Exploring High-Density DOT for Three-dimensional Imaging of Mental State using CNN and Transformer Models

Chen. Jianan, *Student Member, IEEE,* Yang. Huixin, Xia. Yunjia, Li. Jinchen, *Student Member, IEEE,* Wang. Hanyang, Carlson. Tom, *Member, IEEE,* and Zhao. Hubin, *Member, IEEE*

***Abstract*—Mental fatigue and workload are critical to cognitive health, with relevance in BCIs, neurorehabilitation, and psychiatric care. Accurate decoding can support adaptive interventions and assistive system performance. High-density diffuse optical tomography (HD-DOT), an advanced form of functional near-infrared spectroscopy (fNIRS), offers high spatial resolution for non-invasive monitoring of cerebral hemodynamics, making it well-suited for detecting subtle changes in cognitive states. Deep learning models, particularly Convolutional Neural Networks (CNNs) and Transformers, are highly effective in image-based pattern recognition and thus hold promise for analyzing HD-DOT data. This study investigates the effectiveness of CNN- and Transformer models for classifying mental fatigue and workload using HD-DOT data. We collected data from 16 participants during rest, reaction time, and N-back tasks, and reconstructed 3D images of hemodynamic changes. Both two-class (low vs. high fatigue) and four-class (0,1,2,3-back) classification tasks were performed. CNN models, particularly lightweight architectures like MobileNet, demonstrated strong generalization performance under leave-one-out cross-validation, achieving up to 90.9% accuracy. Transformer models also performed competitively, with MobileViT achieving the highest classification accuracy of 98.6% in the four-class task. These findings highlight the feasibility and effectiveness of combining HD-DOT with lightweight deep learning architectures for accurate and generalizable assessment of cognitive states.**

***Clinical Relevance*— This study demonstrates the potential of HD-DOT combined with deep learning models for non-invasive, high-resolution imaging of mental fatigue and workload. Such a framework could support real-time cognitive state assessment in both clinical and occupational settings, enabling mental health monitoring, early detection of cognitive overload or decline in high-risk environments such as surgical theatres, intensive care units, and neurorehabilitation programs.**

Keywords—Mental fatigue, Mental workload, fNIRS, HD-DOT, Deep learning.

I. Introduction

Mental fatigue is a psychobiological condition that arises from prolonged cognitive engagement and is commonly manifested as subjective sensations of exhaustion and diminished vitality [1]. It can impair attention, reaction time, and decision-making [2]. Closely related to mental fatigue is mental workload, which refers to the cognitive demands placed on an individual during task performance [3]. Excessive or sustained mental workload can overwhelm cognitive resources, leading to performance degradation and increased error rates. Mental fatigue and mental workload are critical indicators of cognitive function and mental health, with increasing importance across applications such as brain-computer interfaces (BCIs), neurofeedback, neurorehabilitation, and psychiatric monitoring. Accurate decoding of these mental states is essential for enabling adaptive interventions in clinical settings, enhancing user engagement in BCI systems, and optimizing performance in mobility/cognitive assistive technologies, including smart wheelchairs and fatigue detection systems in transportation. Specifically in emerging applications like BCIs, both mental fatigue and high mental workload may impair system performance by reducing user engagement and compromising the reliability of neural signals. At the same time, BCIs themselves offer a promising approach for monitoring mental states [4], particularly when the headcap is configured to cover relevant cortical regions such as the prefrontal cortex [5].

Neuroimaging techniques have become indispensable for investigating the neural mechanisms underlying mental fatigue. Functional near-infrared spectroscopy (fNIRS), a non-invasive and even wearable modality, has proven particularly effective in real-world contexts. fNIRS offers a favourable balance of temporal and spatial resolution, low cost, and tolerance to movement. However, traditional fNIRS systems, typically consisting of around 100 or fewer channels [6], are limited in their ability to accurately localize neural sources. Although they provide millimeter-level spatial resolution, they are susceptible to contamination from superficial extracerebral hemodynamic signals [7]. This limitation reduces depth sensitivity and hinders precise three-dimensional (3D) mapping of cerebral oxygenation changes.

