COGNITIVE LOAD ESTIMATION USING BRAIN FOUNDATION MODELS AND INTERPRETABILITY FOR BCIS

Deeksha M. Shama[1*]   *Dimitra Emmanouilidou*[2]   *Ivan J. Tashev*[2]

[1]Johns Hopkins University, Baltimore, MD, USA
[2] Microsoft Research, Redmond, WA, USA

ABSTRACT

Accurately monitoring cognitive load in real time is critical for Brain-Computer Interfaces (BCIs) that adapt to user engagement and support personalized learning. Electroencephalography (EEG) offers a non-invasive, cost-effective modality for capturing neural activity, though traditional methods often struggle with cross-subject variability and task-specific preprocessing. We propose leveraging Brain Foundation Models (BFMs), large pre-trained neural networks, to extract generalizable EEG features for cognitive load estimation. We adapt BFMs for long-term EEG monitoring and show that fine-tuning a small subset of layers yields improved accuracy over the state-of-the-art. Despite their scale, BFMs allow for real-time inference with a longer context window. To address often-overlooked interpretability challenges, we apply Partition SHAP (SHapley Additive exPlanations) to quantify feature importance. Our findings reveal consistent emphasis on prefrontal regions linked to cognitive control, while longitudinal trends suggest learning progression. These results position BFMs as efficient and interpretable tools for continuous cognitive load monitoring in real-world BCIs.

Index Terms— Brain Foundation Models, Explainability, EEG, Brain Computer Interfaces, Cognitive Load

1. INTRODUCTION

Cognitive load estimation plays a pivotal role in enabling intelligent systems that adapt to users' mental states. Applications include adaptive learning systems that adjust instructional content based on cognitive state; personalized delivery platforms that respond to user engagement; and neuroadaptive games that modulate difficulty or pacing in real time. These systems show promise towards enhancing user experience, improving learning outcomes, and supporting mental well-being through passive, brain-informed adaptation. In this regard, Brain-Computer Interfaces (BCIs) offer a promising avenue for passive cognitive load monitoring by leveraging physiological signals such as Eye Gaze [1], Heart Rate [2], Electrocardiography (ECG) [3], and Electroencephalography (EEG) [4]. Among these, EEG stands out as a non-invasive, portable, cost-effective modality providing direct access to neural activity.

EEG signals are inherently complex, characterized by high dimensionality, spatio-temporal dynamics, and inter-subject variability [5, 6]. Traditional machine learning approaches rely on handcrafted features such as power spectral density and functional connectivity [7–9], which often require extensive preprocessing and task-specific configurations [10]. While effective in controlled settings, these methods struggle to generalize across tasks, users,

and recording conditions [10, 11]. Deep learning models, including CNNs, recurrent LSTMs, and more recently transformers, have shown promise in learning features directly from raw or minimally processed EEG [12–15]. Yet, these models are typically *trained for specific tasks* from scratch, making them task-specific and thus may lack flexibility and scalability. Their performance often degrades when applied to new users or tasks, limiting their utility in real-world BCIs systems and cross-subject requirements.

In contrast, Brain Foundational Models (BFMs) have emerged as a new paradigm for EEG-based BCIs. These large-scale models are pre-trained on diverse EEG datasets via self-supervised objectives toward *generalizable representations* of brain activity. BFMs can be adaptable to downstream tasks with minimal fine-tuning [16–20]. BFMs show great potential for short-duration EEG classification tasks, but their application to continuous cognitive load monitoring requiring long-term EEG modeling remains underexplored [9]. Moreover, it is crucial to ground these advanced models in established neuroscientific cognitive theory [21]. This mandates interpretable feature analyses, often missing in large-scale BFMs [22].

Herein, we investigate the use of BFMs for continuous cognitive-load monitoring and examine key challenges in scalability, generalization, interpretability. To our knowledge, this is the first work to apply BFMs to continuous cognitive-load estimation and to analyze their behavior in a multi-day training setting. Our contributions are:

- **A scalable and cross-participant pipeline** for long-term cognitive load estimation using BFM-derived features.

