "Workload induces changes in hemodynamics, respiratory rate and heart rate variability," in *2016 IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE)*, 2016, pp. 176–181.
- [11] I. Tashev, C. Beauchene, R. M. Winters, Y. T. Wang, D. Johnston, N. Bridges, and J. R. Estepp, "Workload estimator using EEG and eye-tracking," in *EMBC 2024*, IEEE, IEEE, July 2024. [Online]. Available: <https://www.microsoft.com/en-us/research/publication/workload-estimator-using-eeeg-and-eye-tracking/>
- [12] I. J. Tashev, R. M. Winters, Y. T. Wang, D. Johnston, J. Estepp, and N. Bridges, "Towards a better scoring," *2023 IEEE Research and Applications of Photonics in Defense Conference, RAPID 2023 - Proceedings*, 2023.
- [13] I. J. Tashev, R. M. Winters, Y. T. Wang, D. Johnston, and N. Bridges, "Adaptive training system," *2023 IEEE Research and Applications of Photonics in Defense Conference, RAPID 2023 - Proceedings*, 2023.
- [14] I. J. Tashev, R. M. Winters, Y. T. Wang, D. Johnston, A. Reyes, and J. Estepp, "Modelling the training process," *2022 IEEE Research and Applications of Photonics in Defense Conference, RAPID 2022 - Proceedings*, 2022.
- [15] M. A. Yücel, A. v. Lühmann, F. Scholkmann, J. Gervain, I. Dan, H. Ayaz, D. Boas, R. J. Cooper, J. Culver, C. E. Elwell, A. Eggebrecht, M. A. Franceschini, C. Grova, F. Homae, F. Lesage, H. Obrig, I. Tachtsidis, S. Tak, Y. Tong, A. Torricelli, H. Wabnitz, and M. Wolf, "Best practices for fNIRS publications," *Neurophotonics*, vol. 8, 1 2021. [Online]. Available: [/pmc/articles/PMC7793571/](https://pubmed.ncbi.nlm.nih.gov/3571712/) [/pmc/articles/PMC7793571/?report=abstract](https://pubmed.ncbi.nlm.nih.gov/3571712/) <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7793571/>
- [16] H. J. Foy, P. Runham, and P. Chapman, "Prefrontal Cortex Activation and Young Driver Behaviour: A fNIRS Study," *PLOS ONE*, vol. 11, no. 5, p. e0156512, may 2016.
- [17] H. Aghajani, M. Garbey, and A. Omurtag, "Measuring mental workload with EEG+fNIRS," *Frontiers in Human Neuroscience*, vol. 11, 2017. [Online]. Available: <https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2017.00359>
- [18] J. Cao, E. M. Garro, and Y. Zhao, "EEG/fNIRS based workload classification using functional brain connectivity and machine learning," *Sensors*, vol. 22, no. 19, oct 2022. [Online]. Available: [/pmc/articles/PMC9571712/](https://pubmed.ncbi.nlm.nih.gov/3571712/) [/pmc/articles/PMC9571712/?report=abstract](https://pubmed.ncbi.nlm.nih.gov/3571712/?report=abstract) <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9571712/>