While traditional fNIRS systems often miss these localized dynamics due to limited channel density and depth sensitivity, high-density diffuse optical tomography (HD-DOT), an advanced neuroimaging modality, extends the capabilities of fNIRS by enabling voxel-wise functional mapping. This allows researchers to observe fine-grained patterns of neural

*Research supported by The Royal Society Research Grant (RGS\R2\222333), Engineering and Physical Sciences Research Council Grant 13171178 R00287.

Jianan Chen, Huixin Yang, Yunjia Xia, Jinchen Li, Hanyang Wang, Tom Carlson, and Hubin Zhao are with the HUB of Intelligent Neuro-engineering (HUBIN), CREATe, Division of Surgery and Interventional Science, UCL, HA7 4LP, Stanmore, UK (e-mail: Jianan.chen.22@ucl.ac.uk; huixin.yang.24@ucl.ac.uk; yunjia.xia.18@ucl.ac.uk; jinchen.li.23@ucl.ac.uk; hanyang.wang.23@alumni.ucl.ac.uk; t.carlson@ucl.ac.uk; hubin.zhao@ucl.ac.uk).

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Committee for Research Ethics at UCL, London, under Application No 6860.017; 6860.018; 17599.002.

compensation or disengagement over time [8]. With a dense array of overlapping measurements across multiple source-detector distances [9], HD-DOT enables tomographic reconstruction of haemoglobin concentration changes across the cortical surface, achieving spatial resolution comparable to that of functional Magnetic Resonance Imaging (fMRI) [10]. This makes HD-DOT particularly well-suited for studying mental fatigue, where subtle and spatially heterogeneous changes in prefrontal activity, linked to sustained cognitive effort and attentional decline, must be captured with high fidelity.

Given HD-DOT's ability to capture fine-grained, voxel-wise brain activity, analyzing its high-dimensional outputs requires models capable of extracting and interpreting complex spatial patterns. Deep learning approaches, particularly convolutional neural networks (CNNs) and Transformer architectures, are well-suited for this purpose. Specifically, CNNs are well-suited for processing 3D neuroimaging data due to their ability to extract hierarchical spatial features, whereas Transformer-based models offer the advantage of capturing global contextual relationships through self-attention mechanisms [13]. To date, multiple studies have demonstrated the effectiveness of deep learning such as CNNs in mental states assessment with fNIRS [11], [12], few have explored the potential of deep learning models applied to 3D-reconstructed HD-DOT images. This study aims to bridge this gap by evaluating both CNNs and Transformers for mental state classification using HD-DOT-derived 3D data.

To mitigate overfitting risks associated with limited data, we adopted a transfer learning approach by fine-tuning models pre-trained on ImageNet, which is a large-scale hierarchical image database [14]. Despite the large domain gap, transfer learning from ImageNet has been shown to be effective in medical imaging domains, including fMRI, Positron Emission Tomography (PET), and Electroencephalography (EEG) [15].

By comparing CNN and Transformer models on binary and multi-class tasks, we assess their effectiveness for HD-DOT-based mental state decoding and identify which architecture better captures the spatial-functional patterns in HD-DOT data. As shown in Fig. 1, this analysis offers insights into model-specific strengths and informs future deep learning strategies for optical neuroimaging.

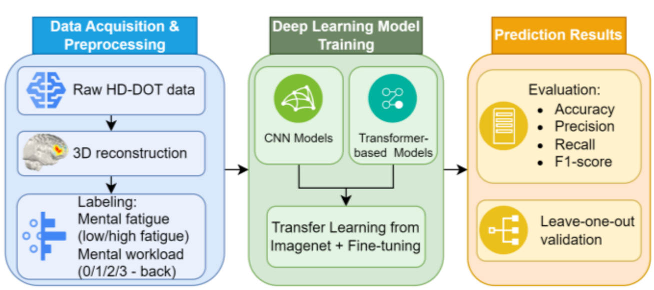

Figure 1. Data process pipeline, including data preparation, model training and evaluation.