- **A flexible group-average channel alignment** for heterogeneous layouts, improving cross-subject generalization.

- **An adaptation of Partition SHAP** to interpret EEG feature and region importance, aligned with neuroscience [23].

- **A longitudinal analysis** across multiple days revealing learning progression w.r.t. cognitive load and other neural markers.

We find that cognitive load decreases over time while prefrontal neural relevance increases. Our results further show that BFMs, particularly LaBraM [18], improve estimation accuracy and consistently emphasize frontal regions linked to working memory and executive function, supporting their use in real-world cognitive monitoring.

2. METHODS

2.1. Data Collection and EEG Preprocessing

Five consecutive data cohorts A, B, C, D, and E were collected over the span of 3 years, each with similar but slightly modified channel configurations or hardware. Recruited participants sat on a 6 DoF chair in a Prepar3D Virtual Reality (VR)-based flight simulator,
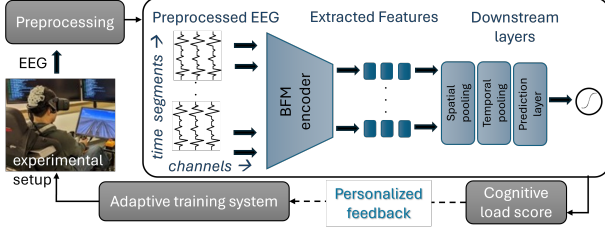

Fig. 1: Overall pipeline of cognitive load estimation with brain foundation model (BFM) in adaptive training systems.

Group	Electrodes *(when present)*
#1 (F3)	F3, F5, AF3, AF7, FP1
#2 (F4)	F4, F6, AF4, AF8, FP2
#3 (FC3)	FC3, FC5, FC1
#4 (FC4)	FC4, FC6, FC2
#5 (C3)	C3, C4, C1
#6 (C4)	C4, C6, C2
#7 (CP3)	CP3, CP4, CP1
#8 (CP4)	CP4, CP6, CP2
#9 (PO)	P1,P3, P5, PO3, P2, P4, P6, PO4, PZ, POZ

Fig. 2: Electrodes and groupings. Green fill: part of 32-chann setup, underline: part of 28-chann setup, purple outline: part of 26-chann.

wearing a Varjo VR3 headset with custom dry electrodes of varying $\{26, 28, 32\}$-channel configurations (besides ground and reference). Custom single-pin, spring-loaded EEG electrodes were developed, gold-pinned or Ag/AgCl-coated depending on the cohort, and were embedded in an adjustable three-plate 3D printed hardware shell. The EEG sensors were connected to a BrainVision Active/Dry electrode interfaces and a LiveAmp sampling at 500 Hz. Eye-tracking data was also collected from the integrated Varjo VR3 headset. The study was approved by the Microsoft Research Ethics IRB.

Most participants were inexperienced in VR or in flying and were presented with two tasks at varying difficulty levels (created by simulated weather conditions of wind, turbulence, visibility) [9]: approach the runway with constant course and speed or maintain constant altitude, speed, and course on a level flight. The study included 30 participants total, who, over 5 consecutive days, completed 90 trials, each lasting ~ 2 minutes (shorter if simulated aircraft crashed). Some sessions were missed due to participant no show or discarded due to poor experimental conditions. Overall, as participants underwent consecutive training sessions, their performance was expected to improve. Participants were grouped into temporally defined cohorts, collected sequentially during the study. Cohorts D and E are considered the most stable in terms of experimental setup, reflecting improvements to the acquisition protocol, custom EEG cap design, and signal quality management. Hence, cohort E is reserved exclusively for evaluation (see more in Section 2.5). The cohort details:

Cohort A:	2 participants,	26 channels
Cohort B:	6 participants,	26 channels
Cohort C:	6 participants,	28 or 32 channels (varied)
Cohort D:	11 participants,	32 channels
Cohort E:	5 participants,	32 channels

Prior to feature extraction, EEG signals were band-pass filtered within $[0.1 - 75]$ Hz, with a 60 Hz notch filter to remove electrical noise, resampled to 200 Hz. On average, most trials were $\sim$100 sec long (maxed at 2 min). We extract 90 sec from the center of each trial, cropping or zero-padding to ensure uniformity (less than 10% of segments were cropped/padded). Segments were then split into 16-sec windows with 50% overlap, chosen as the maximum length that could be processed by the BFMs. The resulting EEG input is denoted as follows, where number of windows $N_T = (90-16)/8+1$ = 10, number of electrodes N_E, and $N_S = 16 \times 200 = 3200$ samples:

$$x \in \mathbb{R}^{N_T \times N_E \times N_S}$$

2.2. Ground-truth Labels from Adaptive Training System

In our study we make use of the comprehensive score from the Adaptive Training System (ATS) for Pilots [24] as the *objective* estimate of cognitive load, designed for pilot training in VR flight simulators. These are **continuous ground truth label scores**, that integrate flight logs, task difficulty, participant skill, and learning rate, aligned with the cognitive load theory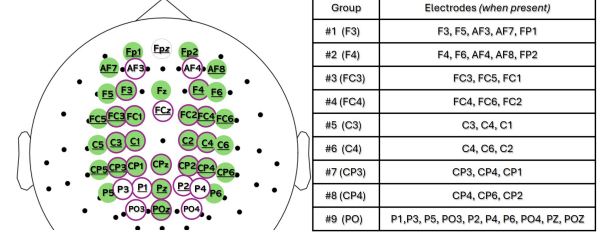: when difficulty and skill are controlled, performance is largely determined by the cognitive effort exerted [21]. The overall process can be seen in Fig. 1.

2.3. Feature Extraction with Brain Foundation Models

We extract features from preprocessed EEG using two recently released BFMs: LaBraM [18] and CBraMod [19]. LaBraM employs a convolutional temporal encoder, spatio-temporal trainable positional embeddings, and 12 transformers with self-attention. In contrast, CBraMod combines temporal and frequency-domain encoders with convolutional positional embeddings and 12 transformers with crisscross attention. Each encoder processes EEG at a fine resolution of one second per channel. This work adapts the encoders to create embeddings over longer, 90-sec, multi-channel EEG. Given input $x \in \mathbb{R}^{N_T \times N_E \times N_S}$, each encoder outputs features $h \in \mathbb{R}^{N_T \times N_E \times N_d}$ per electrode channel per segment, with $N_d = 200$ for both models and $N_T = 10$. Multiple 1-sec outputs within 16-sec segment were compressed by averaging, matching the long-range (slower manifestation) of cognitive load during the learning task, where instantaneous fluctuations are less informative. This yields the final feature vector h of size $2000 \times N_E$, with N_E channels varied per montage.

While BFMs may be able to handle variable electrode configurations, downstream models require fixed-size features. To achieve this, we apply spatial and temporal pooling techniques. These strategies standardize feature dimensions across participants, enabling interpretable and computationally efficient modeling of cognitive load.

Spatial pooling helps reduce variability across electrode configurations. We evaluated two strategies: (1) *Group-average*: Electrode features were averaged within 9 anatomically defined regions, shown in Fig. 2, preserving neuroscientific relevance for cognitive load estimation. (2) *Intersection*: A single electrode per region was retained, reducing dimensionality but potentially discarding informative signals. After spatial pooling, the features were flattened along the temporal dimension, resulting in a vector of size $N_T \times 9 \times N_d = 18000$.