II. METHODS

A. Data Acquisition

Data were collected using a HD-DOT system LUMO (Gowerlabs Ltd, UK) [16], consisting of 12 tiles, each equipped with 3 sources and 4 detectors, yielding a total of 1728 theoretical channels across two wavelengths (735/850 nm). As shown in Fig.2, sixteen participants completed: (1) 2-minute rest, (2) 2-minute reaction time task, (3) 33-minute N-back task (0-, 1-, 2-, 3-back), and (4) 2-minute post-test reaction time task, which is the same protocol as in [17]. All procedures were approved by the UCL Research Ethics Committee. N-back was used to induce mental fatigue.

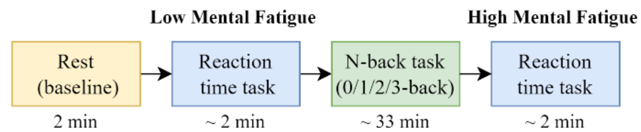

Figure 2. Experiment protocol.

B. Data Preprocessing

Raw data were preprocessed in MATLAB 2022b using the DOTHUB toolbox [18]. Motion artefacts were corrected via Kurtosis-Wavelet filtering [19]. To ensure data quality, channels were excluded from analysis if they exhibited poor signal quality according to three criteria: (1) the mean optical density fell outside a predefined range, indicating signal saturation or insufficiency; (2) the signal-to-noise ratio, computed as the mean divided by the standard deviation, was below a threshold; or (3) the source-detector distance lay outside the physiologically valid range. These criteria were applied to each wavelength. A 0.025-0.15 Hz bandpass filter were applied. Oxygenated (HbO) and deoxygenated hemoglobin (HbR) were extracted after short-channel regression was used to remove superficial blood oxygenation changes caused by the scalp.

C. Image Reconstruction

For HD-DOT, 3D hemoglobin images are reconstructed by solving a model-based inverse problem that links channel-level measurements to volumetric tissue absorption changes. After obtaining channel-wise HbO/HbR changes (derived from raw intensity data), a physics-based forward model of light transport is employed to characterize photon propagation from each source to detector. This forward model (often using the diffusion approximation to the radiative transfer equation on a head mesh) provides a sensitivity matrix (Jacobian) that quantifies how a unit absorption change at each 3D location affects each source-detector channel's measurement [20]. The resulting Jacobian is then used in the inverse model to estimate the spatial distribution of hemoglobin changes that best explain the observed channel data. Due to the ill-posed nature of the tomographic inverse problem, Tikhonov regularization was applied with a regularization parameter $\lambda = 0.01$. The resulting inverse Jacobian was used to reconstruct spatially resolved maps of HbO and HbR concentration changes, which were then projected onto the cortical surface for visualization of task-related hemodynamic responses.

D. Dataset generation

For the mental fatigue classification task, data were extracted from the first and second reaction time sessions. For the mental workload classification task, data from six repetitions of the N-back task (0-, 1-, 2-, and 3-back) were collected for each subject, and the average across trials was computed for each condition to improve signal stability. This trial-averaging approach helps reduce random noise and enhance the reliability of task-evoked responses. However, since each subject contributes only one averaged sample per condition, the overall sample size remains limited. To avoid the risk of overfitting, a multi-view data augmentation strategy was applied. Each 3D HD-DOT image was rendered from multiple horizontal viewing angles to increase spatial variability while preserving the original class label. Specifically, the image was rotated horizontally around the vertical (Y) axis in increments of 5°, generating views from 90° to 270°, as shown in Fig.3. This approach enhances model generalization by exposing it to a broader range of spatial representations, and has been shown effective in similar neuroimaging and multi-view classification tasks [21]. Each image was then cropped and resized to 224×224 pixels to match the input dimensions required by CNN and Transformer models. This process results in a total of 3,259 training samples. The dataset size for the binary classification task and the four-class classification task are 1086 and 2173, respectively. The dataset was split into the training set, validation set, and test set in an 8:1:1 ratio.

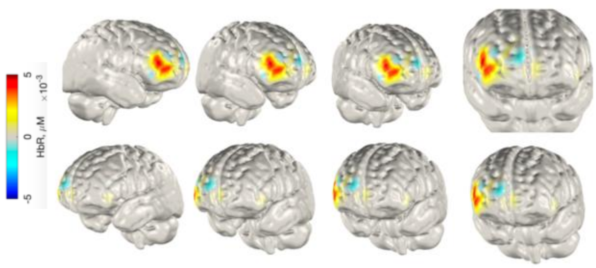

Figure 3. Reconstructed cortical activation maps from HD-DOT data showing HbR changes. Multiple angles are used to provide a full view of spatially distributed prefrontal responses.

E. *Model Training*

1) *Model selection:*

CNN models (AlexNet, GoogLeNet, ResNet, DenseNet, MobileNet) and Transformer models (ViT, Swin Transformer, MobileViT) were trained using PyTorch. These models were selected to compare depth, connectivity, and computational efficiency, all of which are critical design factors in real-time neuroimaging applications. To enhance generalization and mitigate overfitting due to the limited dataset size, transfer learning was adopted. Model weights were initialized using parameters pre-trained on ImageNet, a large-scale dataset of natural images widely used for feature extraction. Pre-trained ImageNet weights were fine-tuned for two classification tasks:

- Task I: two-class (low vs. high fatigue)
- Task II: four-class (0-3 back).

A two-phase fine-tuning strategy was subsequently employed. In the first phase, the early convolutional layers were frozen while the final fully connected layer was retrained, with output dimensions modified to 2 and 4 for binary and four-class classification tasks, respectively. A dropout layer was introduced to further reduce overfitting and improve generalizability. In the second phase, all layers were unfrozen and fine-tuned using a reduced learning rate, allowing the model to progressively adapt to task-specific features while retaining the advantages of the pre-trained representations.

2) *Learning rate and batch size:*

The initial learning rate was set to 0.001 and combined with a learning rate decay schedule to mitigate overfitting and improve convergence. To further enhance generalization, L2 regularization was incorporated into the loss function. A batch size of 32 was used throughout training. Training was conducted over a maximum of 50 epochs, with early stopping applied using a patience value of 5. The cross-entropy loss function was selected to evaluate the model's bias in the classification task.

III. RESULTS AND DISCUSSION

We evaluated the model using four standard metrics: Accuracy, Precision, Recall, and F1-score. Accuracy reflects the overall correctness of predictions, Precision measures the correctness of positive predictions, Recall assesses the ability to identify all positive cases, and F1-score balances Precision and Recall. As shown in Fig. 4, most of the models exceeded 90% accuracy. For binary classification, ViT achieved the highest performance (97.9%), while GoogLeNet led among CNNs (94.5%). In four-class classification, MobileViT outperformed others with 98.6% accuracy.

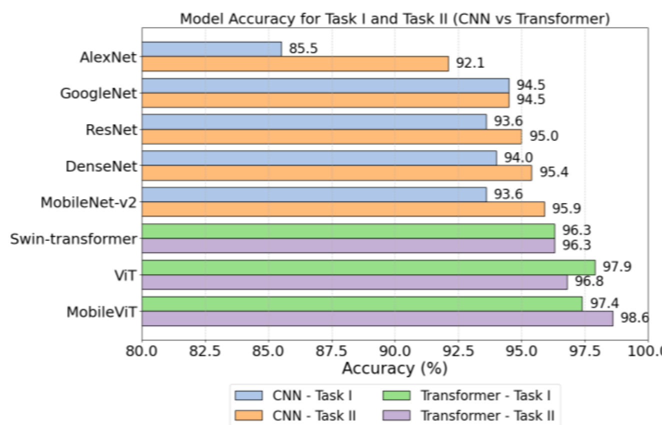

Figure 4. Model accuracy of CNNs and Transformer.