Temporal pooling was applied after spatial pooling to aggregate features over time. We evaluated three strategies: (1) *Global pooling*: Aggregates all time steps into a single vector. (2) *Mean pooling*: Computes the average across time. (3) *Mean-standard deviation (MeanStd)*: Stacks mean and standard deviation over time, capturing central tendency and variability.

2.4. Estimating and Explaining Cognitive Load

Estimation: We estimate cognitive load as a continuous-value, similar to our ground truth, Section 2.2, with the downstream estimators:

1. **Linear Layer (Linear)**: with an $L1$ sparsity constraint to ensure optimal training given the large vector size.
2. **Dense Neural Network (DNN)**: consisting of two layer neural network of sizes ($N_T \times 9 \times N_d$, $N_T \times N_d$, 1) with batch normalization and ReLU activation.

Table 1: Average Pearson correlation comparing BFM pipelines with spectral baselines and prior state-of-the-art deep models (left), and spatial and temporal pooling ablation (right). Best spatial pooling per row is underlined; overall best is bolded.

(a) Feature and Estimator comparison.

Feature	Downstream Estimator layer			
	SVM	Linear	DNN	LSTM
PSD	0.120	0.147	0.133	**0.299**
EEGNet	–	0.143	0.157	–
EEGConf	–	0.110	0.147	–
CBraMod	**0.133**	0.030	0.108	0.115
LaBraM	0.113	**0.281**	**0.230**	0.201

(b) Comparison of Spatial and Temporal pooling strategies.

Pooling dimension		CBraMod			LaBraM		
Temporal	Spatial	SVM	Linear	DNN	SVM	Linear	DNN
Mean	Intersec.	0.031	-0.006	0.058	-0.052	<u>0.094</u>	-0.006
	GroupAvg	<u>0.137</u>	<u>0.005</u>	<u>0.070</u>	<u>0.069</u>	0.078	**0.248**
MeanStd	Intersec.	0.033	-0.054	<u>0.056</u>	-0.043	0.107	-0.053
	GroupAvg	**0.145**	<u>0.015</u>	<u>0.056</u>	<u>0.072</u>	**0.294**	<u>0.239</u>
Global	Intersec.	0.010	**0.067**	0.101	0.007	0.042	-0.025
	GroupAvg	<u>0.133</u>	0.030	**0.108**	**0.113**	<u>0.281</u>	0.230

3. **Support Vector Machine (SVM)**: with a fixed Radial Basis Function kernel (RBF) kernel.

LSTM was also used for comparison using globally pooled features, following its success in prior work [9], but excluded from later experiments due to incompatibility with the collapsed temporal axis.

Explainability: To ensure our models rely on brain-evoked features rather than other artifacts, we assess whether the frozen encoder outputs align with neuroscientific expectations. While many studies emphasize predictive accuracy, we augment our evaluation with a novel Partition SHAP-based probe to interpret model behavior [23, 25]. Partition SHAP is a model-agnostic method that accounts for correlated EEG features using a domain-aware hierarchical tree to compute Owen values, an efficient approximation of Shapley values. For CBraMod and LaBraM encoders we perturb input EEG signals and measure changes in model output to assign relevance scores to electrodes, simulating the effect of signal loss or artifacts. These scores are aggregated per participant and model, as we evaluate alignment with expected neural patterns.

2.5. Experiments

We formulate cognitive load estimation as a supervised regression task, optimizing models with mean squared error (MSE) loss and evaluating performance via Pearson correlation between predicted and ground-truth scores. To assess cross-subject generalization, we used nested cross-validation (CV), where test and validation participants were never seen during training, and were drawn from the final and most stable Cohort E, as it contained five individuals who completed all 90 trials. In total, 20 CV folds were created (20 different ways of selecting one for test and one for validation, out of five). All model feature experiments and pooling comparisons in Sections 3.1- 3.2 used the stable Cohort D for training, with 11 participants with a 32-channel configuration. The EEG variability experiments and the longitudinal study in Sections 3.3- 3.4 incorporated all 25 participants from cohorts A-D for training. All experiments used the nested-CV from Cohort E for evaluation (testing and validation).

We benchmarked against prior work using EEG power spectral density (PSD) features across five frequency bands [9], and deep learning baselines trained end-to-end, including EEGNet [26] and EEGConformer [27], with comparable downstream layers. Unlike our modular framework, these baselines integrate spatial-temporal processing internally, limiting adaptability to classical models like SVMs and LSTMs. All experiments were implemented in Python 3.11 with PyTorch 2.0. A tiny set of hyperparameters was considered; Linear, with $L1$ penalties in $\{0, 0.5, 1\}$; SVM, with fixed RBF kernels; DNN, with Adam optimizer with a cosine annealing scheduler, and lr in $\{5e-4, 1e-4, 5e-5, 1e-5\}$.

3. RESULTS

3.1. Model and Feature performance comparison

We benchmarked LaBraM and CBraMod against state-of-the-art PSD features [9] and deep networks trained from scratch for cognitive load estimation [26, 27]. Table 1 (a) reports average cross-validated Pearson correlation over Cohorts D+E, 16 participants on a 32-electrode setup. BFMs with group-average spatial and global temporal pooling, especially LaBraM, consistently outperform prior methods across all downstream estimator layers. LaBraM surpasses CBraMod despite the latter's asymmetric conditional positional encoding and larger pretraining dataset, likely due to its larger encoder (6.2M vs. 5.2M params) and learned spatio-temporal positional dictionary. BFMs provide high-resolution (per-second, per-channel) encodings, enabling flexible downstream integration and improved performance with end-to-end processing lasting less than a second with all model configurations. In contrast, PSD compresses information early, limiting performance except with LSTMs, while fixed models like EEGNet and EEGConformer require retraining per dataset and offer limited flexibility with marginal gains.

3.2. Ablation Pooling studies

We perform ablation studies to evaluate the impact of spatial and temporal pooling strategies on BFM feature aggregation across SVM, Linear, and DNN downstream estimators. LSTM was excluded, as Mean and MeanStd temporal pooling have a collapsed temporal axis. Table 1 (b) shows detailed results across models and pooling combinations. No temporal pooling method consistently outperformed others: Mean and MeanStd pooling effectively reduced dimensionality, while global pooling preserved more information, offering marginally better performance. In contrast, spatial pooling had a more pronounced effect. Across all downstream layers, group-average spatial pooling consistently outperformed intersection pooling, suggesting that these BFMs capture complementary electrode-level information that intersection pooling smears. In other words, the use of anatomically informed pooling strategies like group-average, allows for more expressive and generalizable representations, despite heterogeneous EEG configurations.