A. *CNN Models*

Among the five CNN models evaluated, GoogLeNet achieved the highest classification accuracy in distinguishing mental fatigue, outperforming ResNet, MobileNet, DenseNet, and AlexNet (Table. I: Task I). Its superior performance may attribute to the architectural design of the Inception modules, which effectively balance network depth and width. In contrast, AlexNet demonstrated the lowest accuracy across all models. Its relatively shallow architecture and use of fixed-size convolutional kernels limit its capacity to extract detailed and hierarchical features. This design constraint reduces its effectiveness in tasks that require fine-grained discrimination, such as detecting nuanced changes in brain activation related to cognitive fatigue.

TABLE I. CNN MODELS' PERFORMANCE ON TASK I AND TASK II

Model	**Task I: Mental Fatigue**			
	Accuracy (%)	***Precision (%)***	***Recall (%)***	***F1-score (%)***
AlexNet	85.5	83.1	89.1	86.0
GoogLeNet	94.5	94.5	94.5	94.5
ResNet	93.6	94.4	93.6	93.6
DenseNet	94.0	94.1	94.1	93.8

Model	Task I: Mental Fatigue			
	Accuracy (%)	*Precision (%)*	*Recall (%)*	*F1-score (%)*
MobileNet-v2	93.6	94.4	93.6	93.6
Model	**Task II: Mental Workload**			
	Accuracy (%)	*Precision (%)*	*Recall (%)*	*F1-score (%)*
AlexNet	92.1	92.0	91.8	91.8
GoogLeNet	94.5	94.7	94.5	94.5
ResNet	95.0	95.2	95.0	95.0
DenseNet	95.4	95.9	95.4	95.4
MobileNet-v2	95.9	95.9	95.9	95.9

Overall, the CNN models demonstrated improved performance on the four-class mental workload classification task compared to the binary mental fatigue task (Table. I: Task II). One contributing factor is the increased dataset size in the four-class setting. This broader training set exposed the models to a wider range of hemodynamic patterns, enhancing their ability to generalize and learn more discriminative features. From a feature space perspective, the four workload levels present more structured and separable activation patterns. The presence of progressive cognitive load across the 0-back to 3-back conditions may introduce useful relational patterns and feature redundancy, which the models can leverage to improve classification performance.

Among the CNNs evaluated, ResNet, DenseNet, and MobileNet exhibited the most robust performance in the four-class task. ResNet's residual connections allow for effective gradient flow in deeper architectures, which is crucial for capturing subtle distinctions across similar cognitive states. DenseNet's densely connected layers promote feature reuse and gradient efficiency, making it particularly suitable for processing the distributed and overlapping activation patterns found in HD-DOT data. MobileNet, while lightweight, achieves strong performance by using depth-wise separable convolutions that maintain expressiveness with fewer parameters, making it highly suitable for resource-constrained or real-time applications such as wearable BCIs or mobile neurofeedback systems.

B. Transformer-based Models

To complement the evaluation of CNN architectures, we also investigated three Transformer-based models: ViT, Swin Transformer, and MobileViT, chosen for their varying trade-offs between global context modeling, local feature aggregation, and computational efficiency. This comparison aimed to assess whether self-attention-based mechanisms can effectively exploit the complex spatial features embedded in HD-DOT neuroimaging data.

In the binary classification of mental fatigue (Table. II: Task I), ViT achieved the highest accuracy at 97.9%, outperforming both MobileViT and Swin Transformer. ViT's use of global self-attention enables it to model long-range dependencies across the entire spatial field, which is particularly advantageous for detecting subtle, distributed changes in cerebral hemodynamics. In contrast, the Swin Transformer adopts a hierarchical structure with local self-attention windows, which improves computational scalability but may restrict its ability to capture global patterns, limiting performance in tasks that require holistic spatial understanding.