3.3. Effect of variable EEG configurations

To assess robustness to data variability, we augmented the training set of Cohort D with 14 additional participants from Cohorts A–C, which feature heterogeneous cap configurations. The same nested cross-validation setup was retained. Both CBraMod and LaBraM pipelines were retrained using the best-performing Linear layer, and
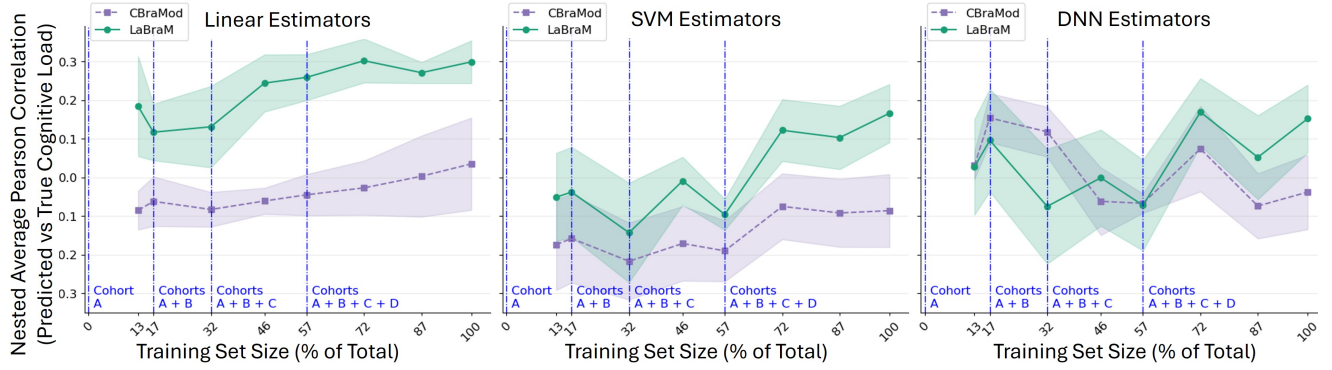
Fig. 3: Average correlation trends with confidence intervals over training size (points of cohort additions shown with blue verticals). LaBraM embeddings remain robust for the cognitive load task, while a Linear layer helps avoid overfitting to specific data cohorts. The x-axis appears not-rounded as each point reflects the addition of full participant trial sets.
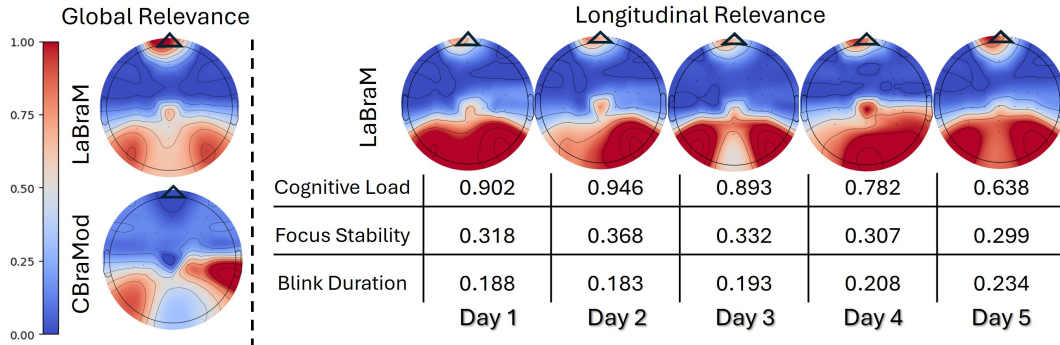

	Day 1	Day 2	Day 3	Day 4	Day 5
Cognitive Load	0.902	0.946	0.893	0.782	0.638
Focus Stability	0.318	0.368	0.332	0.307	0.299
Blink Duration	0.188	0.183	0.193	0.208	0.234

Fig. 4: Topomap of SHAP feature relevance, averaged globally (left) and per day (right); red indicates high importance. Daily averages of cognitive load (decreasing over time), focus stability (declining), blink duration (increasing over time) are also shown.

average correlation scores were plotted as a function of training set size (Fig. 3), expressed as a percentage of the full training set. Each point reflects the cumulative addition of full trial sets from participants, which results in uneven percentage increments (e.g., 17%, 32%, 46%), since not all participants completed all of their daily trials. Blue vertical lines indicating cohort additions. Confidence intervals are shown to illustrate variability across folds. Despite increased variability, both models improved whereas LaBraM achieved larger gains, indicating more robust and generalizable feature representations. The Linear layer appears robust against overfitting to cohort additions, while SVM and DNN seem more affected at the change points. Overall, SVM exhibited similar gaining trends across the x-axis, while DNN performance remained unaffected.

3.4. SHAP explainability results

Fig. 4 (left) shows normalized global SHAP feature relevance, averaged across all days and folds, while Fig. 4 (right) presents daily SHAP maps alongside behavioral metrics (cognitive load, focus stability, blink duration). LaBraM emphasizes frontal and prefrontal regions linked to cognitive control and decision-making [28], as well as parieto-occipital areas linked to visual working memory [29, 30]. In contrast, CBraMod primarily highlights parieto-occipital regions, consistent with the visual nature of the task, but lacks prefrontal emphasis, which may explain its weaker performance. These patterns were consistent across all estimators (SVM, Linear, DNN), underscoring the robustness of the extracted features and SHAP's model-

agnostic design. The longitudinal design, with participants returning over multiple days, enables analysis of temporal trends in cognitive load and training progression. Fig. 4 (right) presents daily SHAP relevance maps alongside behavioral metrics of focus stability and blink duration extracted from Varjo VR3. LaBraM shows increasing prefrontal relevance over time, while cognitive load decreases from 0.90 to 0.64 and blink duration increases from 0.19 to 0.23. This suggests participants became more proficient and cognitively efficient over time, while (potentially) more relaxed. These trends align with behavioral indicators such as reduced focus stability and longer blink durations, supporting the neurophysiological validity of LaBraM's feature representations [31]. In summary, prefrontal and parietal channels dominate SHAP relevance, with prefrontal emphasis differences explaining LaBraM and CBraMod performance gaps.

4. CONCLUSION

We presented a scalable pipeline for cross-subject cognitive load estimation using Brain Foundation Models (BFMs), with LaBraM outperforming CBraMod and other state-of-the-art models even with Linear estimators. Our approach generalizes across heterogeneous EEG setups and captures neurophysiologically meaningful patterns in the frontal regions. The pipeline runs in under a second on standard CPUs, supporting real-time inference with a longer sliding window. Our findings support the use of BFMs for cognitive load estimation, with future work aimed at integrating additional biometric signals and complementary insights from multiple XAI methods.

5. REFERENCES

[1] M. Nasri, M. Kosa, L. Chukoskie, 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, pp. 51–4.

[2] P.-J. Chen, W.-S. Chien, and C.-C. Lee, "Disentangle heart rate signals for improved stress detection," in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025, pp. 1–5.

[3] J. Wang, A. Wang, H. Hu, et al., "Multi-source domain generalization for ECG-based cognitive load estimation: Adversarial invariant and plausible uncertainty learning," in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024, pp. 1631–1635.

[4] N. Beauchemin, P. Charland, A. Karran, et al., "Enhancing learning experiences: EEG-based passive BCI system adapts learning speed to cognitive load in real-time, with motivation as catalyst," *Frontiers in Human Neuroscience*, vol. 18, pp. 1416683, 2024.

[5] A. Routray and S. Kar, "Classification of brain states using principal components analysis of cortical EEG synchronization and HMM," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 641–4.

[6] S. Saha and M. Baumert, "Intra-and inter-subject variability in EEG-based sensorimotor brain computer interface: a review," *Frontiers in computational neuroscience*, vol. 13, pp. 87, 2020.

[7] Y. Liu, Y. Yu, Z. Ye, et al., "Fusion of spatial, temporal, and spectral EEG signatures improves multilevel cognitive load prediction," *IEEE Transactions on Human-Machine Systems*, vol. 53, no. 2, pp. 357–366, 2023.

[8] J. Hassan, S. Reza, S. U. Ahmed, et al., "EEG workload estimation and classification: a systematic review," *Journal of Neural Engineering*, 2024.

[9] I. Tashev, C. Beauchene, R. M. Winters, et al., "Workload estimator using EEG and eye-tracking," in *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2024, pp. 1–5.

[10] K. Kyriaki, D. Koukopoulos, and C. A. Fidas, "A comprehensive survey of EEG preprocessing methods for cognitive load assessment," *IEEE Access*, vol. 12, pp. 23466–23489, 2024.

[11] L. C. Gómez, R. Hervás, I. Gonzalez, et al., "Studying the generalisability of cognitive load measured with EEG," *Biomedical Signal Processing and Control*, vol. 70, pp. 103032, 2021.