TABLE II. TRANSFORMER-BASED MODELS' PERFORMANCE ON TASK I AND TASK II

Model	Task I: Mental Fatigue			
	Accuracy (%)	*Precision (%)*	*Recall (%)*	*F1-score (%)*
ViT	97.9	98.1	96.3	97.2
Swin-transformer	96.3	100	92.6	96.2
MobileViT	97.4	100	94.4	97.1
Model	**Task II: Mental Workload**			
	Accuracy (%)	*Precision (%)*	*Recall (%)*	*F1-score (%)*
ViT	96.8	96.9	96.8	96.8
Swin-transformer	96.3	96.5	96.3	96.3
MobileViT	98.6	98.5	98.6	98.6

In the four-class mental workload classification task, all three Transformer-based models maintained high accuracy, with MobileViT demonstrating the strongest performance (Table. II: Task II). Compared to its performance in the binary task, MobileViT showed notable improvement in the multi-class setting. Unlike ViT, MobileViT combines convolutional inductive bias with efficient self-attention modules, enabling better generalization in small-sample settings. Its hybrid design enhances multi-scale feature learning while maintaining a lightweight architecture—making it particularly suitable for real-time applications such as wearable BCIs, closed-loop neurofeedback, and adaptive cognitive monitoring. MobileViT's balanced architecture thus offers a compelling solution for tasks requiring both fine-grained classification and computational efficiency.

C. Generalization Capability

The generalization performance of CNN and Transformer models was also evaluated using a leave-one-out cross-validation strategy. The four-class mental workload classification task was conducted on the 16 subjects ($N = 16$). In each iteration, one participant's data was held out as the test set, while the remaining $N-1$ subjects formed the training set. This process was repeated N times, ensuring that each subject was used once as the validation set. Final performance metrics were calculated by averaging the results across all folds.

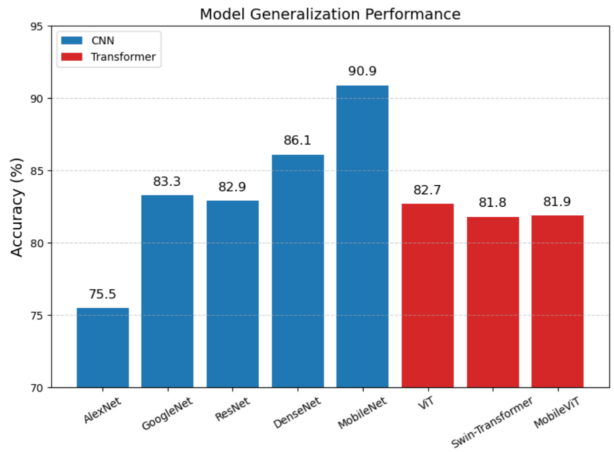

Figure 5. Leave-one-subject-out cross-validation results comparing the generalization performance of CNN and Transformer models in mental workload classification tasks.

As shown in Fig.5, among the CNN models, all except AlexNet achieved over 80% accuracy, with MobileNet demonstrating the best generalization performance at 90.9%. This result highlights the strength of lightweight architectures like MobileNet in small-sample neuroimaging contexts, where reduced parameter counts help mitigate overfitting and training instability. In contrast, AlexNet underperformed, likely due to its shallow architecture and limited feature extraction capacity, which are insufficient for capturing complex spatiotemporal patterns in HD-DOT data.

Transformer-based models also demonstrated strong generalization, with all accuracy scores around 82%, confirming their capacity to handle multi-class classification tasks. However, in this setting, Transformer models did not outperform CNNs, in contrast to earlier experiments. This may be due to the relatively small training dataset, as Transformers typically require larger data volumes to fully leverage their representational power and avoid overfitting, given their higher parameter complexity.