[12] Y. Zhou, X. Xu, and D. Zhang, "Cognitive load recognition in simulated flight missions: an EEG study," *Frontiers in Human Neuroscience*, vol. 19, pp. 1542774, 2025.

[13] Z. Li, R. Zhang, Y. Zeng, et al., "MST-net: A multi-scale swin transformer network for EEG-based cognitive load assessment," *Brain Research Bulletin*, vol. 206, pp. 110834, 2024.

[14] N. Panwar, V. Pandey, R. Tiwari, et al., "Combining spatio-temporal networks and graph attention architectures for EEG-based workload classification," in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025, pp. 1–5.

[15] S. Kuanar, V. Athitsos, N. Pradhan, et al., "Cognitive analysis of working memory load from EEG, by a deep recurrent neural network," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 2576–2580.

[16] D. Zhang, Z. Yuan, Y. Yang, et al., "Brant: Foundation model for intracranial neural signal," *Advances in Neural Information Processing Systems*, vol. 36, pp. 26304–26321, 2023.

[17] C. Yang, M. Westover, and J. Sun, "BIOT: Biosignal transformer for cross-data learning in the wild," *Advances in Neural Information Processing Systems*, vol. 36, pp. 78240–60, 2023.

[18] W.-B. Jiang, L.-M. Zhao, and B.-L. Lu, "Large brain model for learning generic representations with tremendous EEG data in BCI," in *The Twelfth International Conference on Learning Representations*, 2024.

[19] J. Wang, S. Zhao, Z. Luo, et al., "CBraMod: a criss-cross brain foundation model for EEG decoding," in *The Thirteenth International Conference on Learning Representations*, 2025.

[20] W. Cui, W. Jeong, P. Thölke, et al., "Neuro-gpt: Towards a foundation model for EEG," in *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2024, pp. 1–5.

[21] J. Sweller, "Cognitive load theory and educational technology," *Educational technology research and development*, vol. 68, no. 1, pp. 1–16, 2020.

[22] J. Lai, J. Wei, L. Yao, et al., "A simple review of EEG foundation models: Datasets, advancements and future perspectives," *arXiv preprint arXiv:2504.20069*, 2025.

[23] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017.

[24] I. J. Tashev, R. M. Winters, Y.-T. Wang, et al., "Towards a better scoring," in *IEEE Research and Applications of Photonics in Defense Conference (RAPID)*. IEEE, 2023, pp. 1–2.

[25] G. Van den Broeck, A. Lykov, M. Schleich, et al., "On the tractability of SHAP explanations," *Journal of Artificial Intelligence Research*, vol. 74, pp. 851–886, 2022.

[26] V. J. Lawhern, A. J. Solon, N. R. Waytowich, et al., "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," *Journal of neural engineering*, vol. 15, no. 5, pp. 056013, 2018.

[27] Y. Song, Q. Zheng, B. Liu, et al., "Eeg conformer: Convolutional transformer for eeg decoding and visualization," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 710–719, 2022.

[28] D. R. Euston, A. J. Gruber, and B. L. McNaughton, "The role of medial prefrontal cortex in memory and decision making," *Neuron*, vol. 76, no. 6, pp. 1057–1070, 2012.

[29] M. Koenigs, A. K. Barbey, B. R. Postle, et al., "Superior parietal cortex is critical for the manipulation of information in working memory," *Journal of Neuroscience*, vol. 29, no. 47, pp. 14980–14986, 2009.

[30] K. Grill-Spector and R. Malach, "The human visual cortex," *Annual Review Neuroscience*, vol. 27, no. 1, pp. 649–677, 2004.

[31] A. L. Callara, A. Greco, E. P. Scilingo, et al., "Neuronal correlates of eyeblinks are an expression of primary consciousness phenomena," *Scientific Reports*, vol. 13, no. 1, 2023.