IV. Conclusion

This study addresses two key gaps in neuroimaging-based mental state classification. First, while conventional fNIRS systems have been increasingly used for analyzing mental fatigue and workload, their limited spatial resolution restricts their effectiveness in capturing distributed brain activity; meanwhile, HD-DOT offers fine-grained, 3D mapping of cortical hemodynamics, making it well-suited for decoding subtle cognitive states, however, its potential remains underutilized. Second, despite the growing success of Transformer models in computer vision, their application to HD-DOT data, especially in comparison to CNNs, has been largely unexplored.

To bridge these gaps, we systematically evaluated CNN and Transformer models on 3D HD-DOT data across binary (fatigue) and four-class (workload) classification tasks. CNNs, particularly MobileNet, demonstrated strong generalization even with limited data, while Transformer models such as ViT showed competitive performance, especially in binary classification. However, their reliance on larger datasets was evident under cross-validation.

Overall, our findings demonstrate the feasibility and effectiveness of combining HD-DOT with lightweight deep learning models for accurate mental state decoding. This integrated approach offers a powerful tool for real-time, non-invasive imaging of cognitive states, supporting applications in closed-loop neurofeedback, adaptive BCIs, and clinical decision-making. By enabling fine-grained assessment of mental fatigue and workload, such systems hold significant promise for deployment in high-stakes environments such as surgical theatres, intensive care units, and neurorehabilitation settings, where early detection of cognitive decline or overload is critical for safety and therapeutic outcomes.

Future work will focus on addressing current limitations and extending the applicability of our findings. A key direction is to further validate the unique advantages of HD-DOT in classifying mental workload and fatigue. While HD-DOT offers higher spatial resolution than traditional fNIRS and EEG, direct comparisons are needed to quantify its added value in terms of classification performance and cortical specificity. We also aim to identify the most informative cortical regions using explainability techniques such as Grad-CAM and ROI-based analysis. To move beyond static 3D classification, we will explore spatiotemporal modeling to capture the dynamic nature of cognitive states. Additionally, we plan to improve model generalizability in small-sample scenarios by applying lightweight Transformer architectures and domain adaptation techniques. Individual variability will also be addressed through personalized modeling or participant clustering, supporting more robust brain-state decoding in real-world applications.

References

[1] M. A. S. Boksem and M. Tops, "Mental fatigue: Costs and benefits," *Brain Res Rev*, vol. 59, no. 1, pp. 125–139, Nov. 2008, doi: 10.1016/j.brainresrev.2008.07.001.

[2] H. M. Gavelin, A. S. Neely, I. Aronsson, M. Josefsson, and L. Andersson, "Mental fatigue, cognitive performance and autonomic response following sustained mental activity in clinical burnout," *Biol Psychol*, vol. 183, p. 108661, Oct. 2023, doi: 10.1016/j.biopsycho.2023.108661.

[3] A. Hamann and N. Carstengerdes, "Investigating mental workload-induced changes in cortical oxygenation and frontal theta activity during simulated flights," *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-10044-y.

[4] J. Chen *et al.*, "fNIRS-EEG BCIs for Motor Rehabilitation: A Review," *Bioengineering*, vol. 10, no. 12, p. 1393, Dec. 2023, doi: 10.3390/bioengineering10121393.

[5] Y. Yan, Y. Guo, and D. Zhou, "Mental fatigue causes significant activation of the prefrontal cortex: A systematic review and meta-analysis of fNIRS studies," *Psychophysiology*, vol. 62, no. 1, Jan. 2025, doi: 10.1111/psyp.14747.

[6] J. Uchitel, E. E. Vidal-Rosas, R. J. Cooper, and H. Zhao, "Wearable, integrated eeg–fnirs technologies: A review," *Sensors*, vol. 21, no. 18, Sep. 2021, doi: 10.3390/s21186106.

[7] J. Chen *et al.*, "Simultaneous Mental Fatigue and Mental Workload Assessment With Wearable High-Density Diffuse Optical Tomography," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 33, pp. 1242–1251, 2025, doi: 10.1109/TNSRE.2025.3551676.

[8] H. Zhao *et al.*, "Design and validation of a mechanically flexible and ultra-lightweight high-density diffuse optical tomography system for functional neuroimaging of newborns," *Neurophotonics*, vol. 8, no. 01, Mar. 2021, doi: 10.1117/1.nph.8.1.015011.

[9] E. E. Vidal-Rosas, A. von Lühmann, P. Pinti, and R. J. Cooper, "Wearable, high-density fNIRS and diffuse optical tomography

technologies: a perspective," *Neurophotonics*, vol. 10, no. 02, May 2023, doi: 10.1117/1.nph.10.2.023513.

[10] M. D. Wheelock, J. P. Culver, and A. T. Eggebrecht, "High-density diffuse optical tomography for imaging human brain function," *Review of Scientific Instruments*, vol. 90, no. 5, May 2019, doi: 10.1063/1.5086809.

[11] S. Karmakar *et al.*, "Real time detection of cognitive load using fNIRS: A deep learning approach," *Biomed Signal Process Control*, vol. 80, p. 104227, Feb. 2023, doi: 10.1016/j.bspc.2022.104227.

[12] N. E. Mughal, K. Khalil, and M. J. Khan, "FNIRS Based Multi-Class Mental Workload Classification Using Recurrence Plots and CNN-LSTM," in *AIMS 2021 - International Conference on Artificial Intelligence and Mechatronics Systems*, Institute of Electrical and Electronics Engineers Inc., Apr. 2021. doi: 10.1109/AIMS52415.2021.9466084.

[13] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in Vision: A Survey," *ACM Comput Surv*, vol. 54, no. 10s, pp. 1–41, Jan. 2022, doi: 10.1145/3505244.

[14] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2009, pp. 248–255. doi: 10.1109/CVPR.2009.5206848.

[15] Z. Ardalan and V. Subbian, "Transfer Learning Approaches for Neuroimaging Analysis: A Scoping Review," *Front Artif Intell*, vol. 5, Feb. 2022, doi: 10.3389/frai.2022.780405.

[16] Gowerlabs Ltd., "LUMO - fNIRS neuroimaging device." Accessed: May 27, 2025. [Online]. Available: https://www.gowerlabs.co.uk/lumo

[17] J. Chen, Y. Xia, A. Thomas, T. Carlson, and H. Zhao, "Mental Fatigue Classification with High-Density Diffuse Optical Tomography: A Feasibility Study," in *Proceedings of the The Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2024*, Orlando, Florida: IEEE, 2024.

[18] DOT-HUB, "DOT-HUB_toolbox. GitHub." Accessed: May 27, 2025. [Online]. Available: https://github.com/DOT-HUB/DOT-HUB_toolbox

[19] B. Molavi and G. A. Dumont, "Wavelet based motion artifact removal for Functional Near Infrared Spectroscopy.," *Annu Int Conf IEEE Eng Med Biol Soc*, vol. 2010, pp. 5–8, 2010, doi: 10.1109/IEMBS.2010.5626589.

[20] S. Srinivasan, D. Acharya, E. Butters, L. Collins-Jones, F. Mancini, and G. Bale, "Subject-specific information enhances spatial accuracy of high-density diffuse optical tomography," *Frontiers in Neuroergonomics*, vol. 5, Feb. 2024, doi: 10.3389/fnrgo.2024.1283290.

[21] L. Chen, H. Qiao, and F. Zhu, "Alzheimer's Disease Diagnosis With Brain Structural MRI Using Multiview-Slice Attention and 3D Convolution Neural Network," *Front Aging Neurosci*, vol. 14, Apr. 2022, doi: 10.3389/fnagi.2022.871706.